Stability and Efficiency of Two-Sided Matching Markets

Preliminary Draft for Seminar at Collegio Carlo Alberto

Qingmin Liu*

May 9, 2022

Abstract

We study the stability of two-sided markets with incomplete information and propose a program for formulating cooperative concepts that separate belief formation and coalition formation. Belief- based refinements are invoked to show that stability has significant restrictions.

*Department of Economics, Columbia University. E-mail: ql2177@columbia.edu.

Contents

1 Matching Games with Incomplete Information

We build on the complete-information matching games formulated by Gale and Shapley (1962), Shapley and Shubik (1971) and Crawford and Knoer (1981). The economic agents are referred to as *workers* and *firms*, but the model of two-sided markets is obviously applicable more generally.

Let I be a finite set of workers, and J be a finite set of firms. Let T_i and T_j be finite sets of types for worker $i \in I$ and $j \in J$ respectively. We also use $n \in I \cup J$ to denote either a worker or a firm. Let $T = \prod_{n \in I \cup J} T_n$ be the set of type profiles for all workers and firms, with a typical element t. We shall assume that there is a common prior $\beta^0 \in \Delta(T)$ and, for simplicity, that β^0 has a full support.[1] To simply notiation, we shall write $t_{ij} = (t_i, t_j)$ as the profile of types of the pair of worker i and firm j, t_{-ij} as the profile of types of players outside of the pair, and T_{ij} and T_{-ij} as the corresponding set of type profiles. To account for unmatched players, we equate t_{ii} and t_i, and t_{jj} and t_j.

When the type profile is $t \in T$, let $a_{ij}(t), b_{ij}(t) \in \mathbb{R}$ be the **matching values** worker i and firm j receive in a matched pair (i, j), respectively, and let $a_{ii}(t), b_{jj}(t) \in \mathbb{R}$ be the players' payoff from staying single. For generality, we allow the possibility that matching values depend on players' observable attributes summarized by their index i and j, and we also allow the possibility that matching values of a matched pair depend on the entire profile of player types, including players outside of the pair.

A **matching game** $\left(a, b, \beta^0\right)$ with incomplete information is summarized by the matching value function $(a, b) : I \times J \times T \to \mathbb{R}^2$ and the common prior $\beta^0 \in \Delta(T)$.

A **match** is a one-to-one function $\mu : I \cup J \to I \cup J$ that pairs up workers and firms such that the following holds for each $i \in I$ and $j \in J$: (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 $\mu(i) = i \in I$ means that worker i is unmatched; likewise for $\mu(j) = j \in J$.

Let $\mathbb{P} \subset \mathbb{R}$ be the set of **permissible transfers** and denote by $p_{ij} \in \mathbb{P}$ the **transfer** that worker i receives from firm j. We assume $0 \in \mathbb{P}$. If $\mathbb{P} = \{0\}$, the matching game has **non-transferrable utility**. If $\mathbb{P} = \mathbb{R}$, the matching game has **perfectly transferrable utility**. A **transfer scheme** associated with a match μ is a vector $\mathbf{p}$ that specifies a transfer $p_{i\mu(i)} \in \mathbb{P}$ for each $i \in I$ and $p_{\mu(j)j} \in \mathbb{P}$ for each $j \in J$. Without loss of generality, we require $p_{ii} = p_{jj} = 0$. If worker i and firm j are matched together with a transfer p_{ij} when the profile of workers' types is t, worker i's and firm j's ex post payoffs are $a_{ij}(t) + p_{ij}$ and $b_{ij}(t) - p_{ij}$, respectively.

We shall refer to a match together with a transfer scheme $(\mu, \mathbf{p})$ as a **matching outcome**. We shall assume that a matching outcome is publicly observable.

[1] The extension to type spaces without common priors or with heterogenous priors without full support are straightforward; see Liu (2017) for a formulation.

2 Some Classes of Matching Games

Several classes of matching games are of general interests. Let $\pi_I = \{(i,j) : i \in I, j \in J \cup \{i\}\}$ be the set of pairs that involve a worker (including unmatched workers), and let $\pi_J = \{(i,j) : j \in J, i \in I \cup \{j\}\}$ be the set of pairs that involve a firm (including unmatched firms).

2.1 Private Values

Definition 1. *A matching game has* ***private values*** *if for any $t \in T$ we have*

$$\begin{aligned} a_{ij}(t) &= A_{ij}(t_i) + A_i(t) \text{ for all } (i,j) \in \pi_I, \\ b_{ij}(t) &= B_{ij}(t_j) + B_j(t) \text{ for all } (i,j) \in \pi_J, \end{aligned}$$

where $A_{ij} : T_i \to \mathbb{R}$, $B_{ij} : T_j \to \mathbb{R}$, and $A_i, B_j : T \to \mathbb{R}$ are a class of real-valued functions.

In a private-value matching game a player's matching value can depend on the observable attribute of his partner, as well as the types of all other players not in the pair (i,j). Although the matching value a_{ij} depends on t_j through $A_i(t)$, this dependence on t_j behaves more like a private value because $a_{ij'}$, $j' \neq j$, depends on t_j through $A_i(t)$ in the same way. In the special case where the matching values of a pair (i,j) depend only on t_{ij}, the functional forms in the definition of private values simplify to

$$a_{ij}(t) = A_{ij}(t_i) \text{ and } b_{ij}(t) = B_{ij}(t_j),$$

and the terminology of "private value" is justified.[2]

2.2 Comonotonic Differences

Two real-valued functions $f, g : X_1 \times X_2 \to \mathbb{R}$ are **comonotonic** on X_1 if $(f(x_1,x_2) - f(x_1',x_2))\,(g\,(x_1,x_2) -$
0 for any $x_1, x_1' \in X_1$ and $x_2 \in X_2$.

Definition 2. *A matching game has* ***comonotonic differences*** *if $a_{ij} - a_{ij'}$ and $b_{ij} - b_{i'j}$ are comonotonic on T_i and on T_j for any two pairs $(i,j) \in I \times J$ and $(i',j') \in (I \cup \{j\}) \times (J \cup \{i\})$.*

Comonotonicity on T_i and on T_j separately is weaker than comonotonicity on $T_i \times T_j$. Although the property of comonotonic differences clearly places no restriction on complete information matching games, it is central for incomplete information problems. For any

[2]In this case, $a_{ij}(t_{ij}, t_{-ij}) = a_{ij}(t_{ij}, t'_{-ij})$ and hence $A_{ij}(t_i) + A_i(t_{ij}, t_{-ij}) = A_{ij}(t_i) + A_i(t_{ij}, t'_{-ij})$. Therefore, $A_i(t_{ij}, t_{-ij}) = A_i(t_{ij}, t'_{-ij})$. Thus A_i is independent of t_{-ij}. Similar arguments apply to a pair (i,j') and hence A_i is independent of $t_{-ij'}$. Therefore, A_i depends only on t_i, which is a special case of A_{ij}.

putative matching, consider a potential blocking pair i and j whose partners are $j' \neq j$ and $i' \neq i$ respectively. Worker i's gain from the deviation is $a_{ij} - a_{ij'}$ and firm j's gain from the deviation is $b_{ij} - b_{i'j}$. If the game has comonotonic differences, then the incentives for i and j to rematch with each other are aligned.

Some special cases of comonotonic differences are of interests in their own rights.

One-sided Interdependence. A matching game has *one-sided interdependence* if for any $t \in T$, we have either

$$a_{ij}(t) = A_i(t) + A_{ij} \text{ for all } (i,j) \in \pi_I$$

with no restriction placed on b, or

$$b_{ij}(t) = B_j(t) + B_{ij} \text{ for all } (i,j) \in \pi_J$$

with no restriction placed on a, where $A_i, B_j : T \to \mathbb{R}$ are real-valued functions, and A_{ij} and B_{ij} are constants.

One-sided interdependence captures, e.g., applications where workers' cost of production is a function of their own types (but firms' outputs depend on both workers' and firms' private information), or customers' (in J) valuations are their own private information while providers' (in I) actual costs of serving their clients depend both their private information and buyers' valuations.

To verify comonotonic differences, consider the first case where firms' matching values are arbitrary. We have that

$$a_{ij}(t) - a_{ij'}(t) = A_{ij} - A_{ij'}$$

does not depend on t_i and t_j. Therefore, $a_{ij} - a_{ij'}$ and $b_{ij} - b_{i'j}$ are comonotonic on T_i and T_j.

Separable Values. A matching game has *separable values* if for any $t \in T$ we have

$$a_{ij}(t) = A_{ij}(t_{-i}) + A_i(t) \text{ for all } (i,j) \in \pi_I,$$
$$b_{ij}(t) = B_{ij}(t_{-j}) + B_j(t) \text{ for all } (i,j) \in \pi_J,$$

where $A_i, B_j : T \to \mathbb{R}$, $A_{ij} : T_{-i} \to \mathbb{R}$ and $B_{ij} : T_{-j} \to \mathbb{R}$ are a class of real-valued functions.

Separable-value games appears similiar to private-value games in their functional forms, but they are qualitatively different. If matching values of a matched pair (i,j) depend only on t_{ij}, then separable values imply

$$a_{ij}(t) = A_{ij}(t_j) + A_i(t_i) \text{ and } b_{ij}(t) = B_{ij}(t_i) + B_j(t_j)$$

for a class of real-valued functions $A_{ij}, B_j : T_j \to \mathbb{R}$ and $B_{ij}, A_i : T_i \to \mathbb{R}$. If, in addition, players' payoffs do not depend on their observable attributes (i and j), then

$$a_{ij}(t) = A(t_i) + A'(t_j) \text{ and } b_{ij}(t) = B(t_i) + B'(t_j)$$

for a class of real-valued functions $A, B : T_i \to \mathbb{R}$ and $A', B' : T_j \to \mathbb{R}$. So the essence of separable values is that there is no interaction of a player's own type with the matching partner's observable attribute (i.e., the absence of the interaction between t_i and j and the interaction between t_j and i).

To see a separable-value matching game has comonotonic differences, observe that

$$a_{ij}(t) - a_{ij'}(t) = A_{ij}(t_{-i}) - A_{ij'}(t_{-i}),$$

which is independent of t_i, and

$$b_{ij}(t) - b_{i'j}(t) = B_{ij}(t_{-j}) - B_{i'j}(t_{-j}),$$

which is independent of t_j. Therefore, $a_{ij} - a_{ij'}$ and $b_{ij} - b_{i'j}$ are comonotonic on T_i and T_j.

Common Values. Consider a two-player co-ordination game with incomplete information: $I = \{i\}$ and $J = \{j\}$. Also $a_{ij} = b_{ij}$ and $a_{ii} = b_{ii} \equiv 0$. Notice that $a_{ij} - a_{ii} = a_{ij}$ and $b_{ij} - b_{jj} = b_{ij}$ are identical. Hence the game has comonotonic differences (for this conclusion it sufficies that a_{ij} and b_{ij} are comonotonic).

Violation of Comonotonic Differences. Consider a lemon's problem with two players. The buyer's value is $b_{ij}(t_i, t_j) = t_j$ and the seller's reservation value (or production cost) is t_j so $a_{ij}(t_i, t_j) = -t_j$. The no-trade value is 0, $a_{ii} \equiv b_{ii} \equiv 0$. The game does not have comonotonic differences.

3 Stability

3.1 Matching-Belief Configuration

For every type profile $t \in T$, some matching outcome $(\mu, \mathbf{p})$ materializes. The relationship between the underlying uncertainties and the observable outcomes is described by a function $M : t \mapsto (\mu, \mathbf{p})$. We shall call the function M a **matching function** or simply a **matching** for the matching game with incomplete information.

The mapping $M : t \mapsto (\mu, \mathbf{p})$ appears to be deterministic, but this is a matter of interpretation. A non-deterministic matching function can be written as $M : (t, s) \mapsto (\mu, \mathbf{p})$ where s is a profile of private/public signals possibly correlated with t, but in this case we are simply enlarging the type space.[3]

Associated with each matching outcome $(\mu, \mathbf{p}) \in M(T)$, player $n \in I \cup J$ of type t_n has an **on-path belief** $\beta_n(\mu, \mathbf{p}, t_n) \in \Delta(T)$, and associated with each pairwise deviation (i, j, p) from $(\mu, \mathbf{p})$, where $\mu(i) \neq j$ and $p \in \mathbb{R}$, player $n \in \{i, j\}$ of type t_n has an **off-path belief** $\beta_n(\mu, \mathbf{p}, i, j, p, t_n) \in \Delta(T)$. We call $(\mu, \mathbf{p}, i, j, p)$ a pairwise deviation of M at t if $M(t) = (\mu, \mathbf{p})$.

Knowing M and observing $(\mu, \mathbf{p})$, players can infer that the set of type profiles is

$$M^{-1}(\mu, \mathbf{p}) = \{t \in T : M(t) = (\mu, \mathbf{p})\}.$$

Naturally, we shall require that

$$\beta_n(\mu, \mathbf{p}, t_n)(t_n) = \beta_n(\mu, \mathbf{p}, i, j, p, t_n)(t_n) = 1,$$

i.e, player n knows his own type, and

$$\beta_n(\mu, \mathbf{p}, t_n)(M^{-1}(\mu, \mathbf{p})) = \beta_n(\mu, \mathbf{p}, i, j, p, t_n)(M^{-1}(\mu, \mathbf{p})) = 1,$$

i.e., player n's belief does not contradict his knowledge of M.

We do not specify the process that leads to these beliefs, which must require additional assumptions on dynamic interactions; the key observation is that a Bayesian player should have a belief for each on-path and off-path scenario, regardless of the process leading to them. Let $\beta = (\beta_n)_{n \in I \cup J}$ denote a **system of beliefs** and call (M, β) a **matching-belief configuration**.

3.2 Stable Configuration

A configuration (M, β) is **individually rational** at $t \in T$ if, for $(\mu, \mathbf{p}) = M(t)$ and all $i \in I$ and $j \in J$,

$$\mathbf{E}_{\beta_i(\mu,\mathbf{p},t_i)}\left(a_{i\mu(i)}\right) + p_{i\mu(i)} \geq \mathbf{E}_{\beta_i(\mu,\mathbf{p},t_i)}(a_{ii}) \text{ and } \mathbf{E}_{\beta_j(\mu,\mathbf{p},t_j)}\left(b_{\mu(j)j}\right) - p_{\mu(j)j} \geq \mathbf{E}_{\beta_j(\mu,\mathbf{p},t_j)}(b_{jj}).$$

[3] See further discussion of this ideas in Liu (2010, 2015).

A configuration (M, β) is **blocked** at $t \in T$ if there exists a pairwise deviation $(\mu, \mathbf{p}, i, j, p)$ at t such that

$$\begin{aligned} \mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{ij}\right) + p_{ij} &> \mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{i\mu(i)}\right) + p_{i\mu(i)} \\ \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{ij}\right) - p_{ij} &> \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{\mu(j)j}\right) - p_{\mu(j)j} \end{aligned}$$

Equivalently, for each pariwise deviation $(\mu, \mathbf{p}, i, j, p)$ at t, define

$$\begin{aligned} D_i &:= \left\{t_i : \mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{ij}\right) + p > \mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{i\mu(i)}\right) + p_{i\mu(i)}\right\}; \\ D_j &:= \left\{t_j : \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{ij}\right) - p > \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{\mu(j)j}\right) - p_{\mu(j)j}\right\}. \end{aligned} \tag{3.1}$$

Thus D_i and D_j are the set of worker i's types and firm j's types that find the deviation $(\mu, \mathbf{p}, i, j, p)$ profitable.[4] We shall call (D_i, D_j) **blocking sets** of (M, β) with respect to $(\mu, \mathbf{p}, i, j, p)$. A configuration (M, β) is **blocked** by $(\mu, \mathbf{p}, i, j, p)$ if and only if the corresponding blocking sets (D_i, D_j) are non-empty.

Definition 3. *A matching-belief configuration (M, β) is **stable** if it is individually rational and is not blocked at any $t \in T$. If (M, β) is a stable configuration, we say M is a **stable matching** and β is a **stable belief**.*

When T is a singleton, the definition of stability reduces to the familiar notion of complete information. Without any restrictions on beliefs, the concept is restrictive only for very specail games.

Theorem 1. *Suppose the matching game has private values. Then $M(t)$ is a complete-information stable matching for any stable configuration (M, β).*

4 Belief-Based Refinements

4.1 Bayes' Rule with Matching Functions

For each $n \in I \cup J$, we write $M_n^{-1}(\mu, \mathbf{p})$ as the set of player n types that are consistent with observing $(\mu, \mathbf{p})$,

$$M_n^{-1}(\mu, \mathbf{p}) = \{t_n \in T_n : M(t_n, t_{-n}) = (\mu, \mathbf{p}) \text{ for some } t_{-n} \in T_{-n}\}.$$

[4] Since $\beta_i(\mu, \mathbf{p}, i, j, p, t_i)$ and $\beta_j(\mu, \mathbf{p}, i, j, p, t_j)$ are defined only for $t \in M^{-1}(\mu, \mathbf{p})$, we have $D_i \subset M_i^{-1}(\mu, \mathbf{p})$ and $D_j \subset M_j^{-1}(\mu, \mathbf{p})$.

Each player $n \in I \cup J$ in addition observes his private type t_n and Bayes' rule require that his belief on $t = (t_n, t_{-n}) \in M^{-1}(\mu, \mathbf{p})$ is

$$\beta^0(t|\mu, \mathbf{p}, t_n) = \frac{\beta^0(t)}{\beta^0(M^{-1}(\mu, \mathbf{p}) \cap (\{t_n\} \times T_{-n}))}. \tag{4.1}$$

We say (M, β) is **on-path consistent** if

$$\beta_n(\mu, \mathbf{p}, t_n) = \beta^0(\cdot|\mu, \mathbf{p}, t_n). \tag{4.2}$$

If, in addition, player n knows that some other player m's types is in a non-empty subset of types $D_m \subset T_m$, the posterior belief of player n is

$$\beta^0(t|\mu, \mathbf{p}, t_n, D_m) = \frac{\beta^0(t)}{\beta^0(M^{-1}(\mu, \mathbf{p}) \cap (\{t_n\} \times D_m \times T_{-nm}))} \tag{4.3}$$

for any $t = (t_n, t_m, t_{-nm}) \in (\{t_n\} \times D_m \times T_{-nm}) \cap M^{-1}(\mu, \mathbf{p})$. Subsets of types that of interest are the set of types that benefit from the deviation, as defined in (3.1).

4.2 Surplus Maximization

From a planner's perspective, knowing the matching game (a, b, β^0) and the matching function M, and observing $(\mu, \mathbf{p})$, her posterior will be

$$\beta^0(t|\mu, \mathbf{p}) = \frac{\beta^0(t)}{\beta^0(M^{-1}(\mu, \mathbf{p}))}$$

for all $t \in M^{-1}(\mu, \mathbf{p})$. The expected surplus generated from this matching outcome according to this posterior is

$$\sum_{i \in I, j \in J} \mathbf{E}\left(a_{i\mu(i)} + b_{\mu(j)j}|\mu, \mathbf{p}\right)$$

We can ask the following question: can the planner rearrange the matching to improve the expected surplus? That is to say, whether μ is the solution of the following surplus maximization problem:

$$\max_{\mu'} \sum_{i \in I, j \in J} \mathbf{E}\left(a_{i\mu'(i)} + b_{\mu'(j)j}|\mu, \mathbf{p}\right) \tag{4.4}$$

If the answer is in the affirmative for *all* $(\mu, \mathbf{p}) \in M(T)$, we say the matching M is **Bayesian efficient**.

To compute the surplus, an outside observer need to know M and the game (a, b, β^0). This is unrealistic. We instead pursue theorems of the following kind: efficiency is obtained a large class of stable matchings for a large class of games. We identify the class of stable matchings by refinements of off-path beiefs, and identify the class of games by structrual properties of payoffs. So the planner needs to know neither the stable matching nor the exact games.

The surplus maximization problem (4.4) has a dual minimization problem[5]:

$$\min_{(u_i)_{i\in I},(v_j)_{j\in J}} \sum\nolimits_{i\in I} u_i + \sum\nolimits_{j\in J} v_j$$

such that, for any $i \in I$ and $j \in J$,

$$\begin{aligned} u_i + v_j &\geq \mathbf{E}\left(a_{ij} + b_{ij}|\mu, \mathbf{p}\right); \\ u_i &\geq \mathbf{E}\left(a_{ii}|\mu, \mathbf{p}\right); \\ v_j &\geq \mathbf{E}\left(b_{jj}|\mu, \mathbf{p}\right). \end{aligned}$$

Lemma 1. *A matching M is Bayesian efficient if for all $(\mu, \mathbf{p}) \in M(T)$, $i \in I$ and $j \in J$, we have*

$$\mathbf{E}\left(a_{i\mu(i)}|\mu, \mathbf{p}\right) + \mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}\right) \geq \mathbf{E}\left(a_{ij} + b_{ij}|\mu, \mathbf{p}\right); \tag{4.5}$$

$$\mathbf{E}\left(a_{i\mu(i)}|\mu, \mathbf{p}\right) \geq \mathbf{E}\left(a_{ii}|\mu, \mathbf{p}\right); \tag{4.6}$$

$$\mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}\right) \geq \mathbf{E}\left(b_{jj}|\mu, \mathbf{p}\right). \tag{4.7}$$

This is the implication of the theorem of duality. If the conditions in Lemma 1 are satisfied, then $\left(\left(\mathbf{E}\left(a_{i\mu(i)}|\mu, \mathbf{p}\right)\right)_{i\in I}, \left(\mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}\right)\right)_{j\in J}\right)$ is a feasible solution for the dual program and $\sum_{i\in I, j\in J} \mathbf{E}\left(a_{i\mu(i)} + b_{\mu(j)j}|\mu, \mathbf{p}\right)$ is an upper bound for the primal program. Therefore, μ is a solution to the primal.

4.3 Refinement 1: Weak Consistency

Motivation. For each pairwise deviation $(\mu, \mathbf{p}, i, j, p)$ of (M, β), players i and j gain from this deviation if and only if their types are in the blocking sets D_i and D_j, respectively (see (3.1)):

$$\begin{aligned} D_i &:= \left\{t_i : \mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{ij}\right) + p > \mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{i\mu(i)}\right) + p_{i\mu(i)}\right\}; \\ D_j &:= \left\{t_j : \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{ij}\right) - p > \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{\mu(j)j}\right) - p_{\mu(j)j}\right\}. \end{aligned}$$

[5]The primal is the maximization of $\sum_{i\in I}\sum_{j\in J} x_{ij}\mathbf{E}\left(a_{ij} + b_{ij}|\mu, \mathbf{p}\right) + \sum_{i\in I} x_{ii}\mathbf{E}\left(a_{ii}|\mu, \mathbf{p}\right) + \sum_{j\in J} x_{jj}\mathbf{E}\left(b_{jj}|\mu, \mathbf{p}\right)$ over non-negative real vectors $(x_{ij}, x_{ii}, x_{jj})_{i\in I, j\in J}$ subject to $\sum_{j\in J\cup\{i\}} x_{ij} \leq 1$ and $\sum_{i\in I\cup\{j\}} x_{ij} \leq 1$.

If the two players form their beliefs conditional on each other's gain from the deviation, their belief should satisfy

$$\begin{array}{l} \beta_i\left(\mu,\mathbf{p},i,j,p,t_i\right)=\beta^0(\cdot|\mu,\mathbf{p},t_i,D_j) \\ \beta_j\left(\mu,\mathbf{p},i,j,p,t_j\right)=\beta^0(\cdot|\mu,\mathbf{p},t_j,D_i) \end{array}. \tag{4.8}$$

When D_i or D_j is empty, Bayes' rule has no restriction. Intuitively, when i is called to deviate together to j, he needs to assume that j gains from the deviation, i.e., j's type is in D_j, to make his decision (this reasoning is similiar to the familiar one in common value auctions or pival voting). This leads to the following definition.

Definition 4. *The configuration (M,β) is **weakly off-path consistent** if (4.8) is satisfied for each pairwise deviation $(\mu,\mathbf{p},i,j,p)$ at $t\in T$ and its corresponding blocking sets (D_i,D_j) defined in (3.1). The configuration (M,β) is **weakly consistent** if it is on-path consistent and weakly off-path consistent.*

Weakly consistent stable configuration impose strong restrictions on a class of matching games.

Theorem 2. *All weakly consistent stable configurations of a matching game with one-sided interdependence are Bayesian efficient.*

Proof. Individual rationality and on-path consistency of (M,β) imply that (4.6) and (4.7) are satisfied. Suppose to the contrary that (M,β) is inefficient, then by Lemma 1,

$$\mathbf{E}\left(a_{ij}+b_{ij}|\mu,\mathbf{p}\right)>\mathbf{E}\left(a_{i\mu(i)}|\mu,\mathbf{p}\right)+\mathbf{E}\left(b_{\mu(j)j}|\mu,\mathbf{p}\right)$$

and hence there exist $p\in\mathbb{R}$ and $t^*\in M^{-1}\left(\mu,\mathbf{p}\right)$ such that

$$\mathbf{E}\left(a_{ij}|\mu,\mathbf{p},t_i^*\right)+p \;>\; \mathbf{E}\left(a_{i\mu(i)}|\mu,\mathbf{p},t_i^*\right)+p_{i\mu(i)} \tag{4.9}$$

$$\mathbf{E}\left(b_{ij}|\mu,\mathbf{p},t_j^*\right)-p \;>\; \mathbf{E}\left(b_{\mu(j)j}|\mu,\mathbf{p},t_j^*\right)-p_{\mu(j)j} \tag{4.10}$$

By one-sided interdependence, (4.9) takes the form of

$$\mathbf{E}\left(A_i(t)+A_{ij}|\mu,\mathbf{p},t_i\right)+p>\mathbf{E}\left(A_i(t)+A_{i\mu(i)}|\mu,\mathbf{p},t_i\right)+p_{i\mu(i)}$$

and hence, $A_{ij}+p>A_{i\mu(i)}+p_{i\mu(i)}$. Therefore,

$$\begin{aligned} D_i \quad : \quad &= \left\{t_i\in M_i^{-1}(\mu,\mathbf{p}):\mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{ij}\right)+p>\mathbf{E}_{\beta_i(\mu,\mathbf{p},i,j,p,t_i)}\left(a_{i\mu(i)}\right)+p_{i\mu(i)}\right\} \\ &= \left\{t_i\in M_i^{-1}(\mu,\mathbf{p}):A_{ij}+p>A_{i\mu(i)}+p_{i\mu(i)}\right\} \\ &= M_i^{-1}(\mu,\mathbf{p}) \end{aligned}$$

Weak off-path consistency requires that

$$\beta_j\left(\mu, \mathbf{p}, i, j, p, t_j\right) = \beta^0(\cdot|\mu, \mathbf{p}, t_j, D_i) = \beta^0(\cdot|\mu, \mathbf{p}, t_j)$$

Therefore,

$$\begin{aligned} D_j \quad : \quad &= \left\{t_j \in M_j^{-1}(\mu, \mathbf{p}) : \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{ij}\right) - p > \mathbf{E}_{\beta_j(\mu,\mathbf{p},i,j,p,t_j)}\left(b_{\mu(j)j}\right) - p_{\mu(j)j}\right\} \\ &= \left\{t_j \in M_j^{-1}(\mu, \mathbf{p}) : \mathbf{E}\left(b_{ij}|\mu, \mathbf{p}, t_j\right) - p > \mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}, t_j\right) - p_{\mu(j)j}\right\} \end{aligned}$$

Now $D_j \neq \emptyset$ because (4.10). Therefore, (D_i, D_j) are non-empty blocking sets for (M, β), contradicting the assumption of stability. □

4.4 Refinement 2: Strong Consistency

Motivation. Weak off-path consistency computes blocking sets (D_i, D_j) given beliefs $\beta_i(\mu, \mathbf{p}, i, j, p, t_i)$ and $\beta_j(\mu, \mathbf{p}, i, j, p, t_j)$, and then requires that $\beta_i(\mu, \mathbf{p}, i, j, p, t_i)$ be $\beta_i\left(\cdot|\mu, \mathbf{p}, t_i, D_j\right)$ and that $\beta_j(\mu, \mathbf{p}, i, j, p, t_j)$ be $\beta_j\left(\cdot|\mu, \mathbf{p}, t_j, D_i\right)$, following Bayes' rule.

In a different approach, we do not compute the blocking sets. Instead, suppose players i and j in the deviation believe that i's type is in some arbitrary set $D_i \subset T_i$ and j's type is in $D_j \subset T_j$. Following Bayes' rule, their posterior belief will be $\beta_i\left(\cdot|\mu, \mathbf{p}, t_i, D_j\right)$ and $\beta_j\left(\cdot|\mu, \mathbf{p}, t_j, D_i\right)$, respectively. The sets of types that make i and j want to deviate with these posterior beliefs are $d_i\left(D_j\right)$ and $d_j\left(D_i\right)$, respectively, where

$$\begin{array}{l} d_i\left(D_j\right) = \left\{t_i : \mathbf{E}\left(a_{ij}|\mu, \mathbf{p}, t_i, D_j\right) + p > \mathbf{E}\left(a_{i\mu(i)}|\mu, \mathbf{p}, t_i, D_j\right) + p_{i\mu(i)}\right\} \\ d_j\left(D_i\right) = \left\{t_j : \mathbf{E}\left(b_{ij}|\mu, \mathbf{p}, t_j, D_i\right) - p > \mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}, t_j, D_i\right) - p_{\mu(j)j}\right\} \end{array}. \tag{4.11}$$

Players' initial assumptions that their opponent's types are in D_i and D_j are confirmed correct *if and only if*

$$d_i\left(D_j\right) = D_i \text{ and } d_j\left(D_i\right) = D_j. \tag{4.12}$$

Of course, the existence and uniqueness of non-empty (D_i, D_j) that satisfies (4.11) and (4.12) are guaranteed.

This above operation can be intuitively understood as follows: a player, say i, in the deviation makes the following claim: my type is in D_i and I think your type is in D_j; if so, let's deviate together; indeed, conditional on your type being in D_j I benefit from the deviation if and only if my type is in D_i so you should believe that my type is in D_i; if you believe my type is in D_i, you gain from the deviation if and only if your type is in D_j, therefore, I should believe your type is in D_j.

We present two examples to demonstrate the intuitive idea and explain why it is different from weak consistency.

4.4.1 Motivating Examples

Example 1. *There are only one worker i and one firm j. Each of them has two types $T_i = \{t_i, t'_i\}$ and $T_j = \left\{t_j, t'_j\right\}$. The prior β^0 is uniform. The matching value (a_{ij}, b_{ij}) is as follows*

	t_j	t'_j
t_i	$5, 5$	$-1, -1$
t'_i	$-1, -1$	$-1, -1$

and we assume $a_{ii}(\cdot) = b_{jj}(\cdot) = 0$. Consider the matching function M that has both palyers unmatched regardless of types. The on-path belief is uniform. Consider off-path beliefs β_i (resp. β_j) that assigns probability 0.9 to the opponent being t'_j (resp. t'_i). Clearly $D_i = D_j = \emptyset$. Therefore, (M, β) is a weakly consistent stable configuration.

However, for this common interest game, it is quite intuitive that the worker of type t_i and the firm of type t_j can form a coalition to block the no-trade outcome. For instance, worker i of type t_i can make the following announcement to firm j: "I'm type t_i and if you're type t_j, let's match" and firm j of type t_j can make a similar annoucement to worker i: "I'm type t_j and if you're type t_i, let's match". The two announcements are compatible in the following sense: only the worker i of type t_i will gain from the announced plan, so firm j of type t_i has no reason to doubt its sincerity, and vice versa. The takeaway message from this example is that the off-path belief β_i (resp. β_j) that assign 0.9 to the opponent being t'_j (resp. t'_i) is not perfectly reasonable, even if it is weakly consistent. We can strengthen the refinement.

Example 2. *Consider a two-player game, each player has two types: $T_i = \{t_i, t'_i, t''_i\}$ and $T_j = \left\{t_j, t'_j, t''_j\right\}$. The prior $\beta^0 \in \Delta(T_i \times T_j)$ is uniform. The matching value (a_{ij}, b_{ij}) is as follows*

	t_j	t'_j	t''_j
t_i	$1, 1$	$1, 2$	$-1, -1$
t'_i	$2, 1$	$-3, -3$	$-1, -1$
t''_i	$-1, -1$	$-1, -1$	$-1, -1$

and we assume $a_{ii}(\cdot) = b_{jj}(\cdot) = 0$. Consider the matching function M that lets both palyers unmatched regardless of types. The on-path belief is uniform. Consider off-path beliefs β_i (resp. β_j) that assigns probability 0.9 to the opponent being t''_j (resp. t''_i). Consider M that leaves both players alone regardless of types. The blocking sets with respect to the off-path beliefs are empty, so (M, β) is a weakly consistent stable matching-belief configuration. Consider the deviation with $p = 0$ that involves i's types $D_i = \{t_i\}$ and j's types $D_j =$

$\left\{t_j, t'_j\right\}$. *Then for each* $\widetilde{t}_i \in T_i$, $\beta^0(\cdot|\mu, \mathbf{p}, \widetilde{t}_i, D_j)$ *assigns equal probability to* t_j *and* t'_j. *With this belief, only* t_i *will join the deviation. For each* $\widetilde{t}_j \in T_j$, $\beta^0(\cdot|\mu, \mathbf{p}, \widetilde{t}_j, D_i)$ *assigns probability* 1 *to* t_i *and the set of j's types that gain from the deviation is* $D_j = \{t_j, t'_j\}$. *Therefore,* $D_i = \{t_i\}$ *and* $D_j = \left\{t_j, t'_j\right\}$ *are compatible.*

For this game, $D'_i = \{t_i, t'_i\}$ *and* $D'_j = \{t_j\}$ *form the other fixed point. So incorporating the restriction into the off-path beliefs is not straightforward.*

4.4.2 Formulation of Strong Consistency

The idea intuitively described around (4.11) and (4.12) and demonstrated in the two examples above can be formalized as follows.

We approch the refinement as follows.

Definition 5. *We say a configuration* (M, β) *is* ***strongly off-path consistent*** *if*

(i) it is weakly off-path consistent and

(ii) for any pairwise deviation $(\mu, \mathbf{p}, i, j, p)$, *and any* $t_i \in M_i^{-1}(\mu, \mathbf{p})$ *and* $t_j \in M_j^{-1}(\mu, \mathbf{p})$, *the blocking sets with respect to* $\beta_i(\mu, \mathbf{p}, i, j, p, t_i)$ *and* $\beta_j(\mu, \mathbf{p}, i, j, p, t_j)$ *are non-empty if there exists non-empty* (D_i, D_j) *such that* $t_i \in D_i$, $t_j \in D_j$, *and*

$$\begin{array}{l} D_i = \left\{t_i : \mathbf{E}\left(a_{ij}|\mu, \mathbf{p}, t_i, D_j\right) + p > \mathbf{E}\left(a_{i\mu(i)}|\mu, \mathbf{p}, t_i, D_j\right) + p_{i\mu(i)}\right\} \\ D_j = \left\{t_j : \mathbf{E}\left(b_{ij}|\mu, \mathbf{p}, t_j, D_i\right) - p > \mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}, t_j, D_i\right) - p_{\mu(j)j}\right\} \end{array}. \tag{4.13}$$

We say a configuration (M, β) *is* ***strongly consistent*** *if it is on-path conistent and srongly off-path consistent.*

The second requirement says that if the blocking formulated by (4.13) is possible, then blocking should be permitted under $\beta_i(\mu, \mathbf{p}, i, j, p, t_i)$ and $\beta_j(\mu, \mathbf{p}, i, j, p, t_j)$, the belief system in (M, β). The difference between weak and strong consistency is summarized below. Its proof is by comparing definitions and hence omitted.

Lemma 2. *A weakly consistent* (M, β) *is blocked by* $(\mu, \mathbf{p}, i, j, p)$ *only if there exists non-empty sets* D_i *and* D_j *that satisfy (4.13). A strongly consistent* (M, β) *is blocked by* $(\mu, \mathbf{p}, i, j, p)$ *if and only if there exists non-empty non-empty sets* D_i *and* D_j *that satisfy (4.13).*

In dynamic non-cooperative games in which types are independent under prior beliefs, it is common to assume that types remain independent after any history (see **fudenberg1991 fudenberg1991**, p. 237). Naturally, we shall consider *independent* on-path beliefs after any observables; that is, workers' types are independent under $\beta^0(\cdot|M^{-1}(\mu, \mathbf{p}))$ for all $(\mu, \mathbf{p}) \in M(T)$.

Definition 6. *A configuration* (M, β) *is* ***independent*** *if*

$$\beta^0(t|\mu, \mathbf{p}) = \prod_{i \in I \cup J} \beta^0(t_i|\mu, \mathbf{p})$$

for all $t \in T$ *and all* $(\mu, \mathbf{p}) \in M(T)$.

Strong consistency and independence have powerful implications in games with comonotonic differences. The main result of this paper concerns stable matching of games with comonotonic differences.

Theorem 3. *Suppose the matching game has comonotonic differences. A strongly Bayesian consistent stable configuration* (M, β) *with indepedent beliefs is Bayesian efficient.*

The proof proceeds in the following stpes, and the separate results are useful to understand the implications of stability. The first result concerns the implication of comonotonicity and indepenence.

Lemma 3. *Suppose* $f, g : X_1 \times X_2 \to \mathbb{R}$ *are comonotonic on both* X_1 *and* X_2, *and for some constants* c_1 *and* c_2,

$$\mathbf{E}(f) > c_1 \text{ and } \mathbf{E}(g) > c_2, \tag{4.14}$$

where the expectation is with respect to some product measure on $X_1 \times X_2$. *Then there exist non-empty sets* $D_1^* \subset X_1$ *and* $D_2^* \subset X_2$ *such that*

$$\begin{array}{l} D_1^* = \{x_1 : \mathbf{E}(f|x_1, D_2^*) > c_1\} \\ D_2^* = \{x_2 : \mathbf{E}(g|x_2, D_1^*) > c_2\} \end{array}. \tag{4.15}$$

Two-dimensional comonotonicity of f and g implies that the mapping defined on the right-hand side of (4.15) is order-reversing, and an application of Tarski's fixed point theorem to the twice iteration of the mapping has a fixed point, a modification of which is the desired fixed point (D_1^*, D_2^*), and (4.14) ensures its non-emptiness.

Corollary 1 gives simple condition for blocking without the need of computing blocking sets.

Corollary 1. *Suppose the matching game has comonotonic differences and* β_0 *is independent. Then a strongly Bayesian consistent* (M, β) *is blocked by* $(\mu, \mathbf{p}, i, j, p)$ *if*

$$\begin{array}{l} \mathbf{E}\left(a_{ij}|\mu, \mathbf{p}\right) + p > \mathbf{E}\left(a_{i\mu(i)}|\mu, \mathbf{p}\right) + p_{i\mu(i)} \\ \mathbf{E}\left(b_{ij}|\mu, \mathbf{p}\right) - p > \mathbf{E}\left(b_{\mu(j)j}|\mu, \mathbf{p}\right) - p_{\mu(j)j} \end{array} \tag{4.16}$$

Taking $f = a_{ij} - a_{i\mu(i)}$ and $g = b_{ij} - b_{\mu(j)j}$, Lemma 3 establishes the existence of non-empty (D_1^*, D_2^*) that satisfies (4.13), which are blocking sets according to Lemma 2.

Now Theorem 3 follows as follows. Individual rationality of a stable matching (M, β) implies (4.6) and (4.7). By Lemma 1, if (M, β) is not efficient, (4.5) would be violated and hence there exists p such that (4.16) holds. Corollary 1 would imply (M, β) is blocked, a contradicton.

4.5 Bayesian Efficiency and Stability

Example 5. Consider a market with two workers and one firm. The matching values of each worker and the firm are comonotonic, and are as follows:

t_1	t'_1	t_2	t'_2
$(0.5, 5)$	$(1, 6)$	$(-2, 4)$	$(-1.9, 12)$

Suppose that $\beta^0(t_1, t_2) = \beta^0(t'_1, t'_2) = \frac{1}{2}$. Thus, the workers' types are not independent.

Consider a matching M in which the firm hires worker 2 at a price of 2 regardless of the workers' types. In this case, the Bayesian consistent on-path belief is the same as the prior belief β^0. This matching is not Bayesian efficient: it generates an expected total surplus of $\frac{1}{2} \times (-2+4) + \frac{1}{2} \times (-1.9+12) = 6.05$, while the matching in which the firm hires worker 1 generates an expected total surplus of $\frac{1}{2} \times (0.5+5) + \frac{1}{2} \times (1+6) = 6.25$.

But the matching M is stable with Bayesian consistent beliefs. The firm's expected payoff in this matching is $\frac{1}{2} \times 4 + \frac{1}{2} \times 12 - 2 = 6$. Consider a deviating coalition that involves the firm and worker 1 with a price p. No price p is such that only the type t_1 of worker 1 joins the coalition. If the price p is such that both types of worker 1 join the coalition, i.e., $p > -0.5$, then the firm's expected payoff is $\frac{1}{2} \times 5 + \frac{1}{2} \times 6 - p < 6$. In this case the firm rejects the coalition. If the price p is such that only the type t'_1 of worker 1 joins the coalition, then the firm's payoff cannot be higher than 7, the total surplus produced by the pair. But because the two workers' types are correlated, when worker 1's type is t'_1, worker 2's type must be t'_2, and the firm infers that its payoff from M by matching with worker 2 is $12 - 2 = 10$. Therefore, the firm rejects the coalition with worker 1 in this case as well.

4.5.1 Proof of Lemma 3

Suppose without loss of generality that both f and g are non-decreasing with respect to some complete orders $\geq_n$ on X_n. Then consider the class of upper contour sets $B_n(x_n) = \{x'_n : x'_n \geq_n x_n\}$. Let $\mathbb{B}_n = \{B_n(x_n) : x_n \in X_n\} \cup \{\emptyset\}$. Define $d_1 : \mathbb{B}_2 \to \mathbb{R}$ and $d_2 : \mathbb{B}_1 \to \mathbb{R}$

as follows:

$$\begin{array}{l} d_1(D_2) := \{x_1 : \mathbf{E}(f|x_1, D_2) > c_1\} \\ d_1(\emptyset) := X_2 \\ d_2(D_1) := \{x_2 : \mathbf{E}(g|x_2, D_1) > c_2\} \\ d_2(\emptyset) := X_1 \end{array} . \tag{4.17}$$

It follows from $\mathbf{E}(f) > c_1$ and $\mathbf{E}(g) > c_2$ that

$$d_1(X_2) \neq \emptyset \neq d_2(X_1).$$

Define d on $\mathbb{B}_1 \times \mathbb{B}_2$ as $d(D_1, D_2) = (d_2(D_1), d_1(D_2))$. By monotonicity of f and g, we have $d_1(D_2) \in \mathbb{B}_1$ and $d_2(D_1) \in \mathbb{B}_2$. Therefore d is a self-map on $\mathbb{B}_1 \times \mathbb{B}_2$.

For any $x_1' \geq_1 x_1$ and $x' \geq_2 x_2$, we have

$$\begin{array}{l} B_1(x_2') \subset B_1(x_2) \\ B_2(x_1') \subset B_2(x_1) \end{array} . \tag{4.18}$$

By monotonicity of f and g, we have

$$\begin{array}{l} d_1(B_2(x_2)) \subset d_1(B_2(x_2')) \\ d_2(B_1(x_1)) \subset d_2(B_1(x_1')) \end{array} . \tag{4.19}$$

Notice that $\mathbb{B}_1 \times \mathbb{B}_2$ is a complete lattice in the set-inclusion order. It follows from (4.17), (4.18), and (4.19) that d is order-reversing. Therefore $d^2 : \mathbb{B}_1 \times \mathbb{B}_2 \to \mathbb{B}_1 \times \mathbb{B}_2$ is order-preserving. By Tarski's fixed point theorem, d^2 admits a fixed point (D_1, D_2).By definition,

$$\begin{aligned} d^2(D_1, D_2) &= d(d_1(D_2), d_2(D_1)) \\ &= (d_1(d_2(D_1)), d_2(d_1(D_2))) \\ &= (D_1, D_2). \end{aligned}$$

Thus $d_1(d_2(D_1)) = D_1$ and hence $(D_1, d_2(D_1))$ is a fixed point of d. The fixed point cannot be of the form $(\emptyset, D)$ because $D = d_2(\emptyset) = X_2$ but $d_1(X_2) \neq \emptyset$. Similarly, fixed point cannot be of the form $(D, \emptyset)$ because $D = d_1(\emptyset) = X_1$ but $d_Y(X_1) \neq \emptyset$. Therefore, the fixed point of d is non-empty.