Stable Belief and Stable Matching

Qingmin Liu*
Columbia University

October 17, 2017

Abstract

We study matching problems with transferable utility in the presence of adverse selection, and define a notion of stability, i.e., immunity to individual and pairwise deviations, as the consistency of publicly observable matching outcomes and uninformed agents' beliefs over informed agents' private types. The definition incorporates both "off-stability beliefs" conditional on the blocking of any deviating pairs, and "stable beliefs" in the absence of all such deviations. We define a notion of Bayesian efficiency of matching outcomes relative to endogenous stable beliefs, and investigate robust efficiency properties that stable beliefs and stable matchings must jointly satisfy. The idea of stable belief is extended to define the concept of the core, a refinement that differs from the notion of stability only if deviations by larger coalitions exist when no pairwise deviation is possible.

1 Introduction

This paper studies two-sided matching markets with transferable utilities, where agents on one side of the market are privately informed of their payoff-relevant attributes (types), and defines a solution concept of pairwise stability for such environments. A central component of our notion of stability is the consistency between uninformed agents' beliefs over informed agents' *unobservable* types and the matching of *observable* attributes. We call uninformed agents' beliefs under which observable matchings are individually rational and immune to pairwise blocking *stable beliefs*.

*E-mail: qingmin.liu@columbia.edu. For helpful conversations and comments, I am grateful to Pierre-André Chiappori, Alessandro Citanna, Bhaskar Dutta, Yinghua He, Philippe Jehiel, Nicolas Lambert, Jacob Leshno, Xi Zhi Lim, Philip Reny, Debraj Ray, Marzena Rostek, Ariel Rubinstein, Bernard Salanié, Anna Sanktjohanser, Rajiv Vohra, Yu Fu Wong, and seminar audiences at the University of Warwick, Oxford, UCL, Princeton, and Columbia. I thank Michael Borns for professional proofreading.

A theory of stability for two-sided markets with asymmetric information is required for at least two reasons. First, while the solution concept of stability proposed by Gale and Shapley (1962) and Shapley and Shubik (1971) has been successful in analyzing matching problems with complete information, only limited progress has been made on asymmetric information, which is widely recognized as a realistic feature in many two-sided matching markets. Examples include labor markets where firms are not fully informed of workers' abilities, insurance markets where the insured are privately informed of their risks, and marriage markets where the matching qualities of couples are unknown.[1]

Second, although the ideas of asymmetric information and adverse selection (screening and signaling in particular) have revolutionized economics for the past several decades and continue to be central in current theoretical and empirical research, the main analytical frameworks for two-sided markets with adverse selection have been competitive and non-cooperative game-theoretic ones, which are developed on the premise of unilateral deviation and optimization while holding the choices of all other agents fixed. In two-sided markets with pairwise relationships of agents, pairwise optimization and blocking that simultaneously involve two agents from opposite sides of the market are no less plausible than unilateral deviations. This natural consideration calls for a coalitional solution concept such as pairwise stability. In addition, the predictions of cooperative theory are based on assumptions on the primitive payoff structures, whereas modeling coalitional behaviors using non-cooperative game-theoretic models requires a full specification of every detail of their strategic or extensive forms, and their predictions are sensitive to the configuration of actions available to each player, orders of moves, protocols of price negotiation, rules of coalition formation, etc.

In the complete-information framework of Shapley and Shubik (1971) and Crawford and Knoer (1981), two agents from opposite sides of the market block a matching if they are not matched together but can strictly improve their respective payoffs by pairing up at a transfer; however, in a stable matching, no such blocking opportunity exists.[2] If we want to extend the notions of blocking and stability to a problem with asymmetric information, the following conceptual issue must be explained and resolved first. An uninformed player is uncertain about his ex post payoff from pairing up with an informed party who has payoff-relevant private information. As a Bayesian player, the uninformed party will assign a probabilistic belief to the informed party's private types. What should this belief be? First of all, this belief is an "off-stability belief" because no pairwise

[1]Since the seminal analysis of Becker (1973), a large literature has been developed to study various aspects of marriage and household problems using matching models with transfers. See Chiappori (2017) for the state of knowledge on both the theoretical and empirical fronts.

[2]It bears emphasis that the theory of stability abstracts away details such as how two players come together and how the transfer is negotiated (otherwise, ad hoc assumptions on extensive forms would have to be imposed). We shall follow this practice in this paper.

blocking opportunity exists in a stable matching. What then is the on-path, "stable belief" for a stable matching? An informed player's incentive of joining (and hence benefiting from) a blocking pair reveals its payoff-relevant information, and likewise the lack of such a blocking pair also reveals information. There are many blocking pairs to consider. Stability means that all of these pairwise deviations have been exhausted, and no further pairwise blocking opportunity exists, and hence on-path stable beliefs reflect the kind of information that is revealed by the non-existence of any further possible blocking opportunity. Therefore, in a stable matching, beliefs should be *endogenously* defined, as is the matching itself. In addition, off-stability beliefs should not be arbitrary; stable beliefs and off-stability beliefs must be simultaneously determined for all possible configurations of the off-stability blocking of a candidate matching.

In contrast to the normative design approach, our approach is to define a solution concept directly for matching games of asymmetric information. More specifically, we look for a belief and a matching outcome such that no unilateral and pairwise deviations from the matching can occur given the belief, and define stability—immunity to individual and pairwise blocking—as the consistency requirement for stable beliefs and matchings.[3] Our first requirement is that a stable belief be such that each uninformed player finds the matching individually rational in the expected utility sense with respect to this belief and assigns positive probability only to those types with which each informed player finds the matching individually rational. An off-stability belief of an uninformed player when an off-stability blocking opportunity arises is updated from the stable belief conditional on the event that the informed player benefits from the blocking. Our second requirement is that a stable belief be correct in the sense that either it excludes any types for any informed player with which the player has incentives to join some blocking pair, or its updated off-stability belief induces the uninformed players to turn down any blocking opportunity.

Defining stable beliefs over unobservable types consistently with stable matchings of observable attributes in a simple manner is the main conceptual contribution of the present paper. We define Bayesian efficiency as a criterion for a matching function that maximizes the total expected social surplus with respect to the supporting stable belief. We give conditions on matching values under which any stable matching must be Bayesian efficient, and these conditions include familiar models of adverse selection as special cases. The robust efficiency result is an implication of the consistency of stable beliefs and matchings; it does not involve the selection of beliefs or matchings.[4]

[3]In other words, we define a solution concept, but do not study mechanisms to implement the solution concept. This approach is consistent with that of various solution concepts in economics, such as the core, Bayesian Nash equilibrium, and so on.

[4]We call this a robust result for two reasons. First, it holds for all stable matchings of a given matching game. Second, it does not rely on assumptions on non-cooperative game forms.

The theoretical framework is simple and flexible for applied research, including empirical research. The distributions of unobservable types we economists can recover from matching data correspond to the uninformed players' stable beliefs over these unobservable types, as the outside analysts observe no more than the uninformed players in the matching problem. For instance, models of two-sided markets with transfers have found important applications in family economics and labor economics; see the two books devoted to this topic, Browning, Chiappori, and Weiss (2014) and Chiappori (2017). It is worthwhile to emphasize that existing empirical analysis of matching markets assumes a framework of complete information—although researchers may not observe all characteristics of the market, the players are always assumed to have complete information about the problem. If the matching data describe a stable market (as structural microeconomics usually assumes), then Shapley–Shubik complete-information stability theory provides conditions for economists to uncover the empirical distribution of unobservable types; see, e.g., Chiappori and Salanié (2016) for a survey of the econometrics of marriage models with transfers. Nevertheless, the effects of adverse selection have been left entirely unexamined because of the lack of theory. In our theory with asymmetric information, uninformed players have to fulfill the role of researchers to uncover the distribution of unobservable types, and the consistency of observable matching outcomes and stable beliefs over latent variables, together with possible exogenous restrictions that we discuss later in the paper, brings structures and disciplines to data.

The rest of the paper is organized as follows. Section 2 discusses related literature. Although the idea of stable belief distinguishes the present paper from other work, coalitional concepts with asymmetric information have been proposed in several branches of literature for various environments; we put this section upfront to explain our theoretical motivations and to clarify our conceptual contributions. Section 3 demonstrates the role of stable beliefs using two examples. Sections 4 and 5 contain the definitions of stable belief and stable matching. Section 6 defines the notion of Bayesian efficiency with respect to stable beliefs and presents robust efficiency properties of stable matchings. Bayesian efficiency is a natural subject of investigation as it concerns both matching outcomes and beliefs. Section 7 discusses several directions for future research, extends the framework to a richer environment with private beliefs and introduces a strict refinement of stability, the core, which is defined as immunity to deviations by arbitrary coalitions.

2 Related Literature

The restrictiveness of complete information in matching problems is widely recognized. Roth (1989) and Chakraborty, Citanna, and Ostrovsky (2010) are among the first to study matching problems with incomplete information. They look at static or multiple-stage

direct-revelation games of matching with non-transferable utilities.[5] The direct revelation mechanism excludes the crucial feature of stable belief that we emphasize in this paper. We define a solution concept for matching games, without imposing any mechanisms.[6] Recognizing the importance of inferences from the non-existence of blocking, Liu, Mailath, Postlewaite, and Samuelson (2014) take a belief-free approach to circumvent the issue of beliefs. The idea is to iteratively eliminate matching outcomes that are not immune to blocking for any possible belief over outcomes that survive prior eliminations (just as deletion of never-best responses in the definition of rationalizability). The theory critically hinges on the assumption that each uninformed agent knows perfectly his partner's type and hence his ex post payoff in the candidate matching. The elimination process cannot start if the uninformed agent instead has a slight uncertainty that does not restrict the support of his partner's private types.[7] By contrast, beliefs in a stable matching are the subject of investigation in the present paper, which does not impose the assumption that the uninformed agent knows perfectly the type of his match. As a result, the notion of stability in the present paper is neither a refinement nor a generalization of that of Liu et al. (2014), but much simpler.[8]

Prior to the work on matching, there was a large and deep literature on the core in incomplete information environments. In his pathbreaking paper, Wilson (1978) defines "coarse core" and "fine core," which differ in how deviating agents aggregate their private information. The literature after Wilson (1978) focuses on incentive compatibility of information aggregation, either by taking a mechanism design approach to blocking or by incorporating non-cooperative elements into the otherwise cooperative framework; see, e.g., Serrano and Vohra (2007) and Myerson (2007), among many others. Forges and Serrano (2013) survey unresolved questions as well as the state of the art in cooperative

[5]Chakraborty, Citanna, and Ostrovsky (2015) extend their early analysis to study group stability in many-to-one matching environments.

[6]The following analog in non-cooperative games of incomplete information may make the comparison clearer. One approach is to define a Bayesian Nash equilibrium for the incomplete information game. An alternative approach is to study incentive-compatible mechanisms and the implementability of Nash equilibria for the profiles of reported types.

[7]The difficulty is for the uninformed agent to compare his payoff from a candidate matching to that from a deviation. The belief-free approach essentially requires that the agent compare the worst-case payoff from the deviation with the best-case payoff in the candidate matching. If the uninformed player is "almost sure" about his partner's type in a candidate matching, but cannot rule out any other type, then the best-case payoff in the matching is higher than the worst-case payoff from the deviation, and hence it is not possible to remove anything. That is why the theory of Liu et al. (2014) obtains its prediction power from the assumption that each uninformed player knows perfectly the types of his own match.

[8]Even if a firm observes its own worker's type, the solution concept defined in the present paper is different (See Section 7.1). This difference is analogous to that between equilibrium and rationalizability in non-cooperative games, where the former solution concept assumes correct and common beliefs.

games of incomplete information.

Dutta and Vohra (2005) develop a notion of "credible core" in a direct revelation mechanism: the information that a deviating coalition should condition on is precisely the information that makes the deviation profitable; thus, in contrast to Wilson (1978), information aggregation of a deviating coalition is endogenously determined. The notion of credible belief updating was proposed by Grossman and Perry (1986) in their refinement of the sequential equilibrium of Kreps and Wilson (1982): the off-path belief is updated conditional on the set of types that would benefit from the deviation.[9] The basic idea appears in Rothschild and Stiglitz (1976) and Wilson (1977) when they consider off-equilibrium beliefs over types that are attracted by an off-equilibrium contract. We use the same idea to derive off-stability beliefs from stable beliefs in this paper. But our paper differs in crucial aspects from Dutta and Vohra (2005). In their model, agents do not learn from the absence of blocking and hence the notion of endogenous stable belief does not appear. Moreover, in line with the literature on the core with asymmetric information, they assumed that allocations are not observable; in our framework, players observe and make inference from matching outcomes, and the consistency of observables and beliefs over unobservables is key in defining stability.

Several aforementioned papers on stability and the core take a mechanism design approach, which has the advantage of abstracting away many details of the extensive forms; however, the approach is still restrictive for the purpose of defining stability. It is a widely accepted view that coalitional solution concepts capture properties of steady states or equilibria of reasonable dynamic game processes corresponding to the underlying coalition formation problem in the frictionless limit.[10,11] In traditional mechanism design models, the mechanism designer is assumed to have full commitment power, which excludes decentralized matching processes since these are better modeled by limited commitment. For matching problems in particular, the direct revelation game excludes the

[9]Kahn and Mookherjee (1995) adopted this belief refinement in their analysis of coalitional-proof Nash equilibrium, and Bloch and Dutta (2009) adopted it in the definition of coalitional-proof correlated equilibrium.

[10]For example, in the case of complete information, Perry and Reny (1994) show that the core can be implemented by stationary equilibria in a continuous-time coalition formation game; Gul (1989) shows that the Shapley value captures stationary equilibrium outcomes in a dynamic game of pairwise meeting and bargaining. Both papers choose extensive-form games by closely following the essential features captured by cooperative solution concepts. It is also well known that non-cooperative outcomes are highly sensitive to the game form. Asymmetric information adds another layer of complication: belief updating is extremely sensitive to the choice of extensive forms. Therefore, it is crucial that we do not make strong ad hoc restrictions on stable beliefs.

[11]A solution concept capturing some steady state of a dynamic process is not unique to cooperative solution concepts. Indeed, it is advocated as a justification of Nash equilibrium, the fundamental non-cooperative solution concept.

possibility of making inferences from observable matching outcomes and neglects the feature that exhausting all coalitional deviation possibilities, belief becomes an endogenous object consistent with observable outcomes (instead, static revelation games simply assume that there is no belief updating on path, and it is the off-path blocking that refines the set of allocations). The goal of this paper is to define precisely a descriptive solution concept as the consistency of the unobservable type distributions (stable beliefs) and the observable outcomes (matchings and transfers).

Another approach to coalitional deviations with incomplete information studies efficiency in collective decision problems. Holmström and Myerson (1983) consider efficiency in a direct incentive-compatible mechanism and the durability of allocation rules relative to a static voting game. Crawford (1985) discussed the role of imposing exogenous restrictions on game rules in collective decision problems. By contrast, we want to avoid mixing non-cooperative elements (direct mechanisms with full commitments or exogenous game forms) with a cooperative framework. The advantage of a cooperative framework of pairwise stability is precisely that it makes predictions based on the payoff structures alone, without making specific assumptions about the rules of the game. We do not use their notion of efficiency and durability. We propose and study the notion of Bayesian efficiency of matching outcomes relative to endogenous stable beliefs that support stable matching outcomes, without any need for prior beliefs, which are largely irrelevant in our framework.

Since we study stable beliefs in a class of coalitional games, it is worthwhile to compare them with equilibrium beliefs in non-cooperative incomplete-information games. Defining equilibrium beliefs in a consistent manner was the major conceptual task in the classic solution concepts of sequential equilibrium (Kreps and Wilson 1982) and perfect Bayesian equilibrium (Fudenberg and Tirole 1991). Because a game tree specifies every detail about how events during the game unfold, equilibrium beliefs that are consistent with equilibrium strategies (or outcomes) can be derived from priors using Bayes' rule. A purely cooperative approach to stable matching has the advantage of avoiding specifying non-cooperative game forms of decentralized matching processes (it is an overwhelming task to consider all game forms, and ruling out a subset of games requires ad hoc justifications), but the advantage comes with a cost: it is unclear how to derive posteriors from prior beliefs without a fixed game tree. Our strategy is not to specify the prior beliefs. We work with consistency between stable matching outcomes and stable beliefs, without making a priori restrictions on stable beliefs: stability imposes endogenous restrictions on stable beliefs and we derive robust properties from stability. In other words, our approach is prior-free, but not belief-free. We do make the implicit assumption that the underlying matching processes have common priors and observable actions, and hence that stable beliefs are shared by uninformed players; Section 7.1 extends the framework to private

stable beliefs that are consistent with a common prior.

3 Examples

3.1 Stability and Inefficiency

There are two firms with commonly known types: firm 1 is denoted as f_1 and firm 2 is denoted as f_2. There is one worker who privately knows his own type $w \in \{w_1, w_1'\}$. The two firms assign equal probability to the two types (for now, let us assume that the belief is given). The worker can match with at most one firm. An unmatched player's payoff is normalized to 0. The matrix of matching values of the worker and the firms is given below:

	f_1	f_2
w_1	$0, 6$	$2, -3$
w_1'	$4, -5$	$1, 7$

For instance, the (w_1, f_1) entry $(0, 6)$ means that the worker of type w_1 obtains a value of 0 by matching with the firm 1, and the firm obtains a value of 6 in this match. We consider a situation in which ex post matching values are not verifiable or contractible (the same assumption as in adverse selection models); otherwise asymmetric information would not play any role.

First observe that in this example ex post efficiency (i.e., matches that maximize the sum of players' realized payoffs) requires that the worker of type w_1 match with firm 1 and the worker of type w_1' match with firm 2. This is the only matching that has a non-negative realized sum of payoffs for a matched pair. Not surprisingly, it is the stable outcome prescribed by Liu et al. (2014), where it is assumed that a firm observes its worker's private type in a match. However, one would not expect the ex post efficient matching outcome to prevail in a situation where firms' uncertainty regarding the worker's private type is unresolved. The question then is what to expect if firms share a belief that the worker's type is w_1 or w_1' with equal probability.

Consider the following match, which is denoted by μ: regardless of his type, the worker is matched with firm 1 and receives a salary of 0; firm 2 stays unmatched. According to the matrix above, the worker of type w_1 receives a payoff of 0, and the worker of type w_1' receives a payoff of 4. Firm 1's expected payoff is $0.5 \times 6 + 0.5 \times (-5) = 0.5$. Players' expected payoffs are summarized as follows:

μ	w_1	w_1'	f_1	f_2
expected payoffs	0	4	0.5	0

Let us argue that this match is "stable" under the given belief; i.e., it is immune to

individual or pairwise deviations. Individual rationality is clear. Let us check whether firm 2 can lure the worker away from firm 1 and make a profit.

The type w_1 worker wants to switch to firm 2 because his matching value will be increased by 2; but the type w_1' worker's matching value will be reduced by 3. Therefore whenever w_1' finds firm 2's offer attractive, w_1 finds the same offer strictly more attractive; i.e., there is no way firm 2 can attract w_1' alone. Thus firm 2's offer either attracts w_1 alone, or gets both w_1 and w_1'. Nevertheless, firm 2's matching value with type w_1 is -3, so firm 2 does not want to get type w_1 alone. Therefore, in order for firm 2 to lure the worker away from firm 1, it must get both types w_1' and w_1, and the salary offer must be at least 3 to compensate for the type w_1' worker's loss in matching value. But firm 2's expected matching value from getting both w_1 and w_1' is only $0.5 \times (-3) + 0.5 \times 7 = 2$. Therefore firm 2 will not offer a salary of 3 to attract the worker, and hence the firm prefers to stay unmatched.

In the argument above, the on-path stable belief is that the worker is either w_1 or w_1' with equal probability; upon an off-path deviation in which firm 2's offer makes both worker types strictly better off, the firm's off-stability belief about the worker's type is that w_1 and w_1' appear with equal probabilities—a Bayesian updating from the stable belief conditional on the set $\{w_1, w_1'\}$. This updating rule is natural for this example, and is easily extended to general matching problems.

We have just argued that it is stable for firm 1 to hire the worker at a salary offer of 0 and for firm 2 to stay unmatched. We now argue that the following match, denoted by $\overline{\mu}$, is also stable: firm 2 hires the worker with a salary offer of 0; firm 1 is unmatched. According to the matrix of matching values above, players' expected payoffs from the match are as follows:

$\overline{\mu}$	w_1	w_1'	f_1	f_2
expected payoffs	2	1	0	2

Let us see why $\overline{\mu}$ is stable. In this match, if the worker were to switch to firm 1, the matching value for w_1' would be increased by 3 but the matching value for w_1 would be reduced by 2. Therefore, firm 1 cannot lure the type w_1 worker away from firm 2 without attracting type w_1' as well. Firm 1 does not want to attract type w_1' alone, because firm 1's payoff from a (w_1', f_1) match is -5. For firm 1 to attract both types, the salary offer must be at least 2 to compensate for the type w_1 worker's loss in matching value. However, firm 1's expected matching value from hiring the worker is $0.5 \times 6 + 0.5 \times (-5) = 0.5$. Therefore, firm 1 is not willing to pay a salary of 2 to attract the worker. Firm 1 prefers to stay unmatched. Therefore, $\overline{\mu}$ is indeed stable.

In summary, the example has two different stable matchings under the same belief. Both firms will benefit from hiring the single worker: firm 1 derives an expected matching

value of 0.5 and firm 2 derives an expected matching value of 2; but once the worker is assigned to one firm, he cannot be lured away by the other firm. In hindsight, the logic behind the counterintuitive result is clear: it is due to a match-specific adverse selection: players' matching preferences are not aligned with each other, although there is nothing non-generic about the matrix of matching values.

The purpose of this simple three-player example is to demonstrate the intuition of stability and updating of off-stability beliefs. It is also worthwhile to connect the observation with complete-information matching. Shapley and Shubik (1971) show that a stable matching under complete information is ex post efficient; i.e., it maximizes the total surplus. It follows immediately that all stable matchings of a given two-sided market under complete information lead to the same total surplus, and that for generic matching values, the stable matching function is unique.[12] The example demonstrates that none of these conclusions is true under asymmetric information, even though firms have symmetric beliefs.

3.2 Consistency Restrictions of Stable Beliefs

The previous example shows that a given belief can be consistent with a stable matching. The following three-player example shows that not all beliefs can be consistent with stable matchings. Consequently, stability imposes joint consistency restrictions on beliefs and matchings. In other words, stable beliefs must be an endogenous component of the definition of stability.

There is one worker whose type is drawn from $\{w_1, w_1'\}$ according to a prior belief that assigns probability $q \in (0,1)$ to w_1 and $1-q$ to w_1'. There are two firms: f_1 and f_2. The matrix of matching values is given below:

	f_1	f_2
w_1	$-1, 2$	$-3, 5$
w_1'	$0, 2$	$-4, 5$

We claim that it is impossible for the matching outcome (w_1', f_2), i.e., the worker of type w_1' matching with firm 2, to ever appear with positive probability in any stable matching. To prove this claim, note that the salary that the worker of type w_1' received from firm 2 cannot exceed 5, and hence the worker's payoff is at most 1 from this match. But the type w_1' worker could propose to pair up with the unmatched firm 1 and ask for a salary of $p = 1.5$, which would give him a payoff of $1.5 > 1$. Firm 1 would happily accept the proposal, knowing that it would make a positive profit regardless of the worker's type. Therefore, the match (w_1', f_2) cannot ever appear in a stable matching.

[12] Generically the unmatched players are the same across all stable matchings, which is a version of the Rural Hospital Theorem (Roth 1986) for matching games with transferable utilities.

Similarly, it is impossible for the matching outcome (w_1, f_1) to happen with positive probability in any stable matching. If (w_1, f_1) indeed appears in a stable matching, then the worker's payoff cannot exceed 1—because firm 1 will offer a salary of at most 2. The type w_1 worker could propose to pair up with the unmatched firm 2 for a salary of $p = 4.5$, which would give him a payoff of $1.5 > 1$. Firm 2 would accept the worker's proposal, knowing that it would make a positive profit regardless of the worker's types.

Thus we conclude that only the matching outcome (w_1, f_2) or (w_1', f_1) can possibly appear in any stable matching. But then there is a full separation of worker types in these two outcomes, irrespective of the distribution from which the worker's type is drawn—the consistent belief must be such that it assigns probability 1 to w_1 when the worker is matched with firm 2, and that it assigns probability 1 to w_1' when the worker is matched with firm 1. Thus stability forces the stable belief to be fully separating, and consequently the resulting matching is stable and efficient as if there were complete information.

Together, our two examples make clear that a priori restrictions on stable beliefs are problematic. Stability imposes endogenous restrictions on beliefs. The uniqueness of the stable belief in the second example is an implication of the specification of the matching values, which are the primitives assumed in the model. For general problems, it is possible that multiple beliefs and matchings are consistent. A common approach to dealing with equilibrium multiplicity in non-cooperative game theory is refinement. There are other reasons for multiplicity that cannot be easily refined away—the cooperative framework of stability is free of restrictive non-cooperative game forms. Instead of working with refinement, we choose an approach of showing robust properties of stability despite potential multiplicity.

4 The Model

The model is based on the matching games studied by Crawford and Knoer (1981). Let I be a set of n workers, and J be a set of m firms. Let W_i be a finite set of types for worker i. Worker i's type $w_i \in W_i$ is his private information. Denote by $w = (w_1, ..., w_n) \in W = \times_{i=1}^{n} W_i$ a profile of private types for the n workers. Firm j's type is commonly known which is summarized by j. Similarly, each worker i can also have publicly observable, payoff-relevant attributes which are summarized by i.

Let $a_{ij}(w_i) \in \mathbb{R}$ and $b_{ij}(w_i) \in \mathbb{R}$ be the *matching values* worker i (with type w_i) and firm j receive, respectively, when firm j hires worker i.[13] To ease notation, for a profile of workers' types $w = (w_i, w_{-i}) \in W$, we write $a_{ij}(w) := a_{ij}(w_i)$ and $b_{ij}(w) := b_{ij}(w_i)$.

[13]This formulation is general. For instance, it captures the case of $a_{ij}(w_i) = u_i(w_i, t_i, f_j)$, where t_i is worker i's observable characteristic and f_i is firm j's observable type, and there is no restriction on the dimensionality of (w_i, t_i, f_j).

We normalize the matching values of unmatched players to 0 and, with a slight abuse of notation, we write them as $a_{ii}(w) = b_{jj}(w) = 0$. With this notation, a *matching game* is fully summarized by the matching value function $(a, b) : I \times J \times W \to \mathbb{R}^2$.

A *matching* is a one-to-one function $\mu : I \cup J \to I \cup J$ such that the following holds for each $i \in I$ and $j \in J$: (i) $\mu(i) \in J \cup \{i\}$, (ii) $\mu(j) \in I \cup \{j\}$, and (iii) $\mu(i) = j$ if and only if $\mu(j) = i$. Here $\mu(i) = i$ means that worker $i \in I$ is unmatched; likewise for $\mu(j) = j \in J$.

Let $p_{ij} \in \mathbb{R}$ be the transfer that worker i receives from firm j. A *transfer scheme* $\mathbf{p}$ associated with a matching function μ is a vector that specifies a transfer $p_{i\mu(i)} \in \mathbb{R}$ for each $i \in I$ and $p_{\mu(j)j} \in \mathbb{R}$ for each $j \in J$, where $p_{ii} = p_{jj} = 0$. If worker i and firm j are matched together with a transfer p_{ij} when the profile of workers' types is w, worker i's and firm j's ex post payoffs are $a_{ij}(w) + p_{ij}$ and $b_{ij}(w) - p_{ij}$, respectively.

A firm's belief over the profiles of workers' types is a probability measure $\beta \in \Delta(W)$. Denote by β_i the marginal probability measure of β over worker i's types. If $\beta(w) = \times_{i=1}^{n} \beta_i(w_i)$ for all $w = (w_1, ..., w_n) \in W$, we say that the belief β is *independent* over individual workers' types.

Definition 1 A *matching* with asymmetric information is a triple $(\mu, \mathbf{p}, \beta)$ that consists of a matching function μ, a transfer scheme p associated with μ, and a belief β.

Given a matching $(\mu, \mathbf{p}, \beta)$, we ask whether any individual, or firm-worker pair, has incentives to deviate, in other words, whether the matching is *stable*. Before proceeding to the formal definition, we shall clarify several components of the environment.

Observables. Players' identities $i \in I$ and $j \in J$ summarize all of their publicly observable attributes. We assume that $\mathbf{p}$ and μ are publicly observable.[14] In fact, as shall become clear later on, because only uninformed firms need to make inferences from observables, technically speaking, it suffices to assume that they are public to firms, whereas workers observe only matches and prices in their own matches. The matching function $\mu : I \cup J \to I \cup J$, a function from observables to observables, specifies a set of distinct matched pairs of players. The workers' private types $w = (w_1, ..., w_n)$ are not made public in a match, and firms form a belief β about w. Hence, β and μ induce a (random) assignment of workers' types to firms, $(i, w_i) \mapsto j$, and a firm's inference on workers' private types w_i from the matching of i and j is the key component in the definition of stability. For this reason, defining a matching as an assignment of workers' types to firms, $(i, w_i) \mapsto j$, makes no difference than μ.

Common beliefs. All firms share the same beliefs β about the workers' types, because all firms observe exactly the same information: the set of matched pairs, the

[14] See Salanié (2015) for a discussion of the role of observable transfers and matchings from an empirical perspective.

transfer within each pair, and the firms' own publicly observable types. Implicitly, we assume that there is a common prior in the background (it is shared by the players but need not be observed by the analyst). Section 7.1 extends the notion of matching and stability to the environment where uninformed players observe private signals and have private beliefs.

We refrain from discussing the common prior assumption or the existence of a "prior stage" of a game. Interested readers may refer to Aumann (1998) and Gul (1998). The requirement of common stable belief in this case is no different from that of standard equilibrium concepts for non-cooperative games where uninformed players observe identical information and hence have common equilibrium beliefs. Consequently, β should not exclude the true states. Note also that beliefs on w can be correlated with observable characteristics summarized by (i, j). Since these characteristics are observed by all firms, the dependence of β on them should not be confused with the commonality of β.

Matching payoffs. In a matching $(\mu, \mathbf{p}, \beta)$, firm $j \in J$ is assigned a worker $\mu(j)$, but the firm does not observe $w = (w_1, ..., w_n)$. Therefore, from firm j's perspective, $a_{i\mu(i)}(\cdot)$ and $b_{\mu(j)j}(\cdot)$ are random variables defined on the state space W. Firm j's expected payoff from the matching is $\mathbb{E}_\beta\left[b_{\mu(j)j}\right] - p_{\mu(j)j}$. Worker i knows his own type w_i and hence he knows his ex post payoff $a_{i\mu(i)}(\cdot) + p_{i\mu(i)}$, although he is uncertain about other workers' payoffs.

Correlated beliefs. There is no reason to assume that the workers' types in a stable matching are i.i.d. Indeed, the example in Section 3.2 shows that such a restriction is not only unfounded but also problematic. As we shall see in Proposition 2, for the assignment games studied by Shapley and Shubik (1971), an important class of matching games, beliefs over individual workers in a stable matching are generally not identical. However, if the workers' private types are drawn independently from some prior distribution, then it would make sense to assume that the stable belief $\beta \in \Delta(W)$ is a product probability measure, i.e., $\beta(w) = \times_{i=1}^{n} \beta_i(w_i)$ for all $w \in W$. More generally, correlated beliefs add an interesting twist to the definition of stability. Suppose that a firm j, whose assigned partner is $\mu(j)$ from the matching $(\mu, \mathbf{p}, \beta)$, contemplates a deviation with firm $i \neq \mu(j)$. The firm needs to compare its payoffs from matching with i and with $\mu(j)$. The presence of belief correlation means that if the firm makes any inference on worker i's type (from i's willingness to deviate), it will revise its belief about its assigned worker $\mu(j)$ as well.

5 Stability

Stability defines the consistency between matching outcomes $(\mu, \mathbf{p})$ and belief β. The first consistency requirement is that the belief β be such that each uninformed player finds the matching $(\mu, \mathbf{p})$ individually rational in the expected utility sense with respect to this

belief and assigns positive probability only to those types with which the informed player finds the matching individually rational.

Definition 2 A matching $(\mu, \mathbf{p}, \beta)$ is *individually rational* if $a_{i\mu(i)}(w) + p_{i\mu(i)} \geq 0$ for all $i \in I$ and all w in the support of β, and $\mathbb{E}_\beta\left[b_{\mu(j)j}\right] - p_{\mu(j)j} \geq 0$ for all $j \in I$.

The second consistency requirement is that the belief β is correct in the sense that either it excludes any types of any informed player with which the player has incentives to join a blocking pair, or its updated off-stability belief induces the uninformed player to turn down any blocking opportunity. The key is to define the notion of pairwise blocking.

If worker i and firm $j \neq \mu(i)$ were to form a coalition with a transfer p to block the candidate matching—we denote this coalitional deviation by (i, j, p)—it must be that both players expect to be better off. Worker i knows his own type, so for him to join the deviation (i, j, p), we must have $a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}$ where the true type profile is $w = (w_1, ..., w_n)$. Let D_{ijp} be the set of w's with which worker i finds the deviation (i, j, p) profitable, i.e.,

$$D_{ijp} = \left\{w \in W : a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}\right\}.$$

Here D_{ijp} takes the form of $D_i \times W_{-i}$ for some $D_i \subset W_i$.

Knowing that worker i is willing to join the deviation (i, j, p), firm j updates its belief from $\beta(\cdot)$ to $\beta(\cdot|D_{ijp})$ using Bayes' rule.[15,16] Therefore, firm j's expected payoff from the deviation (i, j, p) is

$$\mathbb{E}_\beta[b_{ij}|D_{ijp}] - p. \tag{1}$$

Firm j will compare this payoff with its payoff from matching with $\mu(j)$ at $p_{\mu(j)j}$ in the given matching. Again, worker i's willingness to join the deviation (i, j, p) is the new information that leads firm j to reassess the matching payoff with $\mu(j)$ using the updated belief $\beta(\cdot|D_{ijp})$, because the types of worker i and $\mu(j)$ could be correlated under β. The firm's expected payoff from its original matching, under the new information, is

$$\mathbb{E}_\beta\left[b_{\mu(j)j}|D_{ijp}\right] - p_{\mu(j)j}. \tag{2}$$

[15]Our introduction of the blocking pair may seem to suggest a game form in which the worker first proposes to the firm, and then the firm updates its belief and responds. This is just one way to interpret the formation of a blocking pair. Indeed, in deciding whether to join a coalition, a player must condition on the event that a coalition is formed (and hence beneficial to the other player). This is similar to the winner's curse logic in auctions with interdependent values: a player computes his bid conditional on the information revealed by the event that he is a winner.

[16]As explained in the Introduction, this conditional belief is an "off-stability belief" when β is a "stable belief." The updating does not discriminate types in D_{ijp}; this is similar to the idea of credible belief updating proposed by Grossman and Perry (1986) in signaling games and the idea of off-equilibrium beliefs used by Rothschild and Stiglitz (1976) and Wilson (1977) in competitive insurance applications. Further refinement or coarsening of beliefs can be considered, but the issue is beyond the scope of the present paper.

It is likely that the expected payoff computed with the updated belief in (2) is negative, in which case, firm j would no longer find it individually rational to matching with worker $\mu(j)$. Therefore, (1) being strictly larger than (2) does not ensure that firm j will join the deviation (i,j,p). The deviation (i,j,p) is feasible only if

$$\mathbb{E}_\beta\left[b_{ij}|D_{ijp}\right]-p>\max\left\{0,\mathbb{E}_\beta\left[b_{\mu(j)j}|D_{ijp}\right]-p_{\mu(j)j}\right\}. \tag{3}$$

It is tempting to argue that the matching $(\mu,\mathbf{p},\beta)$ is invalidated as long as (2) is negative because firm j will reject the assigned worker $\mu(j)$, and, therefore, that the "max" operator in (3) is unnecessary for the notion of blocking of a matching. This argument is flawed. There are two cases under which (3) is violated. The first case is when (2) is larger than (1). This is the obvious case where (i,j,p) does not form a blocking coalition. The second case is when $(2)<(1)\leq 0$. In this case, worker i's incentive to work with firm j reveals to firm j that it should fire its worker $\mu(j)$. However, worker i understands that he will not be hired by firm j even after the firm fires its worker (because worker i knows the stable belief β and can replicate firm j's calculation). Therefore, worker i has no incentive to join the blocking in the first place. One might be tempted to argue that firm j can still pay worker i for the purpose of deducing information from him even though the worker will not be hired. But then all types of worker i would want to obtain the payment without matching with firm j, and hence no information would be revealed.

Definition 3 A matching $(\mu,\mathbf{p},\beta)$ is *blocked* by $(i,j,p)\in I\times J\times\mathbb{R}$ if $\beta(D_{ijp})>0$ and

$$\mathbb{E}_\beta\left[b_{ij}|D_{ijp}\right]-p>\max\left\{0,\mathbb{E}_\beta\left[b_{\mu(j)j}|D_{ijp}\right]-p_{\mu(j)j}\right\},$$

where

$$D_{ijp}=\left\{w\in W:a_{ij}(w)+p>a_{i\mu(i)}(w)+p_{i\mu(i)}\right\}.$$

An observation is that the blocking by the coalition (i,j,p) may reveal information about worker $i'\neq i$, when β is not independent, and hence could trigger a further deviation by (i',j',p'). This is indeed a possibility, but further blocking would not make a difference for our purpose, because the matching would already be invalidated as long as there is one blocking possibility.[17] We now summarize the definition of stability.

Definition 4 A matching $(\mu,\mathbf{p},\beta)$ is *stable* if it is individually rational and is not blocked by any $(i,j,p)\in I\times J\times\mathbb{R}$. If $(\mu,\mathbf{p},\beta)$ is a stable matching, we say that β is a *stable belief* that supports the observable matching outcome $(\mu,\mathbf{p})$.

[17]This should not be confused with the notion of farsightedness; see, e.g., Dutta and Vohra (2017). In addition, deviations simultaneously involving multiple workers and firms can be relevant, when opportunities for pairwise blocking do not exist and opportunities for blocking by larger coalitions do, regardless of whether β is independent or not. We shall define this core concept in Section 7.2.

If $(\mu, \mathbf{p}, \beta)$ is stable, and β assigns probability 1 to some type profile w, then by individual rationality (Definition 2), $a_{i\mu(i)}(w) + p_{i\mu(i)} \geq 0$ and $b_{\mu(j)j}(w) - p_{\mu(j)j} \geq 0$ for all $i \in I$ and $j \in J$. In addition, there is no (i, j, p) such that

$$a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)} \text{ and } b_{ij}(w) - p \leq b_{\mu(j)j}(w) - p_{\mu(j)j}.$$

This says that $(\mu, \mathbf{p})$ is stable when there is complete information about worker type profiles w as defined by Crawford and Knoer (1981). Thus, Definition 4 generalizes the standard notion without asymmetric information. The existence of a stable matching is immediate: let β assign probability 1 to any type profile w, and let $(\mu, \mathbf{p})$ be the complete-information stable matching under w; then $(\mu, \mathbf{p}, \beta)$ satisfies Definition 4. However, it is not true that any $\beta \in \Delta(W)$ can be a stable belief that supports some $(\mu, \mathbf{p})$, as demonstrated in Section 3.2; otherwise, the notion of stability would be powerless.

Our goal however is not about existence. Instead, we would like to find robust properties of all stable matchings of a matching problem. To put it differently, we want to uncover the restrictions stability imposes on the relationship between the observables $(\mu, \mathbf{p})$ and the distribution of unobservables given by the stable belief β. Although μ is a function from observables to observables, which alone does not impose a useful restriction, μ together with β determines the matching of types. To go beyond the inequalities that define stability, we need to look deeper into the payoff structures of the game.

A matching game is fully summarized by its matching value function $(a, b) : I \times J \times W \to \mathbb{R}^2$. The following result summarizes the general properties of stable matching and stable beliefs. All omitted proofs are in Appendix.

Proposition 1 *A stable matching exists for each matching game (a, b), and the correspondence from (a, b) to the set of stable matchings is upper hemicontinuous.*[18] *Furthermore, consider any vector of non-negative real numbers $(\lambda^1, ..., \lambda^K)$ such that $\sum_{k=1}^K \lambda^k = 1$. The following convexity properties hold:*

(i) *If $(\mu, \mathbf{p}^k, \beta)$ is stable for $k = 1, ..., K$, and β is independent, then $(\mu, \sum_{k=1}^K \lambda^k \mathbf{p}^k, \beta)$ is stable.*

(ii) *If $(\mu, \mathbf{p}, \beta^k)$ is stable and β^k is independent for $k = 1, ..., K$, then $(\mu, \mathbf{p}, \sum_{k=1}^K \lambda^k \beta^k)$ is stable.*

(iii) *If $(\mu, \mathbf{p})$ is a complete-information stable matching when the profile of workers' types is known to be w^k for each $k = 1, ..., K$, then $(\mu, \mathbf{p}, \beta)$ is a stable matching for any belief β with support $\{w^1, ..., w^K\}$.*

[18] A sequence of matchings $(\mu, \mathbf{p}^k, \beta^k)$ converges to $(\mu, \mathbf{p}, \beta)$ if $\mathbf{p}^k \to \mathbf{p}$ in the Euclidean metric and $\beta^k \to \beta$ if $\beta^k(w) \to \beta(w)$ for all $w \in W$. It is without loss of generality to ignore the convergence of matching functions $\mu^k \to \mu$ in the discrete topology since the set of players is finite. A sequence of matching games (a^k, b^k) converges to (a, b) if $(a_{ij}^k(w), b_{ij}^k(w)) \to (a_{ij}(w), b_{ij}(w))$ in the Euclidean metric for all $w \in W, i \in I$, and $j \in J$.

In the definition of blocking pairs, we work with conditional beliefs of the form $\beta(\cdot|D_{ijp})$, where D_{ijp}, the set of worker types that find (i,j,p) profitable, varies as $\mathbf{p}$ or β changes. This feature introduces non-convexity into the problem. Therefore, the convexity result above is different from that of complete information, and we need to assume independence of β to obtain the result here. In addition, the conditional probability measures of $\sum_{k=1}^{K}\lambda^k\beta^k$ are in general not in the convex hull of the conditional probability measures of β^k's. Although for a fixed subset of type profiles $D \subset W$, $\left(\sum_{k=1}^{K}\lambda^k\beta^k\right)(\cdot|D)$ is a convex combination of $\{\beta^k(\cdot|D)\}$, the convex combination weights vary with D and in general $\left(\sum_{k=1}^{K}\lambda^k\beta^k\right)(\cdot|D) \neq \sum_{k=1}^{K}\lambda^k\beta^k(\cdot|D)$, as explained in the proof of Proposition 1. For this reason, it is problematic to consider $(\mu, \sum_{k=1}^{K}\lambda^k\mathbf{p}^k, \sum_{k=1}^{K}\lambda^k\beta^k)$. Indeed, it does not follow from the stability of $(\mu,\mathbf{p}^k,\beta^k)$ that $(\mu, \sum_{k=1}^{K}\lambda^k\mathbf{p}^k, \sum_{k=1}^{K}\lambda^k\beta^k)$ is stable; it is easy to see that the latter may not even be individually rational.

6 Joint Restrictions on Beliefs and Matchings

6.1 Notions of Efficiency

Definition 5 A matching $(\mu,\mathbf{p},\beta)$ is *full-information efficient* if μ maximizes

$$\sum_{i=1}^{n}\left(a_{i\overline{\mu}(i)}(w)+b_{i\overline{\mu}(i)}(w)\right)$$

over all matchings $\overline{\mu}: I\cup J \to I\cup J$ for any w in the support of β.

Full-information efficiency maximizes total surpluses, and players are given the same weights in the calculation of social welfare. When there is complete information, Shapley and Shubik (1971) show that a stable matching maximizes the sum of individual players' surpluses. In the presence of asymmetric information, the criterion of full-information efficiency is very stringent as it requires that a single matching μ maximize the sum of individual surpluses for every realization of private types in the support of β.

It should be noted that although full-information efficiency is a strong property, satisfying it is not sufficient to render asymmetric information futile. For instance, it is not the case that whenever $(\mu,\mathbf{p},\beta)$ is full-information efficient and stable, $(\mu,\mathbf{p})$ is complete-information stable for any w in the support of β. Suppose that a firm is matched, say at a price of 0, with a worker whose type is w_1 or w_1' with equal probability; the matching values for the two types of the worker are identically 2, and the firm's matching values with the two types are 2 and -1, respectively. This one-firm one-worker matching is stable, but when w_1' is known, the matching at a price of 0 is not individually rational and hence not complete-information stable.

The following notion of efficiency is weaker than the notion of full-information efficiency.

Definition 6 A matching $(\mu, \mathbf{p}, \beta)$ is *Bayesian efficient* if μ maximizes

$$\mathbb{E}_\beta \left[\sum_{i=1}^{n} \left(a_{i\overline{\mu}(i)} + b_{i\overline{\mu}(i)} \right) \right]$$

over all matchings $\overline{\mu} : I \cup J \to I \cup J$.

The notion of Bayesian efficiency again requires that a matching maximize the sum of individual surpluses. The welfare weights for different types of the same worker are given by belief β. In a stable matching, the stable belief β is not the prior belief and each worker knows his realized type; hence $\mathbb{E}_\beta \left[a_{i\mu(i)} \right]$ should not be interpreted as the (ex ante) expected payoff of worker i. But from the perspective of a planner or an analyst (who observe the same information as the firms in our model), the stable belief is the probability with which workers' types appear in a *stable* matching, not in an *arbitrary* matching. This criterion of Bayesian efficiency is thus appealing and useful in evaluating the matching outcome that has reached a stable situation. However, as it is possible that the same matching outcome is stable under multiple beliefs, a conceptual problem arises as to which belief we should use to assess the efficiency of the matching. We avoid this problem by find conditions on the payoff of matching games under which all stable matchings are Bayesian efficiency with respect to all of their corresponding stable beliefs. We do sometimes make the assumption that β being independent.

By definition, if $(\mu, \mathbf{p}, \beta)$ is full-information efficient, then it is Bayesian efficient, but not the other way around. Section 3.1 provides an example in which stability satisfies neither efficiency criterion. Section 3.2 offers an example in which stability implies full-information efficiency and hence Bayesian efficiency. In the following example, there are multiple stable matchings, and only one of them is Bayesian efficient.

Example 1 Consider two workers and two firms. Suppose that $\beta = \beta_1 \times \beta_2$, where $\beta_1(w_1) = \beta_1(w_1') = \beta_2(w_2) = \beta_2(w_2') = \frac{1}{2}$. The value matrix is as follows:

	f_1	f_2
w_1	$0, 0$	$0, 0$
w_1'	$-2, 5$	$-2, 0$
w_2	$-3, 0$	$-3, 4$
w_2'	$-4, 9$	$-4, 4$

The first stable matching is as follows: firm 1 is matched with worker 1 at a price of 2, and firm 2 is matched with worker 2 at a price of 4. The second stable matching is as follows: firm 1 is matched with worker 2 at a price of 4, and worker 1 and firm 2 are unmatched. But the two outcomes differ in total surpluses. Only the second one is Bayesian efficient. ■

6.2 Stability and Efficiency

We now turn to economic assumptions on the matching values (a, b). Our payoff assumptions bring the model closer to the incomplete-information version of Shapley and Shubik (1971), but our model, which includes both multiple-object auctions and adverse selection problems as special cases, is more general.

An example of Shapley and Shubik (1971) is assigning houses with existing owners to new buyers. The owners each have a reservation value that depends only on the quality of their houses, not on the buyers' characteristics, whereas the buyers do care about the matching of houses and their own attributes. With asymmetric information, it matters which side of the market possesses private information. Our model with workers and firms is rich enough: we can interpret the problem as either allocating job positions to workers, i.e., firms are job (house) owners in the language of Shapley and Shubik (1971), or allocating workers to firms, i.e., workers are labor (house) owners. That is, the workers in our model can be interpreted as either "house owners" or "house buyers." Therefore, it is without loss of generality to consider the case where it is always the workers' side that has private information, as in our main model.

6.2.1 Benchmark: Full-Information Efficiency

Assumption 1 *$b_{ij}(w) = b_{ij}(w')$ for any $w, w' \in W$, $i \in I$ and $j \in J$.*

Assumption 1 says that the uninformed player j's matching value b_{ij} is independent of the informed player i's private types, although it can vary with their observable types that are summarized by i and j. A special case is $b_{ij}(\cdot) \equiv 0$, where the uninformed players care only about the transfers, and a privately informed player values the types of both players in a match $(a_{ij}(\cdot) + p)$. One application of this setting is multiple-object auctions in which privately informed bidders (workers) acquire heterogeneous objects (jobs). We do not make any restrictions on a_{ij}, although in auction applications, the bidder's valuation $a_{ij}(w)$ is usually positive. Under Assumption 1, b_{ij} can vary with j, which can be interpreted as the object owners' heterogeneous reservation values.

We show in the following result that full-information efficiency is obtained under Assumption 1. This result is easy to understand. We know from the auction literature that under this assumption a Vickrey–Clarke–Groves mechanism implements ex post efficient allocations. Our stability notion conforms to this classic result.

Proposition 2 *Suppose that Assumption 1 holds. Then a matching $(\mu, \mathbf{p}, \beta)$ is stable if and only if $(\mu, \mathbf{p})$ is a stable matching for any w in the support of β. Hence a stable matching $(\mu, \mathbf{p}, \beta)$ is full-information efficient.*

6.2.2 Bayesian Efficiency

Assumption 2 *$a_{ij}(w) = a_{ij}(w')$ for any $w, w' \in W$, $i \in I$, and $j \in J$.*

Assumption 2 says that $a_{ij}(w) = h(i,j)$ for some function h. There is no restriction on $b_{ij}(\cdot)$. That is, the privately informed players do not directly care about their own types, which are payoff-relevant for the uninformed players (the informed parties care about their types indirectly because they affect the matching outcomes).

A special case of Assumption 2 that is of applied interest is that $a_{ij}(\cdot) \equiv 0$. This case captures a labor market in which workers care only about the salaries they receive ($a_{ij}(\cdot) + p = p$), while firms value the workers' private types ($b_{ij}(\cdot) - p$).

A different assumption is that all public and private attributes are directly payoff-relevant for the informed players, but $a_{ij}(w)$ is separable in w and j:

Assumption 3 *$a_{ij}(w) = g(i,w) + h(i,j)$ for some functions g and h.*

A special case of Assumption 3 is the following familiar assumption adopted in many classic adverse-selection models such as signaling and screening.

Assumption 4 *$a_{ij}(w) = a_{ij'}(w)$ for any $w \in W$, $i \in I$, and j, $j' \in J$.*

This is to say, a worker does not value which firm he works for, but his own types may affect his reservation utilities or costs of effort, etc. This assumption allows $a_{ij}(w)$ to vary with the worker's private type w and the worker's identity i, which summarize all of his observable attributes, but the value is not allowed to vary with the firm's type, which is summarized in j.

The following result shows that Bayesian efficiency of stable matchings obtains under Assumptions 2 or 3.

Proposition 3 *A stable matching $(\mu, \mathbf{p}, \beta)$ is Bayesian efficient if one of the following properties is satisfied:*

(i) *Assumption 2 holds.*

(ii) *Assumption 3 holds and workers are fully matched: $\mu(i) \neq i$ for all $i \in I$.*

(iii) *Assumption 3 holds, $a_{ij}(\cdot)$ and $b_{ij}(\cdot)$ are co-monotonic[19] in w_i for all $i \in I$ and $j \in J$, and β is independent.*

Under condition (i), all stable matchings of a given game must be Bayesian efficient with respect to their respective stable beliefs. Bayesian efficiency under condition (ii)

[19]We say that $a_{ij}(\cdot)$ and $b_{ij}(\cdot)$ are co-monotonic in w_i if there exists a linear order on W_i such that both $a_{ij}(\cdot)$ and $b_{ij}(\cdot)$ are non-decreasing in w_i. This linear order naturally extends to W. It is allowed that one of these two functions is strictly increasing and the other is constant over a subset of W_i. Since the linear order is arbitrary, it does not matter whether we require the two functions to be non-decreasing or non-increasing.

above can be restated as follows: if Assumption 3 holds, then constrained Bayesian efficiency obtains for all stable matchings if the welfare comparison is restricted only to matched agents. It is easy to come up with assumptions to ensure that the short side of the market is fully matched. The co-monotonicity in condition (iii) is an appealing property. In the special case where w_i is a real variable that measures the worker's ability with the natural order of "greater than or equal to," $a_{ij}(w)$ can be interpreted as worker i's disutility from work and $b_{ij}(w)$ as the worker's output at the firm. The monotonicity of a_{ij} and b_{ij} says that the worker's disutility is decreasing in his ability and his output is increasing in his ability. Note that the co-monotonicity condition in condition (iii) is more general than this simple interpretation: the orders over W with respect to which a_{ij} and b_{ij} are monotonic can vary with i and j.

The proof utilizes a duality characterization of Bayesian efficiency. The subtlety arises because surplus maximization (as well as its dual minimization problem) involves unconditional expected payoffs $\mathbb{E}_\beta[a_{ij}]$ and $\mathbb{E}_\beta[b_{ij}]$, whereas stability must involve deviations of individual types w in the support of β as well as conditional expected payoffs conditional an arbitrary deviation. The three conditions are used to overcome this discrepancy.

The following two examples show, respectively, that the full-match restriction in condition (ii) and the independent belief assumption in condition (iii) are tight.

Example 2 There are two workers and one firm. Worker 1's types are w_1 and w_1'. Worker 2's type is w_2. Suppose that the matching values before transfers are made are as follows:

w_1	w_1'	w_2
$-1, 5$	$1, -2$	$0, 1$

Here $(-1, 5)$ means that by matching with the firm, worker 1 of type w_1 obtains a payoff of -1, and the firm obtains a payoff of 5. Suppose the firm assigns equal probabilities to w_1 and w_1'. This belief β supports a stable matching in which the firm hires worker 2 at a price of 0 and the total surplus is 1. This matching is stable because any deviation acceptable to the firm must involve worker 1 of type w_1, and the price must be at least 1 which attracts both types of worker 1. So the firm's expected payoff would be bounded above by 0.5, lower than what it gets in the matching with worker 2.

This stable matching is not Bayesian efficient. Bayesian efficiency requires that the firm be matched with worker 1, and that the weighted total surpluses be 1.5. But this efficient matching is not stable with belief β for any transfers. To see this, note that the firm must pay at least 1 to worker 1 and its expected payoff is at most 0.5. But the firm can switch to the unmatched worker 2 to obtain a larger payoff.

In this example, the firm needs to pay a high price to recruit worker 1 of type w_1 (which is more productive for the firm), but transfers between players are not counted toward the social surplus. Thus the example illustrates the conflict of incentives and

Bayesian efficiency. Note also that the matching values are co-monotonic in the worker's types. ■

Example 3 Consider a market with two workers and one firm. The matching values of each worker and the firm are co-monotonic, and are as follows:

$\underline{w}_1$	$\overline{w}_1$	$\underline{w}_2$	$\overline{w}_2$
$(0.5, 5)$	$(1, 6)$	$(-2, 4)$	$(-1.9, 12)$

Suppose the firm's belief is such that $\beta(\underline{w}_1, \underline{w}_2) = \beta(\overline{w}_1, \overline{w}_2) = \frac{1}{2}$. This belief is not independent. Bayesian efficiency requires that the firm be matched with worker 1, with an expected total surplus of 6.25. However, this belief supports the following inefficient stable matching: the firm hires worker 2 at a price of 2. In this stable matching, the firm's expected payoff is 6 and the total surplus is 6.05. To see that this matching is stable, notice that if the firm joins a deviating coalition with both types of worker 1, or with only type $\underline{w}_1$, its expected payoff is strictly lower than 6. If the firm deviates with type $\overline{w}_1$, the most it can get is 7, but when worker 1 is type $\overline{w}_1$, worker 2 must be type $\overline{w}_2$, from which the firm gets a much larger payoff. ■

7 Discussion and Extension

The main purpose of this paper is to provide a definition of stability in two-sided markets with asymmetric information. Specifically, we point out that stable beliefs must be a component of this definition. There are several ways in which future work can build on this model.

First of all, refinements of off-stability beliefs are natural research questions. It should be emphasized that, without a more structured model, it is unclear whether a further refinement of off-stability beliefs will make blocking easier or not. For instance, the worker type that has the strongest incentive to deviate might be the firm's least preferred type. Second, the assumption of perfectly transferable utility is important for our result of efficiency; it will be interesting to consider imperfectly transferable utilities or costly signaling that must entail some inefficiency. Third, if we consider a model with a continuum of agents with no aggregate uncertainty, stable beliefs must be further restricted, and some strong implications of stability could be obtained. In what follows, we shall offer two other extensions which we regard as important.

7.1 Private Beliefs

In our model, firms share a common belief in a matching, which is reasonable because all firms are assumed to observe the same thing. A natural question is to model private

observations of the uninformed players and their private beliefs. The approach introduced in this paper can be extended to this richer environment.

7.1.1 Modeling Private Beliefs with Private Signals

Let S_j be the finite set of signals of firm $j \in J$. We denote by $s = (s_1, ..., s_m)$ the profile of signals of the m firms, and write $S = \times_{j \in J} S_j$. Firm j observes its own signal $s_j \in S_j$, but is uncertain about workers' types $w = (w_1, ..., w_n)$ and other firms' signals $s_{-j} = (s_1, ..., s_{j-1}, s_{j+1}, ..., s_m)$. The firm's private belief, as a function of s_j, should be an *endogenous* object defined in stability, just as the stable belief studied in previous sections. Let us write this belief as $\sigma(\cdot|s_j)$. With the usual abuse of notation, we treat $\sigma(\cdot|s_j)$ as a probability measure in $\Delta(W \times S)$, with the understanding that it is degenerate on S_j.

We assume that all firms' beliefs are consistent in the following sense: the types of the n workers and the signals of the m firms, (w, s), follow a common probability measure $\sigma \in \Delta(W \times S)$, and hence firm j's private belief over the type profiles of all workers and the signal profiles of all firms should be updated from σ. That is, after observing s_j, firm j's private belief $\sigma(\cdot|s_j) \in \Delta(W \times S)$ is the conditional probability measure $\sigma(\cdot|W \times \{s_j\} \times S_{-j})$.

To ease notation, we write $\sigma(w)$ as $\sigma(\{w\} \times S)$, write $\sigma(s_j)$ as $\sigma(W \times \{s_j\} \times S_{-j})$, write $\sigma(w|s_j)$ as $\sigma(\{w\} \times S|s_j)$, and write $\sigma(w_i|s_j)$ as $\sigma(\{w_i\} \times W_{-i} \times S|s_j)$.

Definition 7 A *matching with private signals* is a triple $(\mu, \mathbf{p}, \sigma)$, where $\mu : I \cup J \to I \cup J$ is a matching, $\mathbf{p}$ is a price scheme associated with μ, and each firm $j \in J$ has a private belief $\sigma(\cdot|s_j) \in \Delta(W \times S)$ when its private signal is $s_j \in S_j$. We say that a belief σ is *independent* if $\sigma(w|s_j) = \times_{i \in I} \sigma(w_i|s_j)$ for all $w \in W$, $s_j \in S_j$, and $j \in J$.

7.1.2 Stability with Private Beliefs

We shall define the stability of $(\mu, \mathbf{p}, \sigma)$ as the consistency between the *matching outcome* $(\mu, \mathbf{p})$ and the *system of beliefs* $(\sigma(\cdot|s_j))_{s_j \in S_j, j \in J}$. We write as

$$\mathbb{E}_\sigma [b_{ij}|s_j] = \sum_{w \in W} b_{ij}(w) \sigma(w|s_j)$$

firm j's expected matching value when matched with worker i conditional on observing a private signal s_j.

Definition 8 A matching with private signals $(\mu, \mathbf{p}, \sigma)$ is *individually rational* if $a_{ij}(w) + p_{i\mu(i)} \geq 0$ for any w such that $\sigma(w) > 0$ and $\mathbb{E}_\sigma [b_{ij}|s_j] - p_{\mu(j)j} \geq 0$ for all s_j such that $\sigma(s_j) > 0$.

Let $D_{ijp} = \left\{w : a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}\right\}$ be a set of types such that worker i prefers to deviate to firm j at a price p rather than stay in the candidate matching with $\mu(i)$. Observing this event D_{ijp}, firm j with a private belief $\sigma(\cdot|s_j)$ will update its belief to $\sigma(\cdot|s_j, D_{ijp})$ by Bayes' rule. We denote by $\mathbb{E}_\sigma[b_{ij}|s_j, D_{ijp}]$ the expectation of $b_{ij}(\cdot)$ with respect to belief $\sigma(\cdot|s_j, D_{ijp})$.

Definition 9 A matching with private signals $(\mu, \mathbf{p}, \sigma)$ is *blocked* by $(i, j, p) \in I \times J \times \mathbb{R}$ if $\sigma(D_{ijp}) > 0$ and

$$\mathbb{E}_\sigma[b_{ij}(w)|s_j, D_{ijp}] - p > \max\left\{0, \mathbb{E}_\sigma\left[b_{\mu(j)j}(w)|s_j, D_{ijp}\right] - p_{\mu(j)j}\right\}$$

for some s_j such that $\sigma(s_j) > 0$, where

$$D_{ijp} = \left\{w : a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}\right\}.$$

Since workers have no uncertainty about payoff-relevant parameters, their incentives to deviate are formulated the same way as in the case of no private signals. A firm's incentive is slightly more complicated. The definition above requires that firm j want to deviate for *some* s_j in the support of σ, but not for all s_j. This is a reasonable requirement. As long as the firm wants to deviate for some realization of its private signals not ruled out by the belief σ, the matching should be considered blocked. If a matching is not blocked by (i, j, p), then it must be that either $\sigma(D_{ijp}) = 0$ or

$$\mathbb{E}_\sigma[b_{ij}|s_j, D_{ijp}] - p \leq \max\left\{0, \mathbb{E}_\sigma\left[b_{\mu(j)j}|s_j, D_{ijp}\right] - p_{\mu(j)j}\right\}$$

for *all* s_j such that $\sigma(s_j) > 0$.

Definition 10 A matching with private signals $(\mu, \mathbf{p}, \sigma)$ is *stable* if it is individually rational and is not blocked by any $(i, j, p) \in I \times J \times \mathbb{R}$. If $(\mu, \mathbf{p}, \sigma)$ is stable, we say that $\sigma \in (W \times S)$ is a *stable belief* that supports the observable matching outcome $(\mu, \mathbf{p})$.

7.1.3 Exogenous Restrictions on Stable Beliefs

The definition of stable beliefs in Definition 10 is general and flexible; depending on applications, context-specific restrictions can be imposed on stable beliefs in addition to consistency. We present two examples to illustrate this point. In some applications, firms observe some attributes of their *own* workers after a match is formed. To model this partial observability, let $W_i = W_i^1 \times W_i^2$, where the set W_i^1 denotes the attributes observable to worker i's employer, and W_i^2 denotes the unobservables. For each $j \in J$, write as

$$S_j = \begin{cases} W_{\mu(j)}^1 & \text{if } \mu(j) \neq j \\ \{\varnothing\} & \text{otherwise} \end{cases}$$

the set of private signals observed by firm j. Firms' partial observation of their own workers require that the support of σ be

$$\left\{(w,s) \in W \times S : w^1_{\mu(j)} = s_j \text{ for each } j \neq \mu(j)\right\}.$$

In this example, the signal space depends on the observable matching function μ.

In some other applications, every firm in a cohort knows the types of workers in a cohort. To model the cohorts of players, we define partitions $J = \cup_{k=1}^K J_k$ and $I = \cup_{k=1}^K I_k$; to model the feature that each cohort J_k observes the types of cohort I_k, we let $S_j = \times_{i \in I_k} W_i$ for each $j \in J_k$ and require that the support of σ be

$$\left\{(w,s) \in W \times S : (w_i)_{i \in I_k} = s_j \text{ for each } j \in J_k \text{ and } k = 1, ..., K\right\}.$$

7.1.4 Efficiency

We now turn to the notion of efficiency, which imposes joint restrictions on μ and σ. Full-information efficiency is defined exactly as in Definition 5. Proposition 2 still holds because under Assumption 1 private signals provide no payoff-relevant information.

With the understanding that $\mathbb{E}_\sigma[a_{ij}] = \sum_{s \in S} \sum_{w \in W} a_{ij}(w) \sigma(w,s)$ and $\mathbb{E}_\sigma[b_{ij}] = \sum_{s \in S} \sum_{w \in W} b_{ij}(w) \sigma(w,s)$, the expected total surplus with respect to belief σ from a matching $(\mu, \mathbf{p}, \sigma)$ can again be written as

$$\mathbb{E}_\sigma\left[\sum_{i=1}^n \left(a_{i\mu(i)} + b_{i\mu(i)}\right)\right].$$

Hence the notion of Bayesian efficiency in Definition 6 immediately carries over here.

Definition 11 A matching with private signals $(\mu, \mathbf{p}, \sigma)$ is *Bayesian efficient* if μ maximizes

$$\mathbb{E}_\sigma\left[\sum_{i=1}^n \left(a_{i\overline{\mu}(i)} + b_{i\overline{\mu}(i)}\right)\right]$$

over all matchings $\overline{\mu} : I \cup J \to I \cup J$.

It should be noted that Assumptions 2 and 3 in Proposition 3 are about $a_{ij}(w)$, and hence Proposition 3 continues to hold. Nevertheless, the exposition of the proof needs to be adjusted because a blocking condition is defined for each s_j.

7.2 Core: Beyond Pairwise Deviations

In our notion of stability, a deviation involves at most two agents, one from each side of the market. Pairwise deviations are natural in two-sided markets. Conceptually, it is interesting to consider deviations by a coalition of multiple firms and multiple workers. This leads us to consider the concept of the core. In complete information matching

games, the core and stability coincide, as established by Shapley and Shubik (1971). However, the two concepts differ under asymmetric information.

As before, beliefs should still be a component of the definition of the core. Given a matching $(\mu, \mathbf{p}, \beta)$, individual rationality with respect to belief β is defined in Definition 2. Consider the following deviation. A subset of workers $\bar{I} \subset I$ and a subset of firms $\bar{J} \subset J$ walk away from the given matching and rematch among themselves with a matching function $\bar{\mu} : \bar{I} \cup \bar{J} \to \bar{I} \cup \bar{J}$ and a transfer scheme $\bar{\mathbf{p}} = \left(\bar{p}_{i\bar{\mu}(i)}\right)_{i \in \bar{I}}$.[20] They block the matching if each of them is strictly better off from this rematching, conditional on the information revealed by their agreement to participate in the deviation; again, there is no timing issue as to when the information is revealed, because each player, in evaluating his expected payoff from joining the coalition, must condition on the (anticipated) event that the coalition will be formed, i.e., everyone's agreeing to join the coalition.

Definition 12 A matching $(\mu, \mathbf{p}, \beta)$ is *blocked* by a coalition $\left(\bar{I}, \bar{J}, \bar{\mu}, \bar{\mathbf{p}}\right)$ if $\beta(D) > 0$ and

$$\mathbb{E}_\beta \left[b_{\bar{\mu}(j)j} | D \right] - \bar{p}_{\bar{\mu}(j)j} > \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D \right] - p_{\mu(j)j} \right\} \tag{4}$$

for all $j \in \bar{J}$, where

$$D = \left\{ w \in W : a_{i\bar{\mu}(i)}(w) + \bar{p}_{i\bar{\mu}(i)} > a_{i\mu(i)}(w) + p_{i\mu(i)} \text{ for all } i \in \bar{I} \right\}.$$

In this definition, D is the set of workers' types $(w_1, ..., w_n)$ with which all workers in $\bar{I}$ find the rematch profitable. The "max" operator in the definition of a firm's blocking needs a remark. The definition excludes the possibility of $\bar{\mu}(j) = j$ for some $j \in \bar{J}$ because otherwise the left-hand side of (4) becomes $\mathbb{E}_\beta \left[b_{\bar{\mu}(j)j} | D \right] - \bar{p}_{\bar{\mu}(j)j} = 0$. But this exclusion is without loss of generality because an unmatched firm j does not contribute any information or value to the blocking by other players.

We summarize the definition of the core below.

Definition 13 A matching $(\mu, \mathbf{p}, \beta)$ is in the *core* if it is individually rational and is not blocked by any coalition $\left(\bar{I}, \bar{J}, \bar{\mu}, \bar{\mathbf{p}}\right)$, where $\bar{I} \subset I$, $\bar{J} \subset J$, $\bar{\mu} : \bar{I} \cup \bar{J} \to \bar{I} \cup \bar{J}$ is a matching, and $\bar{p} = \left(\bar{p}_{i\bar{\mu}(i)}\right)_{i \in \bar{I}}$ is a transfer scheme. If $(\mu, \mathbf{p}, \beta)$ is in the core, we say that β is a *core belief* that supports the observable matching outcome $(\mu, \mathbf{p})$.

If $(\mu, \mathbf{p}, \beta)$ is in the core, then it is not blocked by any coalition including a pairwise deviation $(i, j, p) \in I \times J \times \mathbb{R}$; therefore, it is stable. It is immediate that the core is a refinement of the concept of stability, and hence our previous results on efficiency still hold for the new solution concept.

[20] We can further generalize the notion so that a player receives payment from someone outside of his own match.

Proposition 4 If $(\mu, \mathbf{p}, \beta)$ is in the core, then it is stable.

The reverse of Proposition 4 is not true. A stable matching is not necessarily in the core. The following example demonstrates the subtle reason for the core being a strict refinement of stability even when β is independent: a blocking by a larger coalition can be found when a pairwise blocking does not exist. The example has a pair of a firm and a worker who are matched together in the given matching, but both deviate to rematch with other players. It is precisely its own worker's incentive to join a deviating coalition that reveals to the firm that its payoff from the putative matching is actually lower than it has thought, which incentivizes the firm to rematch with the other worker; meanwhile, the deviation of the firm's own worker is made possible precisely for the same reason: the other firm accepts him because the other worker's deviation reveals information. This coalitional deviation must simultaneously involve two pairs of workers and firms.

Example 4 Consider two workers and two firms. Suppose that $\beta = \beta_1 \times \beta_2$, where $\beta_1(w_1) = \beta_1(w_1') = \beta_2(w_2) = \beta_2(w_2') = \frac{1}{2}$. The matrix of matching values is as follows:

	f_1	f_2
w_1	$0, -1$	$1, 1$
w_1'	$0, 7$	$-2, 0$
w_2	$1, 1$	$0, -1$
w_2'	$-2, 0$	$0, 7$

The following matching is stable under belief β: worker i is assigned to firm $j = i$, and the salaries of both workers are 0. In this matching, the expected payoffs for both firms are 3. To see this is stable, note that for each $i = 1, 2$, worker i of type w_i has an incentive to deviate to firm $3 - i$. But firm $3 - i$ has no incentive to accept this worker: its payoff from hiring the worker is at most 2.

This stable matching is not in the core. The deviation involves a rematching of both workers when their types are w_i with a transfer of 0. Given that worker $i = 1, 2$ finds it profitable by switching to firm $j = 3 - i$, both firms infer that worker i has type w_i instead of w_i'. With this information, firm $j = i$ knows that its payoff in the original matching is actually -1. For this reason, firm i is willing to accept worker $3 - i$. ■

Appendix

A Proof of Proposition 1

A.1 Upper Hemicontinuity

Suppose that $(\mu, \mathbf{p}^k, \beta^k)$ is a stable matching for the matching game (a^k, b^k), with $(\mu, \mathbf{p}^k, \beta^k) \to (\mu, \mathbf{p}, \beta)$ and $(a^k, b^k) \to (a, b)$ as $k \to \infty$. By the definition of stability, $(\mu, \mathbf{p}^k, \beta^k)$ is in-

dividually rational, i.e.,

$$a^k_{i\mu(i)}(w) + p^k_{i\mu(i)} \geq 0 \text{ if } \beta^k(w) > 0$$

and

$$\sum_{w\in W} \beta^k(w)\, b^k_{\mu(j)j}(w) - p^k_{\mu(j)j} \geq 0.$$

Taking the limit $k \to \infty$, it follows immediately that

$$a_{i\mu(i)}(w) + p_{i\mu(i)} \geq 0 \text{ if } \beta(w) > 0$$

and

$$\sum_{w\in W} \beta(w)\, b_{\mu(j)j}(w) - p_{\mu(j)j} \geq 0.$$

That is, $(\mu, \mathbf{p}, \beta)$ is individually rational for the matching game (a, b).

We would like to show that $(\mu, \mathbf{p}, \beta)$ is not blocked by any $(i, j, p) \in I \times J \times \mathbb{R}$. Let

$$D_{ijp} = \left\{w \in W : a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}\right\}. \tag{5}$$

Since W is finite, there exists $\varepsilon > 0$ such that

$$D_{ijp} = \left\{w \in W : a_{ij}(w) + p - 2\varepsilon > a_{i\mu(i)}(w) + p_{i\mu(i)}\right\}. \tag{6}$$

Define

$$D^k_{ij(p-\varepsilon)} := \left\{w \in W : a^k_{ij}(w) + p - \varepsilon > a^k_{i\mu(i)}(w) + p^k_{i\mu(i)}\right\}.$$

Since $(a^k, b^k) \to (a, b)$ and $\mathbf{p}^k \to \mathbf{p}$, it follows that there exists $K_1 > 0$ such that if $k > K_1$, then

$$\left|(a^k_{ij}(w) - a^k_{i\mu(i)}(w) - p^k_{i\mu(i)}) - \left(a_{ij}(w) - a_{i\mu(i)}(w) - p_{i\mu(i)}\right)\right| < \varepsilon \tag{7}$$

for any $w \in W$. It follows that if $w \in D^k_{ij(p-\varepsilon)}$ and $k > K_1$, then

$$a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}.$$

It then follows from (5) that $D^k_{ij(p-\varepsilon)} \subset D_{ijp}$ for $k > K_1$. Moreover, by (6) and (7), if $w \in D_{ijp}$ then $w \in D^k_{ij(p-\varepsilon)}$ for $k > K_1$. We conclude that for any $k > K_1$, $D_{ijp} = D^k_{ij(p-\varepsilon)}$.

Consider $k > K_1$. Since the matching $(\mu, \mathbf{p}^k, \beta^k)$ is stable for the matching game (a^k, b^k), it is not blocked by $(i, j, p-\varepsilon) \in I \times J \times \mathbb{R}$, i.e., if $\beta^k(D^k_{ij(p-\varepsilon)}) > 0$ then

$$\sum_{w\in W} \beta^k(w|D^k_{ij(p-\varepsilon)}) b^k_{ij}(w) - (p-\varepsilon) \leq \max\left\{0, \sum_{w\in W} \beta^k(w|D^k_{ij(p-\varepsilon)}) b^k_{\mu(j)j}(w) - p^k_{\mu(j)j}\right\}. \tag{8}$$

Note that $D_{ijp} = D^k_{ij(p-\varepsilon)}$. Since $\beta^k \to \beta$ by assumption, there exists $K_2 > K_1$ such that $\beta(D_{ijp}) > 0$ implies that $\beta^k(D_{ijp}) > 0$ for any $k > K_2$.

The above condition implies that

$$\sum_{w\in W} \beta^k(w|D_{ijp}) b^k_{ij}(w) - (p-\varepsilon) \leq \max\left\{0, \sum_{w\in W} \beta^k(w|D^k_{ijp}) b^k_{\mu(j)j}(w) - p^k_{\mu(j)j}\right\}.$$

Taking $k \to \infty$,

$$\sum_{w \in W} \beta(w|D_{ijp}) b_{ij}(w) - (p - \varepsilon) \leq \max\left\{0, \sum_{w \in W} \beta(w|D_{ijp}) b_{\mu(j)j}(w) - p_{\mu(j)j}\right\}.$$

Hence

$$\sum_{w \in W} \beta(w|D_{ijp}) b_{ij}(w) - p \leq \max\left\{0, \sum_{w \in W} \beta(w|D_{ijp}) b_{\mu(j)j}(w) - p_{\mu(j)j}\right\}.$$

The proof is completed. ■

A.2 Convexity

Since in general $\sum_{k=1}^{K} \lambda^k \beta^k(\cdot|D) \neq \left(\sum_{k=1}^{K} \lambda^k \beta^k\right)(\cdot|D)$, where $D \subset W$, we first prove the following property.

Lemma 1 *Let $D \subset W$ such that $\beta^k(D) > 0$ for $k = 1, ..., K$. Let $\beta = \sum_{k=1}^{K} \lambda^k \beta^k$, where $\lambda^k \geq 0$, $k = 1, ..., K$, and $\sum_{k=1}^{K} \lambda^k = 1$. Then there exist non-negative real numbers $\alpha^1, ..., \alpha^K$ with $\sum_{k=1}^{K} \alpha^k = 1$ such that*

$$\mathbb{E}_\beta[f(w)|D] = \sum_{k=1}^{K} \alpha^k \mathbb{E}_{\beta^k}[f(w)|D]$$

for any function $f : W \to \mathbb{R}$.

Proof. By definition,

$$\begin{aligned}
\mathbb{E}_\beta[f(w)|D] &= \sum_{w \in D} \beta(w|D) f(w) \\
&= \sum_{w \in D} \frac{\sum_{k=1}^{K} \lambda^k \beta^k(w)}{\sum_{k=1}^{K} \lambda^k \beta^k(D)} f(w) \\
&= \sum_{w \in D} \sum_{k=1}^{K} \frac{\lambda^k \beta^k(D) \beta^k(w|D)}{\sum_{k=1}^{K} \lambda^k \beta^k(D)} f(w).
\end{aligned}$$

Define

$$\alpha^k = \frac{\lambda^k \beta^k(D)}{\sum_{k=1}^{K} \lambda^k \beta^k(D)}.$$

By definition, $\alpha^k \geq 0$ and $\sum_{k=1}^{K} \alpha^k = 1$. Therefore,

$$\begin{aligned}
\mathbb{E}_\beta[f(w)|D_{ijp}] &= \sum_{w \in D} \sum_{k=1}^{K} \alpha^k \beta^k(w|D_{ijp}) f(w) \\
&= \sum_{k=1}^{K} \alpha^k \mathbb{E}_{\beta^k}[f(w)|D_{ijp}].
\end{aligned}$$

This establishes the lemma. ■

The following lemma concerns independent stable beliefs.

Lemma 2 *Suppose that a matching $(\mu, \mathbf{p}, \beta)$ is individually rational and β is independent. Then the matching is not blocked by $(i, j, p) \in I \times J \times \mathbb{R}$ if $\beta(D_{ijp}) > 0$ implies that*

$$\mathbb{E}_\beta [b_{ij} | D_{ijp}] - p \leq \mathbb{E}_\beta \left[b_{\mu(j)j} \right] - p_{\mu(j)j},$$

where

$$D_{ijp} = \left\{ w \in W : a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)} \right\}.$$

Proof. The matching is not blocked by (i, j, p) if

$$\mathbb{E}_\beta [b_{ij} | D_{ijp}] - p \leq \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp} \right] - p_{\mu(j)j} \right\}.$$

Since $D_{ijp} = D_i \times W_{-i}$ for some $D_i \subset W_i$, and β is independent, $\mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp} \right] = \mathbb{E}_\beta \left[b_{\mu(j)j} \right]$. Moreover, individual rationality implies that $\mathbb{E}_\beta \left[b_{\mu(j)j} \right] - p_{\mu(j)j} \geq 0$. The conclusion follows immediately. ■

We proceed to prove the three desired convexity properties.

Proof of Property (i). It suffices to show that the stability of $(\mu, \mathbf{p}^1, \beta)$ and $(\mu, \mathbf{p}^2, \beta)$ implies the stability of $(\mu, \lambda \mathbf{p}^1 + (1 - \lambda) \mathbf{p}^2, \beta)$ for each $\lambda \in (0, 1)$, and the conclusion follows from induction. The individual rationality of $(\mu, \lambda \mathbf{p}^1 + (1 - \lambda) \mathbf{p}^2, \beta)$ is immediate.

Define

$$D_{ijp}(\lambda) = \left\{ w \in W : a_{ij}(w) + p > a_{i\mu(i)}(w) + \lambda p^1_{i\mu(i)} + (1 - \lambda) p^2_{i\mu(i)} \right\}.$$

Then

$$\begin{aligned} D_{ijp}(\lambda) &= \left\{ w \in W : a_{ij}(w) + p - \left(\lambda p^1_{i\mu(i)} + (1 - \lambda) p^2_{i\mu(i)} \right) + p^1_{i\mu(i)} > a_{i\mu(i)}(w) + p^1_{i\mu(i)} \right\} \\ &= \left\{ w \in W : a_{ij}(w) + p - \left(\lambda p^1_{i\mu(i)} + (1 - \lambda) p^2_{i\mu(i)} \right) + p^2_{i\mu(i)} > a_{i\mu(i)}(w) + p^2_{i\mu(i)} \right\}. \end{aligned}$$

By the stability of $(\mu, \mathbf{p}^1, \beta)$ and $(\mu, \mathbf{p}^2, \beta)$, we have

$$\begin{aligned} \mathbb{E}_\beta [b_{ij} | D_{ijp}(\lambda)] - \left[p - \left(\lambda p^1_{i\mu(i)} + (1 - \lambda) p^2_{i\mu(i)} \right) + p^1_{i\mu(i)} \right] &\leq \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp}(\lambda) \right] - p^1_{\mu(j)j} \right\}; \\ \mathbb{E}_\beta [b_{ij} | D_{ijp}(\lambda)] - \left[p - \left(\lambda p^1_{i\mu(i)} + (1 - \lambda) p^2_{i\mu(i)} \right) + p^2_{i\mu(i)} \right] &\leq \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp}(\lambda) \right] - p^2_{\mu(j)j} \right\}. \end{aligned}$$

Multiplying the first inequality by λ and the second inequality by $1 - \lambda$, and then adding them up, we obtain

$$\begin{aligned} \mathbb{E}_\beta [b_{ij} | D_{ijp}(\lambda)] - p &\leq \lambda \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp}(\lambda) \right] - p^1_{\mu(j)j} \right\} \\ &\quad + (1 - \lambda) \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp}(\lambda) \right] - p^2_{\mu(j)j} \right\} \\ &= \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp}(\lambda) \right] - \left(\lambda p^1_{\mu(j)j} + (1 - \lambda) p^2_{\mu(j)j} \right), \end{aligned}$$

where the last equality follows from the independence of β and the individual rationality of stable matching. Thus $(\mu, \lambda \mathbf{p}^1 + (1 - \lambda) \mathbf{p}^2, \beta)$ is not blocked by (i, j, p).

Proof of Property (ii). Since $\lambda\beta^1 + (1-\lambda)\beta^2$ is not necessarily independent even though β^1 and β^2 are independent beliefs, we cannot proceed by induction as in the previous proof. Let $\beta = \sum_{k=1}^{K} \lambda^k \beta^k$. The individual rationality of $(\mu, \mathbf{p}, \beta)$ is immediate. We only need to show that $(\mu, \mathbf{p}, \beta)$ is not blocked by any $(i, j, p) \in I \times J \times \mathbb{R}$. Consider

$$D_{ijp} = \left\{ w \in W : a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)} \right\}.$$

Suppose that $\beta(D_{ijp}) > 0$. We need to show that

$$\mathbb{E}_\beta [b_{ij} | D_{ijp}] - p \leq \max \left\{ 0, \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp} \right] - p_{\mu(j)j} \right\}. \tag{9}$$

To prove the claim, it is without loss of generality to assume that $\beta^k(D_{ijp}) > 0$ for each $k = 1, ..., K$. By Lemma 2, the stability of $\left(\mu, \mathbf{p}, \beta^k\right)$ implies that

$$\mathbb{E}_{\beta^k} [b_{ij} | D_{ijp}] - p \leq \mathbb{E}_{\beta^k} \left[b_{\mu(j)j} | D_{ijp} \right] - p_{\mu(j)j}.$$

Let $(\alpha^1, ..., \alpha^K)$ be the weights obtained in Lemma 1 by setting $D = D_{ijp}$. Multiplying the previous inequality by α^k and then summing over k, we obtain

$$\sum_{k=1}^{K} \alpha^k \mathbb{E}_{\beta^k} [b_{ij} | D_{ijp}] - p \leq \sum_{k=1}^{K} \alpha^k \mathbb{E}_{\beta^k} \left[b_{\mu(j)j} | D_{ijp} \right] - p_{\mu(j)j}.$$

By Lemma 1, the above inequality is equivalent to

$$\mathbb{E}_\beta [b_{ij} | D_{ijp}] - p \leq \mathbb{E}_\beta \left[b_{\mu(j)j} | D_{ijp} \right] - p_{\mu(j)j}.$$

It follows immediately that inequality (9) holds.

Proof of Property (iii). Notice that δ_{w^k}, the probability measure that assigns probability 1 to w^k, is independent: $\delta_{w^k}(w) = \delta_{w_1^k}(w_1) \times \cdots \times \delta_{w_n^k}(w_n)$ for all $w = (w_1, ..., w_n) \in W$. The claim then immediately follows from (i). ■

B Proof of Proposition 2

The "if" part follows from Part (iii) of Proposition 1. To prove the "only if" part, consider a stable matching $(\mu, \mathbf{p}, \beta)$ and fix a type profile w in the support of β.

By the individual rationality of $(\mu, \mathbf{p}, \beta)$, we have

$$a_{i\mu(i)}(w) + p_{i\mu(i)} \geq 0 \text{ for all } i \in I \tag{10}$$

and $\mathbb{E}_\beta \left[b_{\mu(j)j} \right] - p_{\mu(j)j} \geq 0$ for any $j \in J$. By Assumption 1, $b_{\mu(j)j}(w)$ is independent of w, and hence $\mathbb{E}_\beta \left[b_{\mu(j)j} \right] = b_{\mu(j)j}(w)$. Thus,

$$b_{\mu(j)j}(w) - p_{\mu(j)j} \geq 0 \text{ for all } j \in J. \tag{11}$$

Hence, (10) and (11) imply that $(\mu, \mathbf{p})$ is individually rational when there is complete information about w.

Consider any $(i,j,p) \in I \times J \times \mathbb{R}$ such that $a_{ij}(w) + p > a_{i\mu(i)}(w) + p_{i\mu(i)}$. Then $D_{ijp} \neq \emptyset$ and $\beta(D_{ijp}) > 0$. Since $(\mu, \mathbf{p}, \beta)$ is not blocked by (i,j,p), we have

$$\mathbb{E}_\beta[b_{ij}|D_{ijp}] - p \leq \max\left\{0, \mathbb{E}_\beta\left[b_{\mu(j)j}|D_{ijp}\right] - p_{\mu(j)j}\right\}. \tag{12}$$

By Assumption 1, $\mathbb{E}_\beta[b_{ij}|D_{ijp}] = b_{ij}(w)$ and $\mathbb{E}_\beta\left[b_{\mu(j)j}|D_{ijp}\right] = b_{\mu(j)j}(w)$ for any w in the support of β. Inequality (12) can be rewritten as

$$\begin{aligned} b_{ij}(w) - p &\leq \max\left\{0, b_{\mu(j)j}(w) - p_{\mu(j)j}\right\} \\ &= b_{\mu(j)j}(w) - p_{\mu(j)j}, \end{aligned}$$

where the last equality follows from (11). Therefore, (i,j,p) does not block $(\mu, \mathbf{p})$ when there is complete information about w. We have thus proved that $(\mu, \mathbf{p})$ is complete-information stable for any w in the support of β.

It is well-known that a stable matching under complete information maximizes the sum of surpluses. Hence the stable matching $(\mu, \mathbf{p}, \beta)$ is full-information efficient. ■

C Proof of Proposition 3

C.1 Duality of Bayesian Efficiency

Primal. We introduce a vector of non-negative real variables $x = (x_{ij})_{i\in I, j\in J}$. Consider a problem that maximizes the objective

$$V(x) := \sum_{i\in I}\sum_{j\in J} x_{ij}\left(\sum_{w\in W}\beta(w)\, a_{ij}(w) + \sum_{w\in W}\beta(w)\, b_{ij}(w)\right)$$

subject to

$$\begin{aligned} &\textstyle\sum_{j\in J} x_{ij} \leq 1; \\ &\textstyle\sum_{i\in I} x_{ij} \leq 1; \\ &x_{ij} \geq 0,\ i \in I,\ j \in J. \end{aligned}$$

It is well known that this linear programming problem has an optimal solution x^* with all $x^*_{ij} = 0$ or 1. Such (x^*_{ij}) can be equivalently written as a matching function μ: $\mu(i) = j$ if and only if $x^*_{ij} = 1$, and the objective function of the linear program can be viewed as the sum of surpluses weighted by a probability $\beta \in \Delta(W)$. Therefore, Bayesian efficiency of a stable matching is ensured if the corresponding matching function is an optimal solution to the linear programming problem.

Dual. The dual of this linear programming problem is to choose real variables $u = (u_i)_{i\in I}$ and $v = (v_j)_{j\in J}$ to minimize the objective

$$U(u,v) := \sum_{i\in I} u_i + \sum_{j\in J} v_j$$

subject to

$$\begin{aligned} u_i + v_j &\geq \sum_{w \in W} \beta(w)\, a_{ij}(w) + \sum_{w \in W} \beta(w)\, b_{ij}(w),\ i \in I,\ j \in J; \qquad (13) \\ u_i &\geq 0,\ i \in I; \\ v_j &\geq 0,\ j \in J. \end{aligned}$$

Denote the optimal value of the dual by $U_{\min}$ and the optimal value of the primal by $V_{\max}$. By the strong duality theorem, $V_{\max} = U_{\min}$.

If there is complete information, the duality analysis is well known: the dual problem links the stable matching, and the strong duality theorem says that a stable matching is (full-information) efficient. With asymmetric information, the linkage of the dual to stable matching is not immediate because conditional probability measures $\beta(\cdot|\cdot)$ are used to define stability whereas the unconditional probability β appears in the dual problem (that is, a firm needs to update its beliefs before joining a blocking coalition).

C.2 Proof of the Proposition

Consider a stable matching $(\mu, \mathbf{p}, \beta)$. Define $u^* = (u_1^*, ..., u_n^*)$, $v^* = (v_1^*, ..., v_m^*)$, and $x^* = (x_{ij}^*)_{i \in I, j \in J}$ as follows:

$$\begin{aligned} u_i^* &= \sum_{w \in W} \beta(w)\, a_{i\mu(i)}(w) + p_{i\mu(i)}; \\ v_j^* &= \sum_{w \in W} \beta(w)\, b_{\mu(j)j}(w) - p_{\mu(j)j}; \\ x_{ij}^* &= \begin{cases} 1 & \text{if } \mu(i) = j \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

By definition, x^* is feasible for the primal problem; we need to show that x^* is the optimal solution to the primal problem under various conditions. We proceed in two steps.

Step 1. We shall establish the following claim: If (u^*, v^*) is a feasible solution to the dual problem, then x^* is an optimal solution to the primal problem, and the matching $(\mu, \mathbf{p}, \beta)$ is Bayesian efficient.

To prove this claim, note that

$$U(u^*, v^*) \geq U_{\min} = V_{\max} \geq V(x^*),$$

where the first relation follows from the assumption that (u^*, v^*) is a feasible solution for the dual problem, the second relation follows from the strong duality theorem, and the third relation follows because x^* is feasible for the primal problem.

Note also that $V(x^*) = U(u^*, v^*)$ because each of them is the total expected payoff from $(\mu, \mathbf{p}, \beta)$. Therefore,

$$U(u^*, v^*) = U_{\min} = V_{\max} = V(x^*).$$

This proves that x^* is an optimal solution to the primal problem. It follows immediately from the definition of the primal problem and the definition of x^* that $(\mu, \mathbf{p}, \beta)$ is Bayesian efficient.

Step 2. We shall show that (u^*, v^*) is a feasible solution to the dual problem, if one of the three conditions is satisfied.

By definition, (u^*, v^*) is non-negative. It remains to show that (u^*, v^*) satisfies the constraint (13) in the dual problem.

We claim that for any w in the support of β, and any $i \in I$ and $j \in J$,

$$\begin{aligned} & a_{i\mu(i)}(w) + p_{i\mu(i)} + \sum\nolimits_{w \in W} \beta(w)\, b_{\mu(j)j}(w) - p_{\mu(j)j} \\ \geq\ & a_{ij}(w) + \sum\nolimits_{w \in W} \beta(w)\, b_{ij}(w). \end{aligned} \tag{14}$$

The claim is trivially true if $\mu(i) = j$. To prove this claim, suppose by way of contradiction that (14) does not hold for some $\overline{w}$ in the support of β and some pair $(i, j) \in I \times J$, $\mu(i) \neq j$. Then, there exists $p \in \mathbb{R}$ such that

$$a_{i\mu(i)}(\overline{w}) + p_{i\mu(i)} < a_{ij}(\overline{w}) + p \tag{15}$$

and

$$\sum\nolimits_{w \in W} \beta(w)\, b_{\mu(j)j}(w) - p_{\mu(j)j} < \sum\nolimits_{w \in W} \beta(w)\, b_{ij}(w) - p. \tag{16}$$

Inequality (15) captures worker i's incentive to form a blocking coalition with firm j. Consider the set

$$D_{ijp} = \left\{ w \in W : a_{i\mu(i)}(w) + p_{i\mu(i)} < a_{ij}(w) + p \right\}.$$

By assumption, D_{ijp} is a non-empty set that contains $\overline{w}$.

Under condition (i)—Assumption 2, $a_{i\mu(i)}(w)$ and $a_{ij}(w)$ are independent of w. Therefore, $D_{ijp} = \left\{ w \in W : p_{i\mu(i)} < p \right\}$ if $\mu(i) \in J$, and $D_{ijp} = \left\{ w \in W : p_{i\mu(i)} < h(i,j) + p \right\}$ if $\mu(i) = i$, where $h(i,j) = a_{ij}(w)$ and $a_{ii}(w) = 0$. In either case, since $\overline{w} \in D_{ijp}$, $D_{ijp} = W$.

Under condition (ii), $\mu(i) \neq i$, and by Assumption 3, $a_{i\mu(i)}(w) = a_{ij}(w) = g(i,w) + h(i,j)$. Hence,

$$D_{ijp} = \left\{ w \in W : h(i, \mu(i)) + p_{i\mu(i)} < h(i,j) + p \right\}.$$

Again, since $\overline{w} \in D_{ijp}$, $D_{ijp} = W$.

Under both conditions (i) and (ii), $\beta(D_{ijp}) = 1$. If we replace $\beta(w)$ by $\beta(w|D_{ijp})$ in (16), the inequality is unchanged. Therefore, (16) implies that firm j is willing to deviate with worker i. That is, (i, j, p) blocks $(\mu, \mathbf{p}, \beta)$, a contradiction.

Suppose that condition (iii) holds, and $\mu(i) = i$ (the case of $\mu(i) \neq i$ is covered by the proof under condition (ii) already). Then

$$D_{ijp} = \left\{ w \in W : p_{i\mu(i)} < a_{ij}(w) + p \right\}.$$

Since $a_{ij}(\cdot)$ and $b_{ij}(\cdot)$ are co-monotonic, there exists some linear order on W_i that is specific to the pair (i,j), such that both $a_{ij}(w_i)$ and $b_{ij}(w_i)$ is non-decreasing in w_i (note that since a_{ij} and b_{ij} depends only on w_i, the linear order naturally extends to an order on W). Therefore, D_{ijp} contains all w's such that w_i is larger than a cutoff according to the linear order. It follows from the monotonicity of $b_{ij}(w)$ in w_i that

$$\sum_{w\in W}\beta(w)\,b_{ij}(w)-p\leq\sum_{w\in W}\beta(w|D_{ijp})\,b_{ij}(w)-p.$$

Since $\mu(i)\neq j$, it follows from the independence of β that

$$\sum_{w\in W}\beta(w)\,b_{\mu(j)j}(w)-p_{\mu(j)j}=\sum_{w\in W}\beta(w|D_{ijp})\,b_{\mu(j)j}(w)-p_{\mu(j)j}.$$

The above two inequalities together with (16) imply that

$$\sum_{w\in W}\beta(w|D_{ijp})\,b_{\mu(j)j}(w)-p_{\mu(j)j}<\sum_{w\in W}\beta(w|D_{ijp})\,b_{ij}(w)-p.$$

That is, firm j is willing to deviate with worker i. Thus, (i,j,p) blocks $(\mu,\mathbf{p},\beta)$, a contradiction. This establishes the claim that (14) holds.

Multiplying both sides of (14) by $\beta(w)$ and summing it up over w, we obtain

$$u_i^*+v_j^*\geq\sum_{w\in W}\beta(w)\,a_{ij}(w)+\sum_{w\in W}\beta(w)\,b_{ij}(w).$$

That is, (u^*,v^*) satisfies (13). Therefore, (u^*,v^*) is feasible for the dual problem. ■

References

[1] Aumann, Robert J. (1998). "Common priors: A reply to Gul." *Econometrica* 66(4): 929–938.

[2] Becker, Gary S. (1973). "A theory of marriage: Part I." *Journal of Political Economy* 81(4): 813–846.

[3] Bloch, Francis, and Bhaskar Dutta (2009). "Correlated equilibria, incomplete information and coalitional deviations." *Games and Economic Behavior* 66(2): 721–728.

[4] Browning, Martin, Pierre-André Chiappori, and Yoram Weiss (2014). *Economics of the Family.* Cambridge University Press.

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

[6] Chakraborty, Archishman, Alessandro Citanna, and Michael Ostrovsky (2015). "Group stability in matching with interdependent values." *Review of Economic Design* 19(1): 3–24.

[7] Chiappori, Pierre-André (2017). *Matching with Transfers: The Economics of Love and Marriage.* Princeton University Press.

[8] Chiappori, Pierre-André, and Bernard Salanié (2016). "The econometrics of matching models." *Journal of Economic Literature* 54(3): 832–861.

[9] Crawford, Vincent P. (1985). "Efficient and durable decision rules: A reformulation." *Econometrica* 53(4): 817–835.

[10] Crawford, Vincent P., and Elsie Marie Knoer (1981). "Job matching with heterogeneous firms and workers." *Econometrica* 49(2): 437–450.

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

[12] Dutta, Bhaskar, and Rajiv Vohra (2017). "Rational expectations and farsighted stability." *Theoretical Economics* 12(3), 1191–1227.

[13] Forges, Françoise, and Roberto Serrano (2013). "Cooperative games with incomplete information: some open problems." *International Game Theory Review* 15(2): 1–17.

[14] Fudenberg, Drew, and Jean Tirole (1991). "Perfect Bayesian equilibrium and sequential equilibrium." *Journal of Economic Theory* 53(2): 236–260.

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

[16] Grossman, Sanford J., and Motty Perry (1986). "Perfect sequential equilibrium." *Journal of Economic Theory* 39(1): 97–119.

[17] Gul, Faruk (1989). "Bargaining foundations of Shapley value." *Econometrica* 57(1): 81–95.

[18] Gul, Faruk (1998). "A comment on Aumann's Bayesian view." *Econometrica* 66(4): 923–927.

[19] Holmström, Bengt, and Roger Myerson (1993). "Efficient and durable decision rules with incomplete information." *Econometrica* 51(6): 1799–1819.

[20] Kahn, Charles M., and Dilip Mookherjee (1995). "Coalition proof equilibrium in an adverse selection insurance economy." *Journal of Economic Theory* 66(1): 113–138.

[21] Kreps, David M., and Robert Wilson (1982). "Sequential equilibria." *Econometrica* 50(4): 863–894.

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

[23] Myerson, Roger B. (2007). "Virtual utility and the core for games with incomplete information." *Journal of Economic Theory* 136(1): 260–285.

[24] Perry, Motty, and Philip J. Reny (1994). "A noncooperative view of coalition formation and the core." *Econometrica* 62(4): 795–817.

[25] Roth, Alvin E. (1986). "On the allocation of residents to rural hospitals: a general property of two-sided matching markets." *Econometrica* 54(2): 425–427.

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

[27] Rothschild, Michael and Joseph Stiglitzand Stiglitz (1976). "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information." *The Quarterly Journal of Economics*, 90(4): 629–649.

[28] Salanié, Bernard (2015). "Identification in Separable Matching with Observed Transfers." Working Paper, Columbia University.

[29] Serrano, Roberto, and Rajiv Vohra (2007). "Information transmission in coalitional voting games." *Journal of Economic Theory* 134(1): 117–137.

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

[31] Wilson, Charles (1977). "A model of insurance markets with incomplete information." *Journal of Economic Theory*, 16(2): 167–207.

[32] Wilson, Robert (1978). "Information, efficiency, and the core of an economy." *Econometrica* 46(4): 807–816.