MATCHING WITH INCOMPLETE INFORMATION

By

Quingmin Liu, George J. Mailath,
Andrew Postlewaite, and Larry Samuelson

August 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1870

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Matching with Incomplete Information*,†

Qingmin Liu
Department of Economics
Columbia University
New York, NY 10027
qingmin.liu@columbia.edu

George J. Mailath
Department of Economics
University of Pennsylvania
Philadelphia, PA 19104
gmailath@econ.upenn.edu

Andrew Postlewaite
Department of Economics
University of Pennsylvania
Philadelphia, PA 19104
apostlew@econ.sas.upenn.edu

Larry Samuelson
Department of Economics
Yale University
New Haven, CT 06520
larry.samuelson@yale.edu

August 26, 2012

Abstract A large literature uses matching models to analyze markets with two-sided heterogeneity, studying problems such as the matching of students to schools, residents to hospitals, husbands to wives, and workers to firms. The analysis typically assumes that the agents have complete information, and examines core outcomes. We formulate a notion of stable outcomes in matching problems with one-sided asymmetric information. The key conceptual problem is to formulate a notion of a blocking pair that takes account of the inferences that the uninformed agent might make from the hypothesis that the current allocation is stable. We show that the set of stable outcomes is nonempty in incomplete information environments, and is a superset of the set of complete-information stable outcomes. We provide sufficient conditions for incomplete-information stable matchings to be efficient.

*For helpful comments and suggestions, we thank Yeon-Koo Che, Prajit Dutta, Nicole Immorlica, Fuhito Kojima, Dilip Mookherjee, Andrea Pratt, Bernard Salanie, Roberto Serrano, and Rajiv Vohra.

†We thank the National Science Foundation (grants SES-0350969, SES-0549946, SES-0648780, and SES-1153893) for financial support.

Matching with Incomplete Information

by

Qingmin Liu, George J. Mailath, Andrew Postlewaite, and Larry Samuelson

Contents

Matching with Incomplete Information

by

Qingmin Liu, George J. Mailath, Andrew Postlewaite, and Larry Samuelson

1 Introduction

A large literature uses the matching models introduced by Gale and Shapley (1962) and Shapley and Shubik (1971) to analyze markets with two-sided heterogeneity, studying problems such as the matching of students to schools, residents to hospitals, husbands to wives, and workers to firms.[1] The typical analysis in this literature assumes that the agents have complete information, and then examines stable outcomes. A proposed outcome that matches each firm to a worker (for example), along with a specification of a transfer from the firm to the worker, is *stable* if there is no unmatched worker-firm pair that could both increase their payoffs by matching with each other and making an appropriate transfer.

The assumption of complete information makes the analysis tractable but is implausible.[2] This paper examines matching models in which the agents on one side of the market cannot observe the characteristics of those on the other side. To what extent does ignoring the asymmetry of information in a matching problem alter outcomes? What does it mean for an outcome to be stable under incomplete information? What are the properties of stable outcomes?

Our first order of business is to formulate an appropriate modification of stability for problems in which there is asymmetric information. Incorporating incomplete information raises subtle issues. Consider a worker/firm matching problem in which each worker and each firm has a nonnegative index that is their "quality," and any matched worker-firm pair can generate a surplus that is increasing in both of their qualities. Suppose that firms' qualities are commonly known, but workers' qualities are not. Each firm knows the quality of the worker it is matched with, and knows the transfers in other worker-firm pairs, but not the workers' qualities in those pairs. As in the complete information framework, we would say that the outcome is not stable if there is an unmatched worker-firm pair that can deviate and increase the payoff to each. But with incomplete information, how does a

[1] See Roth and Sotomayer (1990) for a survey of two-sided matching theory.

[2] Moreover, there is no mechanism yielding stable matchings under which the truthful revelation of preferences is a dominant strategy for all agents (Roth, 1982), and hence incomplete information will in general have substantive behavioral implications.

firm know whether it will be better off by deviating to match with a worker whose quality isn't known? At first pass, one might think that the answer is simple: the firm should maximize its expected payoff in deciding whether a potential deviation is profitable. But what beliefs should the firm have? Both parties must agree on a proposed deviation, and a firm may be able to make inferences about a worker's quality by his willingness to deviate. In addition, the firm may make further inferences from the fact that no other worker-firm pairs wish to deviate. We formulate a notion of stable outcomes that takes account of all such inferences that firms might make. This notion is defined in terms of an iterative belief-formation process reminiscent of rationalizability.

We are not the first to study these kinds of questions, and we discuss the related literature in Section 5.

We emphasize that our notion of pairwise stability for incomplete information environments is aimed at understanding the stability of possible outcomes. We start with the presumption that a matching exists, and ask whether this matching is stable. To answer this question, we assume that each firm knows the quality of the worker with whom it is matched, that each firm supposes the proposed match is stable, and that firms make use of all of the inferences they can draw from this information, and then ask whether any worker and firm can form a blocking pair. By considering every possible outcome (matching and transfer) in this way, we construct the set of stable outcomes. We do not address *how* stable outcomes might arise. This is in keeping with the vast majority of work on complete-information matching. However, under complete information, one can at least imagine an underlying matching *process*: unmatched agents randomly meet each other and make proposals, with the process stopping when no unmatched pair can improve on their situation by matching. It is less obvious how one would construct a matching process leading to stable outcomes, because agents make inferences from intermediate outcomes during the matching process, so the set of possible incomplete-information stable outcomes becomes a "moving target." Providing noncooperative foundations for incomplete-information stable matchings is an obviously interesting problem but separate problem. We return to this issue in Section 5.

Our notion of stability precludes profitable pairwise deviations, but does not consider deviations by groups of agents other than a single worker and a single firm. Under complete information, it is easy to show that pairs can block any outcome that could be blocked by larger coalitions, and hence restricting attention to pairwise stability sacrifices no generality. That need not be the case with incomplete information. Given our assumption that in

a proposed matching firms know the quality of the worker with whom they are matched, a coalition that includes more than a single firm potentially has more information—"potentially" because one would have to specify the process by which firms communicated, presumably accounting for incentives, in order to characterize the information at their disposal. We allow only pairs to deviate to avoid the problems that arise when larger coalitions are allowed. Section 5 explains why many of our main results would carry over to models that allowed larger coalitions.

Sections 2 and 3 develop our stability concept for matching problems with incomplete information. The key is to identify all the inferences a firm can make from what it knows directly and from the hypothesis that the proposed match is stable. We think of firms repeatedly observing a matching outcome, leading them to conclude that there are no blocking pairs, and in turn to draw conclusions about various workers' types.

Section 4 explores the implications of our notion of incomplete information stability. Under quite general conditions, pairwise stable outcomes exist in incomplete-information environments. Under intuitive sufficient conditions, these outcomes are efficient, but in general they can fail equal treatment of equals. Incomplete-information pairwise stable outcomes are a superset of complete-information stable outcomes. Agents' payoffs in stable incomplete-information problems with "little" asymmetry of information are close to the payoffs to those with no asymmetry.

2 Matching with Incomplete Information

2.1 The Environment

We generalize the complete-information matching models studied by Shapley and Shubik (1971) and Crawford and Knoer (1981). There is a finite set of workers, I, with an individual worker denoted by $i \in I$. There is also a finite set of firms, J, with an individual firm denoted by $j \in J$. Indices identify agents, but do not play a direct role in production. We use male pronouns for workers and female for firms.

The productive characteristics of an agent are described by the agent's *type*, with W being the set of possible worker types and F being the set of possible firm types. The function mapping each firm to her type is denoted $\mathbf{f} : J \to F$. The function mapping each worker to his type is denoted $\mathbf{w} : I \to W$.

Value is generated by matches. We take as primitive the aggregate match value each agent receives in the absence of any transfers between the agents.

Following Mailath, Postlewaite, and Samuelson (2012a,b), we call these values *premuneration values*. For example, the firm's premuneration value may include the net output produced by the worker with whom the firm is matched, the cost of the unemployment insurance premiums the firm must pay, and (depending on the legal environment) the value of any patents secured as a result of the worker's activities. The worker's premuneration value may include the value of the human capital the worker accumulates while working with the firm, the value of contacts the worker makes in the course of his job, and (again depending on the legal environment) the value of any patents secured as a result of the worker's activities.

A match between worker type $w \in W$ and firm type $f \in F$ gives rise to the worker premuneration value $\nu_{wf} \in \mathbb{R}$ and firm premuneration value $\phi_{wf} \in \mathbb{R}$. We call the sum of the premuneration values, $\nu_{wf} + \phi_{wf}$, the *surplus of the match*. We avoid having to continually make special note of nuisance cases by also defining the premuneration values of an unmatched worker and an unmatched firm, which we take (without loss of generality) to be zero, denoting these values by $\nu_{w(\varnothing),f(j)}$ for the worker and $\phi_{w(i),f(\varnothing)}$ for the firm.

Each firm's index is commonly known, as is the function $\mathbf{f}$, and hence each firm's type is common knowledge. On the other hand, while a worker's index is common knowledge, the function $\mathbf{w}$ (and hence workers' types) will in general not be known (though workers will know their own types). We assume the worker type assignment $\mathbf{w}$ is drawn from some distribution with support $\Omega \subset W^I$. As will be clear, while the support plays an important role in the analysis, the distribution does not. The functions $\nu : W \times F \to \mathbb{R}$ and $\phi : W \times F \to \mathbb{R}$ are common knowledge.

Given a match between worker i (of type $\mathbf{w}(i)$) and firm j (of type $\mathbf{f}(j)$), the worker's payoff is

$$\pi_i^w := \nu_{\mathbf{w}(i),\mathbf{f}(j)} + p,$$

while the firm's payoff is

$$\pi_j^f := \phi_{\mathbf{w}(i),\mathbf{f}(j)} - p,$$

where $p \in \mathbb{R}$ is the transfer paid to worker i by firm j. We often refer to this transfer as a wage, though it might be negative.

A *matching function* is a function $\mu : I \to J \cup \{\varnothing\}$, one-to-one on $\mu^{-1}(J)$, that assigns worker i to firm $\mu(i)$, where $\mu(i) = \varnothing$ means that worker i is unemployed and $\mu^{-1}(j) = \varnothing$ means that firm j does not hire a worker. The outcome of such a function is a *matching*.

A *transfer scheme* $\mathbf{p}$ associated with a matching function μ is a vector that specifies a transfer $\mathbf{p}_{i,\mu(i)} \in \mathbb{R}$ for each $i \in I$ and $\mathbf{p}_{\mu^{-1}(j),j} \in \mathbb{R}$ for each $j \in J$. To again avoid nuisance cases, we associate zero transfers with unmatched agents, setting $\mathbf{p}_{\varnothing j} = \mathbf{p}_{i\varnothing} = 0$.

Definition 1 *An* allocation $(\mu, \mathbf{p})$ *consists of a matching function* μ *and a transfer scheme* $\mathbf{p}$ *associated with* μ. *An* outcome *of the matching game* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *specifies a realized type assignment* $(\mathbf{w}, \mathbf{f})$ *and an allocation* $(\mu, \mathbf{p})$.

2.2 An Example

We illustrate the environment and preview our stability notions. There are three workers and firms ($I = J = \{1, 2, 3\}$). The set of possible worker types is $W = \{1, 2, 3\}$ and the set of possible firm types is $F = \{2, 4, 5\}$. The firm type assignment is given by $\mathbf{f}(1) = 2$, $\mathbf{f}(2) = 4$, and $\mathbf{f}(3) = 5$. A worker of type w and a firm with type f generate a premuneration value wf to each agent, i.e., $\nu_{wf} = \phi_{wf} = wf$.

2.2.1 Complete Information

Suppose the worker type assignment is $\mathbf{w}(1) = 1$, $\mathbf{w}(2) = 3$, and $\mathbf{w}(3) = 2$, and that this is commonly known. The notion of stability for this complete-information setting is familiar from Gale and Shapley (1962). Because the surplus function is supermodular, the only stable matching must be positively assortative in type (Shapley and Shubik, 1971), which is the efficient matching (in the sense of maximizing total surplus).

To illustrate the reasoning behind this result, consider the matching shown in Figure 1, which is not assortative. Since the matching of worker 2 (who has type 3) with firm 2 (who has type 4) generates a surplus of 24, we have $\pi_2^w + \pi_2^f = 24$, and similarly $\pi_3^w + \pi_3^f = 20$. But the surplus generated by a positively assortative matching by type of the top two workers and firms is 46. In the candidate match of Figure 1, either $\pi_2^w + \pi_3^f < 30$ or $\pi_3^w + \pi_2^f < 16$, and hence either worker 2 and firm 3, or worker 3 and firm 2, can form a blocking coalition (i.e., can match and make a transfer under which both receive more than under the candidate match).

2.2.2 Incomplete Information

Now suppose that the firms know the workers' indices, know the set of possible worker types $W = \{1, 2, 3\}$, and know the type of worker with whom they are matched, but do not know the function $\mathbf{w}$ assigning types

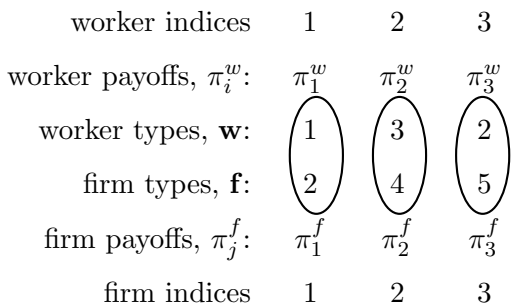

Figure 1: A matching that cannot be complete-information stable. Workers and firms are indexed by column. The matching of types is indicated by the ovals: $\mu(i) = i$, for $i \in \{1, 2, 3\}$.

to indices. Suppose the realized types and the matching of firms to workers are as in Figure 1, with the transfers and payoffs shown in Figure 2. Firms believe the set Ω of possible vectors $(\mathbf{w}(1), \mathbf{w}(2), \mathbf{w}(3))$ is (in this example) the set of permutations of $(1, 2, 3)$. Hence, each firm knows there is one worker of type 1, one worker of type 2, and one worker of type 3, and knows the type of her own worker, but does not know the types of the other two workers.

We consider a stability notion analogous to that of the complete-information case, namely that there be no unmatched pair who can find an agreement that both prefer to the proposed outcome. Consider a candidate blocking pair consisting of worker 3, firm 2, and some transfer $\tilde{p} \in (-2, 0)$. Under complete information, this would indeed be a blocking pair. Under incomplete information, it is again immediate that any such agreement makes a worker of type 2 better off than in the proposed outcome, and hence satisfies one condition for being a blocking pair. However, firm 2 does not know whether worker 3 is of type 1 or type 2. The proposed deal is advantageous for firm 2 if the worker is type 2, but not if the worker is type 1.

Is this a blocking pair? To answer this question, both here and in general, we must take a stand on what beliefs the firm is likely to have about the type of worker in a proposed blocking pair. Our requirement will be that a pair can block only if both agents expect higher payoffs, given *any reasonable* beliefs the firm might have over the support of possible worker types. Could the firm reasonably expect worker 3 to be type 1? It initially appears that this is the case, since firm 2 knows only that worker 3 is not of type 3.

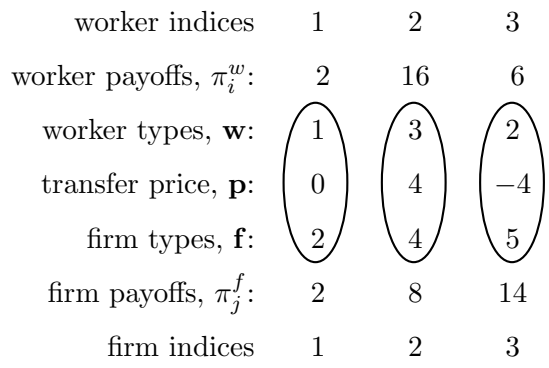

Figure 2: A possible outcome of the worker type assignment under incomplete information, with a matching outcome, transfers, and payoffs. Types and premuneration values match those of Figure 1; workers and firms are indexed by column, and the matching is by index (indicated by the ovals).

However, the firm may be able to refine her beliefs on the strength of the fact that worker 3 is willing to participate in the block. To pursue this, notice that if worker 3 were type 1, his current payoff would be 1, while he would receive a payoff of $4+\tilde{p}$ in the candidate blocking pair. Since $4+\tilde{p} > 1$ for all $\tilde{p} \in (-2, 0)$, the candidate blocking pair is also advantageous for a worker of type 1. Firm 2 then cannot be sure whether the proposal involves a worker of type 1 or type 2. Hence, the firm could reasonably believe the worker is of type 1, making the proposed deal disadvantageous for the firm. The allocation illustrated in Figure 2 thus appears to be incomplete-information stable.

However, the argument does not end here. The "reasonable" requirement we place on the firms' beliefs is that the support of these firm's beliefs be consistent with *all* of the inferences the firm can draw, using the firm's information and the hypothesis that the candidate allocation is stable. In this case, firm 2 can reason as follows: Suppose worker 3 *were* of type 1. Then firm 3 would receive a payoff of 9, worker 1 would be of type 2 and would receive payoff 2, and firm 3 would know that worker 1 was of type at least 2. Worker 1 and firm 3 could the match at wage 0 (for example), giving each a higher payoff than the candidate stable allocation and thus constituting a blocking pair. But firm 2's working hypothesis is that the proposed allocation is stable, and hence that there is no such blocking pair.

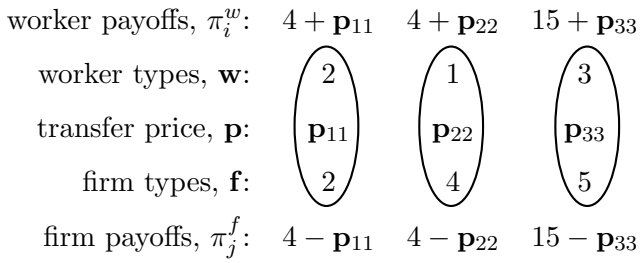

Figure 3: A matching in which the lowest type worker does not match with the lowest type firm. Workers and firms are indexed by column. Types and premuneration values are from Figure 1. The matching of types is indicated by the ovals: $\mu(i) = i$, for $i \in \{1, 2, 3\}$. The worker type assignment is from Section 2.2.3.

If the candidate allocation is to be stable, then worker 3 cannot be of type 1, and hence worker 3 must be of type 2. This ensures that the proposed block is profitable for firm 3, and hence that we indeed have a blocking coalition. The allocation illustrated in Figure 2 is thus *not* incomplete-information stable.

The central issue addressed in this paper is to make precise and then explore the implications of this belief-formation process.

2.2.3 Incomplete Information: Inference

Firm 2's inference in the preceding section does not hinge critically on the strong assumptions made about the possible worker type distributions. In particular, we preview a general result: if premuneration values are increasing and strictly supermodular, then only positive assortative matchings can be stable.

The firms' types are again given by $\mathbf{f}(1) = 2$, $\mathbf{f}(2) = 4$, and $\mathbf{f}(3) = 5$. Assume nothing more about worker types than that the set of possibilities is $W = \{1, 2, 3\}$. Workers' types may be drawn independently from this set, or may be drawn according to any other procedure. Premuneration values are given by $\nu_{wf} = \phi_{wf} = wf$.

We first argue that the lowest type worker must be matched with the lowest type firm. Consider the matching in Figure 3, which pairs the worker of the lowest type with the firm of the second lowest type.

Suppose first that

$$\mathbf{p}_{11} > 4 + \mathbf{p}_{22},$$

and consider a candidate blocking pair involving worker 2, firm 1, and transfer $p = (\mathbf{p}_{11} + \mathbf{p}_{22})/2$. Worker 2 finds the resulting payoff strictly preferable to the current matching, since $2 + p > 4 + \mathbf{p}_{22}$. Moreover, a lower bound on firm 1's payoff under such an offer is provided by assuming that worker 2 has type 1, and so firm 1 also finds such an offer strictly preferable to the current matching, since $2 - p > 4 - \mathbf{p}_{11}$.

Suppose instead that

$$\mathbf{p}_{11} \leq 4 + \mathbf{p}_{22},$$

and consider a candidate blocking pair involving worker 1, firm 2, and transfer $p = \mathbf{p}_{11} - 3$. Worker 1 finds the resulting payoff strictly preferable to the current matching, since $8 + p > 4 + \mathbf{p}_{11}$. In computing a lower bound on her payoff under such an offer, firm 2 should understand that worker 1 of type 1 does not find such a match attractive, since $4 + p < 2 + \mathbf{p}_{11}$. Under the belief that worker 1 has type at least 2, firm 2 then also finds the offer strictly preferable to the current matching, since $8 - p > 4 - \mathbf{p}_{22}$.

This ensures that the lowest types of firm and worker must be matched. This logic can be iterated to show that the lowest two types of workers must be matched with the lowest two types of firms, the lowest three types of workers with the lowest three types, and so on, giving the result that only assortative matchings can be stable.

3 Stability

3.1 Individual Rationality

A matching is *individually rational* if each agent receives at least as high a payoff as provided by the outside option of remaining unmatched, i.e., receives at least zero. Since firms observe the types of workers with whom they are matched at the interim stage, the notion of individual rationality is the same for complete and incomplete information.

Definition 2 *An outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* individually rational *if for all* $i \in I$ *and* $j \in J$,

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \geq 0 \quad \textit{and}$$
$$\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j} \geq 0.$$

3.2 Complete Information Stability

The notion of stability in matching games with transferable utility was first formulated by Shapley and Shubik (1971), who also established existence. Crawford and Knoer (1981) provide a constructive proof of existence by applying a deferred acceptance algorithm to a model with discrete transfers.

Definition 3 *A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* complete-information stable *if it is individually rational, and there is no worker-firm combination* (i, j) *and transfer* $p \in \mathbb{R}$ *from* j *to* i *such that*

$$\nu_{\mathbf{w}(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}$$

and

$$\phi_{\mathbf{w}(i),\mathbf{f}(j)} - p > \phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}.$$

If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is a complete-information stable outcome, the allocation* $(\mu, \mathbf{p})$ *is a* complete-information stable allocation *at* $(\mathbf{w}, \mathbf{f})$.

It is well-known that for each type assignment $(\mathbf{w}, \mathbf{f})$, a complete-information stable allocation exists, is efficient, and agents on the same side of the market obtain the same payoffs if they have the same types (equal treatment of equals).

3.3 Incomplete Information

We are interested in the stability of a matching when each worker knows his type, but the worker type assignment is not known by any agent. We view stability as capturing a notion of steady state: a matching is stable if once established, it remains in place. Think of workers and firms in the labor market observing a particular matching (together with its associated transfers). If the matching is stable, then we should expect to see the same matching when next the labor market opens, and each subsequent time the labor market opens. To make this operational, we characterize the implications of having the stability of the matching be common knowledge. Importantly, we do not examine the process by which the matching comes into being. Rather, we view the matching as already in place, identify the implications of agents understanding that the current match is stable, and then check whether it is indeed stable.

We emphasize what a firm can observe: the types of all firms, the distribution from which the function assigning workers' types is drawn, the type of the firm's current worker, and which worker is matched with which firm

at which price. Hence, a firm assessing a candidate block involving worker i knows the identity and type of the employer with whom i is matched in the supposed stable allocation.

We model the firms' inferences via a procedure of iterated elimination of unstable matching outcomes. This formulation resembles the game-theoretic notion of rationalizability (Bernheim (1984) and Pearce (1984)), obtained via iterated elimination of strategies that are never best responses, though a better analogy may be the deductive iterations that arise in the classic "colored hats" problem with which discussions of common knowledge are often introduced (Geanakoplos, 1994, p. 1439). Similar reasoning lies behind the no-trade theorem of Milgrom and Stokey (1982).

Definition 4 *Fix a non-empty set of individually rational matching outcomes, Σ. A matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma$ is* Σ-stable *if there is no worker-firm combination (i, j) together with a transfer $p \in \mathbb{R}$ such that*

$$\nu_{\mathbf{w}(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}, \tag{1}$$

and

$$\phi_{\mathbf{w}'(i),\mathbf{f}(j)} - p > \phi_{\mathbf{w}'(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j} \tag{2}$$

for all $\mathbf{w}' \in \Omega$ satisfying

$$(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \Sigma, \tag{3}$$

$$\mathbf{w}'(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j)), \quad \textit{and} \tag{4}$$

$$\nu_{\mathbf{w}'(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}. \tag{5}$$

Inequality (1) requires that worker i receives a higher payoff in the proposed deviation than in the putative stable matching. Inequality (2) requires that firm j expect a higher payoff in the proposed deviation than in the putative stable matching, for *any* reasonable beliefs the firm might have over worker type assignments. Our notion of "reasonable" only restricts the supports of such beliefs, and so we suppress the beliefs, describing the restrictions on the supports directly. To qualify as reasonable, a type assignment must satisfy three criteria, given by (3)–(5): (3) the type assignment must be consistent with matching outcomes in the set Σ, a restriction that will become operational in the iterative argument we construct next; (4) the type assignment must not contradict what the firm j already knows at the interim stage, i.e., it must be consistent with the type of firm j's current worker $\mu^{-1}(j)$; and (5) the type of worker i with whom j is matched in the

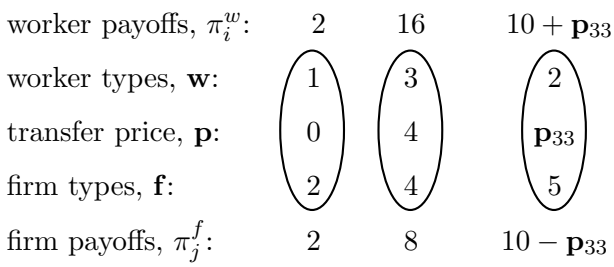

Figure 4: A matching outcome that is not Σ^0-stable (where Σ^0 is the set of individually rational matching outcomes) for the transfer $\mathbf{p}_{33} = -2$, but is Σ^0-stable for the transfer $\mathbf{p}_{33} = -4$ (the outcome from Figure 2). Types and premuneration values are from Figure 1.

proposed deviation must be consistent with i's incentives (i.e., the type of this worker should be better off than under that worker's current match).

The argument in Section 2.2.3 shows that the matching outcome of Figure 3 is not Σ^0-stable, where Σ^0 is the set of all individually rational matching outcomes, irrespective of the level of transfers. That argument is general, and shows that if premuneration values are increasing and strictly supermodular, *no* matching outcome in which a matched lowest type of worker and a lowest type of firm are not matched with each other is Σ^0-stable (Lemma A.3). In other cases, the precise nature of the transfers determines whether the matching outcome is Σ^0-stable. For example, the matching outcome in Figure 4 may or may not be Σ^0-stable, depending on $\mathbf{p}$. Suppose first that $\mathbf{p}_{33} = -2$, and consider a candidate blocking pair consisting of worker 2 (who has type 3) and firm 3, with transfer $p \in (1, 2)$. Worker 2 prefers this resulting match to the proposed equilibrium outcome. Moreover, firm 3 can calculate that a worker matched with firm 2 would prefer such an alternative match if and only if the worker is of type 3, ensuring that firm 3 also strictly prefers the proposed block and hence that the candidate outcome is not Σ^0-stable. In contrast, the outcome with $\mathbf{p}_{33} = -4$ (the outcome from Figure 2) is Σ^0-stable: Note first that worker 2 and firm 3 can no longer block because the total payoff of the pair equals their surplus were that pair to match. Moreover, it is an implication of the discussion in Section 2.2.2 that worker 3 and firm 2 cannot form a blocking pair

While Definition 4 suppresses the role of beliefs, our preferred interpretation is that firms are expected profit maximizers. In particular, when

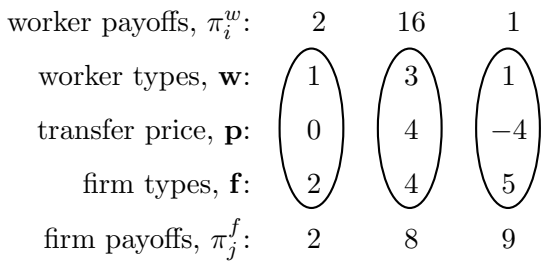

Figure 5: The transfers and matching from Figure 2 with a different worker type realization.

evaluating a potential blocking match with worker i, firm j evaluates the profitability from such a match using her beliefs over worker i's possible type to calculate expected profits. However, we are interested in understanding the basic nature of the possible restrictions implied by incomplete information stability without assuming too much about the nature of the protocols (extensive form) of firm-worker interactions. That is, we only allow blocking if we are confident that firm j believes she benefits. We accordingly require that the firm believe the blocking is profitable under all reasonable beliefs, restricted only by the common knowledge of the structure of the matching.

Definition 5 *Let Σ^0 be the set of all individually rational outcomes. For $k \geq 1$, define*

$$\Sigma^k := \left\{ (\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k-1} : (\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \text{ is } \Sigma^{k-1}\text{-stable} \right\}.$$

The set of incomplete-information stable outcomes *is given by*

$$\Sigma^\infty := \bigcap\nolimits_{k=1}^{\infty} \Sigma^k.$$

If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is an incomplete-information stable outcome, the allocation $(\mu, \mathbf{p})$ is an incomplete-information stable allocation *at $(\mathbf{w}, \mathbf{f})$.*

Consider the outcome in Figure 2. We argued earlier that this outcome is Σ^0-stable. Hence that outcome is in Σ^1. However, the outcome is not Σ^1-stable and hence is not contained in in Σ^2, because outcomes with $\mathbf{w}'(3) = 1$ (such as the one displayed in Figure 5) are not contained in Σ^1 (for Figure 5, consider an offer of $p = -\frac{1}{2}$ by worker 3 to firm 1).

The sequence Σ^k is a (weakly) decreasing sequence of sets of outcomes. As stated in the next proposition, it is straightforward to see that the limit of the sequence, Σ^∞, is nonempty.

Proposition 1 *For each type assignment* $(\mathbf{w}, \mathbf{f})$*, there is an incomplete-information stable outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$*, and so the set of incomplete-information stable allocations is non-empty.*

Proof. If $(\mu, \mathbf{p})$ is a complete-information stable allocation at $(\mathbf{w}, \mathbf{f})$, then by definition $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^k$ for each $k \geq 0$. ■

3.4 Fixed-Point Characterization

The iterative procedure of Definition 5 describes an algorithm for obtaining the set of *all* incomplete-information stable allocations. This set has a fixed-point characterization, which is often more convenient for verifying that a given matching outcome is stable.

Definition 6 *A nonempty set of matching outcomes* E *is* self-stabilizing *if every* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in E$ *is individually rational and if every* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in E$ *is* E*-stable. The set* E stabilizes a given matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *if* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in E$ *and* E *is self-stablizing. A set of worker-type assignments* $\Omega^* \subset \Omega$ stabilizes an allocation $(\mu, \mathbf{p})$ *if* $\{(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) : \mathbf{w} \in \Omega^*\}$ *is a self-stabilizing set.*

We now summarize several useful properties of a self-stabilizing set of matching outcomes (the proof is in Appendix A.1).

Lemma 1

1. *The singleton set* $\{(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})\}$ *is self-stabilizing if and only if* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is a complete-information stable outcome.*
2. *If both* E_1 *and* E_2 *are self-stabilizing, then* $E_1 \cup E_2$ *is self-stabilizing.*
3. *If* E *is self-stabilizing, then its closure* $\overline{E}$ *is self-stabilizing.*[3]
4. *If* E *is a self-stabilizing set and* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in E$*, then* $E \cap \{(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) : \mathbf{w}' \in \Omega\}$ *is also a self-stabilizing set.*

[3] Given any set of outcomes E, the outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is in the closure of E if there is a sequence $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f}^n) \in E$ such that $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f}^n) \to (\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ pointwise. Since $\mu, \mathbf{w}$ and $\mathbf{f}$ are drawn from finite sets, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \overline{E}$ if and only if there exists a sequence $\mathbf{p}^n \to \mathbf{p}$ such that $(\mu, \mathbf{p}^n, \mathbf{w}, \mathbf{f}) \in E$.

The following proposition provides a fixed-point characterization of the set of stable outcomes (the proof is in Appendix A.2):

Proposition 2

1. *If E is a self-stabilizing set, then $E \subset \Sigma^{\infty}$.*
2. *The set of incomplete-information stable outcomes, Σ^{∞}, is a self-stabilizing set, and hence the largest self-stabilizing set.*
3. *The set Σ^{∞} is closed.*

One immediate implication of Proposition 2 is that to show $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is a stable outcome, it suffices to construct a subset Ω^* containing $\mathbf{w}$ that could stabilize the allocation $(\mu, \mathbf{p})$.

4 Implications of Incomplete-Information Stability

4.1 Allocative Efficiency

4.1.1 Payoff Assumptions

Stable matchings have particularly strong implications under some economically relevant constraints on premuneration values:

Assumption 1 (Monotonicity) *The worker premuneration values ν_{wf} and firm premuneration values ϕ_{wf} are strictly increasing in w and f.*

Assumption 2 (Supermodularity) *The worker premuneration values ν_{wf} and firm premuneration values ϕ_{wf} are strictly supermodular in w and f.*

The assumption of supermodularity is common in the literature on labor markets and marriage markets. Its sorting implications in matching markets were first studied by Becker (1973). Note that the supermodularity assumption is imposed on premuneration values, not only on the total surplus.

4.1.2 Information-Revealing Prices

Under supermodularity, a firm faced with evaluating its participation in a potential blocking pair can draw relatively sharp inferences about the type of worker from the worker's willingness to participate in the blocking coalition

at the associated transfer. The following lemma identifies conditions under which a firm entertaining a deviation to match with a worker of unknown type can be certain of a lower bound on the worker's type (the proof is in Appendix A.3).

Lemma 2 *Suppose Assumption 2 (supermodularity) holds, and* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is individually rational. If a type-*w^* *worker is matched with a type-*f^* *firm at a transfer* p^**, then for any firm with type* $f > f^*$*, there exists* $\varepsilon > 0$ *such that for any* $p \in (\nu_{w^*f^*} + p^* - \nu_{w^*f},\ \nu_{w^*f^*} + p^* - \nu_{w^*f} + \varepsilon]$*,*

$$\nu_{wf} + p > \nu_{wf^*} + p^*, \qquad \text{for any } w \geq w^*, \tag{6}$$

$$\nu_{wf} + p \geq 0, \qquad \text{for any } w \geq w^*, \text{ and} \tag{7}$$

$$\nu_{wf} + p \leq \nu_{wf^*} + p^*, \qquad \text{for any } w < w^*. \tag{8}$$

If w^* *is unmatched in an individually rational matching outcome, then for any firm type* f*, there exists* $\varepsilon > 0$ *such that for any* $p \in (-\nu_{w^*f},\ -\nu_{w^*f} + \varepsilon]$*,*

$$\nu_{wf} + p > 0, \qquad \text{for any } w \geq w^*, \text{ and}$$

$$\nu_{wf} + p \leq 0, \qquad \text{for any } w < w^*.$$

The interpretation is as follows. Suppose a worker is willing to participate in a blocking pair with a firm of type $f > f^*$, where f^* is the type of the worker's current match, at a transfer of p just above $\nu_{w^*f^*} + p^* - \nu_{w^*f}$. The type f firm can then infer that the worker would benefit from this proposal if and only if his type is at least w^*. Condition (6) says that all worker types higher than or equal to w^* prefer working for a type f firm under a transfer p to remaining in the old match; (7) says that matching with a type f firm is individually rational; (8) says that if worker type is lower than w^*, then the worker prefers to stay in the candidate matching.

4.1.3 Efficiency

Under supermodularity, an outcome is efficient if and only if it features assortative matching. No additional assumptions are needed to ensure efficiency of stable outcomes. Section A.4 proves the following.

Proposition 3 *If Assumptions 1 (monotonicity) and 2 (supermodularity) hold, then an incomplete-information stable outcome is efficient.*

We now describe an example that demonstrates that without supermodularity stable outcomes may be inefficient. There are two workers and two

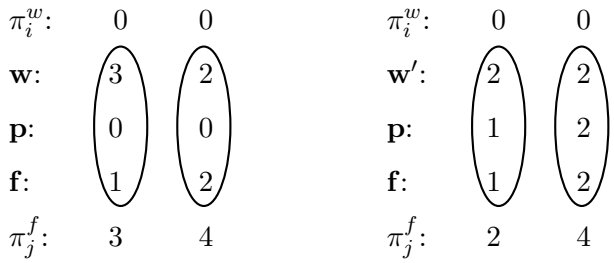

Figure 6: A failure of efficiency in the absence of strict spermodularity. The first matching is incomplete information stable, stabilized by the second complete information stable outcome. In this example, $W = \{2, 3\}$, $F = \{1, 2\}$, $\nu_{wf} = 0$, and $\phi_{wf} = wf$.

firms. It is commonly known that $\mathbf{f}(1) = 1$, $\mathbf{f}(2) = 2$, and $\mathbf{w}(2) = 2$. We suppose that worker 1's type could be either 2 or 3, and the realized value is 3. The worker's premuneration value is zero in any match (thus violating strict supermodularity), while the firm's premuneration value is the product of the types in a match, wf.

We claim that the first matching outcome in Figure 6 is an inefficient, incomplete-information stable outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$. To show $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is an incomplete-information stable outcome, we use Proposition 2 and show that the set $E = \{(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}), (\mu, \mathbf{p}, \mathbf{w}', \mathbf{f})\}$ is a self-stabilizing set, where $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f})$ is given by the second matching.

First note that $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f})$ is a complete-information stable outcome, and hence is self-stablizing as a singleton set. This outcome stabilizes $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ as follows. The only potential blocking pair at type assignment $(\mathbf{w}, \mathbf{f})$ must involve worker 1 and firm 2. However, firm 2 believes that worker 1 could be either type 2 or type 3. There is no nonnegative (so that the worker is willing to participate) transfer at which firm 2 could gain from such a proposed block if worker 1 is type 1. The allocation then cannot be blocked, and hence we have incomplete-information stability.

In this example, the conclusion of the sorting lemma (Lemma 2) fails because worker type 3 has a constant premuneration value and consequently doesn't have an individually rational and incentive compatible way to "convince" the second firm that his type is 3.

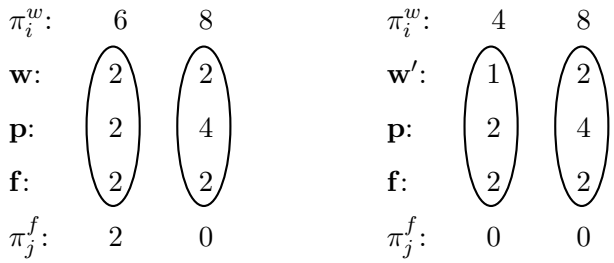

Figure 7: A failure of equal treatment. The first matching is incomplete information stable, stabilized by the second complete information stable outcome. In this example, $W = \{1, 2\}$, $F = \{2\}$ and $\nu_{wf} = \phi_{wf} = wf$.

4.2 Failure of Equal Treatment of Equals

The equal treatment of equals is a basic notion of fairness, and is trivially satisfied by stable outcomes in complete information environments. We have shown that under strict supermodularity and monotonicity, incomplete-information stable matchings exhibit a strong efficiency property. A natural question is whether we also obtain fairness, in the sense of equal treatment of equals.

We now show by example that equal treatment of equal worker types can fail. There are two firms, each of type 2, and two workers, with types drawn independently from the set $\{1, 2\}$. Premuneration values are wf for both workers and firms. Consider the first matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ in Figure 7. This matching outcome violates equal treatment of equals, since the workers are of the same type but receive different payoffs. If there were complete information, the first worker and the second firm would form a blocking pair.

To establish incomplete-information stability, we construct an argument reminiscent of that used in the previous example. We show that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is part of a self-stabilizing set. The easiest way to do so is to consider a set of two outcomes, the second of which $((\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}))$ is complete-information stable.

Consider the self-stabilizing set $E = \{(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}), (\mu, \mathbf{p}, \mathbf{w}', \mathbf{f})\}$, where $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f})$ is given by the second matching outcome in Figure 7. The latter is complete-information stable, so we need only show that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is incomplete-information stable, for which it suffices to show that a coalition consisting of worker 1 and firm 2 cannot block $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$. This follows from

the fact that firm 2 cannot be sure that such a worker is of type 2 rather than type 1, since any transfer inducing a type-2 worker to participate in such a blocking pair would also induce a type-1 worker to participate.

4.3 Relation to Complete-Information Stability

Proposition 1 established that any complete-information stable outcome is incomplete-information stable. The examples in Sections 4.1.3 and 4.2 present incomplete-information stable outcomes that are not complete-information stable. The set of incomplete-information stable outcomes is thus a strict superset of the set of complete-information outcomes. This section describes settings in which the two concepts coincide, or are close.

4.3.1 Almost Complete Information: Continuity

We first seek a continuity result. The motivation for such a result is straightforward. We believe that matching environments invariably involve at least *some* asymmetry of information. At the same time, complete-information models are convenient. It would then be similarly convenient if the equilibrium outcomes of our complete information matching models are "close" to the outcomes of incomplete information matching models when the asymmetry of information is small. Since our notion of incomplete information stability depends only on the support of the distribution determining worker-type assignments, our notion of close is necessarily strong. In particular, it requires the supports to be close (which is not an implication of standard notions of distance on distributions).

We cannot expect such a continuity result without continuity in premuneration values:

Assumption 3 (Continuity) *The premuneration values ν_{wf} and ϕ_{wf} are continuous in w.*

Fix a type assignment $\mathbf{w} \in \mathbb{R}^{|I|}$ and fix $\delta > 0$, and denote by $\xi_\delta(\mathbf{w})$ a δ-neighborhood of $\mathbf{w}$ in the Euclidean metric. Since we will be varying the support Ω, we make the dependence of the set of incomplete information stable outcomes on the support Ω explicit by denoting that set by $\Sigma^\infty(\Omega)$. Note that the set of complete information stable outcomes for a given worker-type assignment $\mathbf{w}$ can be written as $\Sigma^\infty(\{\mathbf{w}\})$. Let $\pi(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \mathbb{R}^{|I|+|J|}$ be the vector of payoffs that workers and firms receive in the matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$, and denote by $\pi(\Sigma^\infty(\Omega)) \in \mathbb{R}^{|I|+|J|}$ the set of payoff

vectors associated with the set of matching outcomes $\Sigma^{\infty}(\Omega)$. Denote by $\xi_{\delta}(\pi(\Sigma^{\infty}(\Omega)))$ the δ-neighborhood of the set $\pi(\Sigma^{\infty}(\Omega))$, that is,

$$\xi_{\delta}(\pi(\Sigma^{\infty}(\Omega))) = \bigcup_{(\mu,\mathbf{w},\mathbf{f},\mathbf{p})\in\Sigma^{\infty}(\Omega)} \xi_{\delta}(\pi(\mu,\mathbf{p},\mathbf{w},\mathbf{f})).$$

We then have that if there is almost complete information about a worker-type assignment $\mathbf{w}$, then the set of incomplete-information stable outcomes is close to the set of complete-information outcomes in terms of payoffs.

Proposition 4 *Suppose Assumption 3 holds. Fix a type assignment* $\mathbf{w} \in \mathbb{R}^{|I|}$. *For any* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that* $\pi(\Sigma^{\infty}(\Omega)) \subset \xi_{\varepsilon}(\pi(\Sigma^{\infty}(\{\mathbf{w}\})))$ *for any finite set* $\Omega \subset \xi_{\delta}(\mathbf{w})$.

4.3.2 Restrictions of Workers' Types

The examples in Sections 4.1.3 and 4.2 present incomplete-information stable outcomes that are not complete-information stable. In these examples, workers' types are determined by independent draws. One's intuition is that firms are able to infer relatively little about workers' types in such an environment. Firms might be able to draw stronger inferences, and the set of incomplete-information stable outcomes might be close to the set of complete-information stable outcomes, if there is correlation among workers' types.

This section considers a very strong form of correlation in workers' types:

Definition 7 *The support* Ω *is a set of* permutations *if for any* $\mathbf{w}, \mathbf{w}' \in \Omega$ *there exists a one-to-one mapping* $\iota : I \to I$ *such that* $\mathbf{w}(i) = \mathbf{w}'(\iota(i))$.

The types were drawn from a set of permutations in Section 2.2.2.

For the result in this subsection, we focus on the case where $|I| = |J|$ and assume that $\nu_{wf} > 0$ and $\phi_{wf} > 0$ for any $w \in W$ and $f \in F$.

A plausible conjecture is that when there are at least as many distinct types of firms as workers, assortative matching identifies worker types from the firm types with which they are matched, and hence incomplete-information stability implies complete-information stability. We now show by example that this is not the case. The matching outcome in Figure 8 illustrates that an incomplete-information stable matching need not be complete-information stable, even though there are equal numbers of worker and firm types, and Ω is a set of permutations. There are 2 types of firms

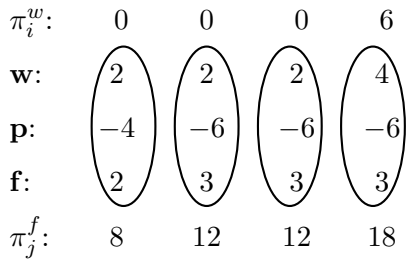

Figure 8: An incomplete information stable matching outcome (when Ω is a set of permutations) that is not complete information stable.

and 2 types of workers. As usual in our examples, $\nu_{wf} = \phi_{wf} = wf$. This is not complete-information stable, as worker 4 and firm 3 can form a blocking pair. In the incomplete-information setting, any transfer at which worker 4 is willing to match with firm 3 also makes worker 2 willing to match with firm 3. Firm 3 thus cannot preclude the possibility that the worker type in a candidate blocking pair is 2, and hence cannot be sure of the profitability of the proposed block. This in turn ensures that the outcome is incomplete-information stable.

The difficulty is that the observables, namely firms' types and transfers, are the same for all firms of type 3. As a result, neither an outside observer who knows only that a firm is type 3, or a different firm of type 3, can ascertain the type of worker with whom the firm is matched.

This difficulty is eliminated if either all firms or all workers have different types (the proof is in Appendix A.6):

Proposition 5 *Suppose Assumptions 1 and 2 hold, and assume Ω is a set of permutations. Incomplete-information stability coincides with complete-information stability if either*

1. *different firms have different types, or*

2. *different workers have different types.*

The first case (different firms have different types) is straightforward, since now (observable) firm types perfectly reveal worker types in an assortative matching. For the second case (different workers have different types), while worker types are not observable, when different workers have different

types, the transfer $\mathbf{p}$, which is observable, is fully informative about worker type regardless of firm types.

5 Discussion

In a world of complete information, it is natural to combine the study of stable matchings with the study of the process by which such matchings are formed. The deferred acceptance algorithm of Gale and Shapley (1962), for example, can be used to construct direct mechanisms with stable equilibrium outcomes.[4] Alternatively, one might consider decentralized procedures. For example, Lauermann and Nöldeke (2012) examine a model in which the members of two populations are continually matched into pairs, with each pair either agreeing to form (and leaving them market) or returning to the unmatched pool. One's intuition (with Lauermann and Nöldeke (2012) identifying the conditions under which this intuition can be made precise) is that as matching frictions become negligible, the equilibrium outcomes of this process should converge to stable outcomes. If not, there exists a pair of agents who would prefer matching with one another rather than settle for their equilibrium outcomes, and the absence of matching frictions suggests that they should be able to find another, vitiating the convergence to an equilibrium outcome that is not stable.

Under incomplete information, the connection between stable matches and the process by which stable matches are formed is less obvious. In the process of encountering others and making and accepting or rejecting matches, the agents are likely to learn about their environment. As a result, the information structure prevailing at the end of the matching process will typically differ from that which characterizing the beginning of the process. Explaining the process leading to a stable matching thus requires specifying the matching mechanics as well as the initial configuration of incomplete information. Our intuition provides few clues as to the relationship between the concluding specification of information, the original information configuration, and the intervening process.

One branch of the literature has responded by focussing on the process

[4]For example, if preferences are strict and the direct mechanism maps announced preferences into the outcome computed via the deferred acceptance algorithm, then it is a dominant strategy in this mechanism for "proposers" to announce their preferences truthfully (Gale and Shapley, 1962). There is no stable matching mechanism under which truthful revelation of preferences is a dominant strategy for all agents (Roth, 1982), but every Nash equilibrium outcome of this deferred-acceptance-based mechanism is stable with respect to the agents' true preferences (Roth, 1984).

by which matches form under incomplete information. For example, one could again consider a direct revelation mechanism in which the announced preferences are inputs to the deferred acceptance algorithm. Rather than considering Nash equilibria, one now examines Bayes Nash equilibria of the incomplete information game. Roth (1989) does this for the case that agents know their own preferences for partners, but do not know potential partners' preferences. He shows that some important qualitative features of the equilibria in complete information do *not* carry over to incomplete information. There exists no mechanism with the property that at least one of its equilibria is always stable with respect to the true preferences at every realization of the game. In other words, any mechanism that might be employed will sometimes result in a match in which there will be an unmatched pair, each of whom knows they would prefer that match to the mechanism's match. Thus even in what would seem to be the simplest extension to incomplete information, in which all agents know the value to them of potential partners, the link between the strategic issue of how matches are formed and the stability of matches is broken.

A number of papers have examined decentralized procedures for forming matches. This work shares with ours the necessity of identifying the inferences agents can draw from the behavior of other agents. Chade (2006) analyzes a model in which agents observe a noisy signal of the true type of any potential mate. In this environment, agents' matching decisions must incorporate not only information about a partner's attribute conveyed by the noisy signal, but also information about a partner's type given their acceptance decision. Chakraborty, Citanna, and Ostrovsky (2010) study a two-sided matching problem with incomplete information and interdependent valuations on one side of the market. They cast their model as one of matching students to colleges when students have complete information about colleges. Colleges care about students' characteristics, but get only noisy signals about those characteristics. Other colleges also get signals about students' characteristics, and as a consequence, the set of offers a student gets conveys information about his or her characteristics. Chakraborty, Citanna, and Ostrovsky (2010) show that when the entire realized matching outcome is publicly observable, stable mechanisms do not generally exist. The instability stems from colleges learning about student qualities from the observable match, *given the mechanism*. In their model, colleges may learn differently under different mechanisms, hence a matching may be stable under some mechanisms but not under others. Their approach is to define stability of matching mechanisms rather than stable matches. We similarly assume in our work that the match is publicly observable, but define stability

for a match without reference to any mechanism from which the matching arose.[5]

Our approach differs from much of the work on matching with incomplete information in that we start with a notion of stability of a match in an incomplete information environment rather than with a process by which matches form. We build into our stability notion the requirement that agents make use of all of the information they infer from the common knowledge that the matching is stable. Our interpretation of this common knowledge is that the agents see a match that persists over time, infer that there are no blocking opportunities, that others also know there are no blocking opportunities, and so on. In a similar spirit, Forges (1994) and Holmström and Myerson (1983) study mechanism design problems with the constraint that the outcome should be free from objections players might make based on information revealed by the mechanism.

We view this notion of stability as a metaphor, capturing necessary conditions for an outcome to be the potential product of a matching process. We do not have an explicit model of origination of the match, nor of the blocking process testing the stability of a match. In particular, our model does not explicitly capture the dynamic and intertemporal considerations that such an explicit model (of either the origination or blocking process) would require.

Our work is most closely related to the literature on the core in incomplete information problems. There are different definitions of the core with incomplete information, each of which is meant to capture the idea that a core outcome should not be subject to objections that coalitions of the agents might raise. Different definitions make different assumptions about what information a coalition might use in evaluating outcomes and formulating objections.

Wilson (1978) proposed two polar cases, the first being that agents could share all information any member of the coalition had and the second being that agents could share only the information that was common knowledge.

[5] A number of other papers study specific dynamic matching games with uncertainty about the valuation of others. Lee (2004) shows that interdependencies in valuations can lead to adverse selection in a college admission problem. Chade, Lewis, and Smith (2011) and Nagypal (2004) analyze college application models when students are uncertain about their own quality and applications are costly. Hoppe, Moldovanu, and Sela (2009) study a model in which agents have private information about their own qualities and are matched assortatively based on costly signals they send. Ehlers and Masso (2007) study mechanisms for matching when preferences are unknown, showing that truth telling is an equilibrium only if every possible preference profile implies a singleton core under complete information.

The first of these ignores the incentive constraints that might inhibit complete sharing, while the second seems overly restrictive about what information might be shared. Dutta and Vohra (2005) consider a middle ground in which coalitions are allowed to coordinate their objections by inferring information from the objection being contemplated. That is, if a coalition contemplates a coordinated objection to a proposed outcome, each agent in the coalition understands that his objection is irrelevant unless all other agents in the coalition agree to the coordinated objection. In essence, agents are able to make "conditional" offers to other agents that have no effect unless the offer is accepted by the other agents.[6] While our analysis is quite similar in spirit, there are important differences in the inferences in Dutta and Vohra (2005) and the inferences in our model. In Dutta and Vohra (2005), the inferences come only from the hypothesis that other agents are willing to participate in a blocking coalition. In our terms, these are "first-round" inferences. In contrast, our model also allows agents to make second round inferences—they may make inferences from the fact that *other* agents do *not* block (which is not the case in Dutta and Vohra (2005)). Our agents continue, making third-round inferences, and fourth-round inferences, and so on. In the end, our agents make all possible inferences consistent with the common knowledge of the stability of the matching. We believe that this feature, distinguishing our work from the literature, is vital to achieving a stability notion that both captures a suitably rich process of information inference and is consistent with existence.

In keeping with our interpretation of stability as characterizing a persisting outcome, we assume that firms know the type of their current partners. In contrast, it is common in the literature to assume that the market contains only unmatched agents, with no distinguished pairs that know one another's identities, and with matched agents leaving the market to be replaced by new agents (as in, for example, Myerson (1995)).

Mailath, Postlewaite, and Samuelson (2012b) examine a model in which a continuum of sellers (the counterpart of firms) and a continuum of buyers (the counterpart of workers) simultaneously invest in attributes (the counterpart of types), and then competitively match, with payoffs determined by premuneration values adjusted by a transfer. Seller attributes are public, while buyer attributes are private. Seller attributes are priced by a *uniform pricing* function that, since buyer attributes are private, is independent of buyer attributes. The equilibrium combines market clearing under uniform pricing with a pairwise stability notion close in spirit to incomplete infor-

[6] See Serrano and Vohra (2007) and Myerson (2007) for similar models.

mation stability. Premuneration values are supermodular and the model incorporates (the continuum counterpart) of worker types being drawn from a set of permutations, with no two workers and no two firms having identical types. In contrast to Proposition 5, however, equilibria in Mailath, Postlewaite, and Samuelson (2012b) need not satisfy complete-information stability. In addition to a collection of modeling and technical differences, it is not assumed in Mailath, Postlewaite, and Samuelson (2012b) that each seller exogenously knows the attribute of the buyer he is matched with, though in equilibrium each seller is able to correctly infer the buyer's attribute. More importantly, sellers in Mailath, Postlewaite, and Samuelson (2012b) cannot identify the current match of buyers appearing as part of potential blocking coalitions.

Our notion of incomplete-information stability allows for deviations by a single pair, but not deviations by larger coalitions. Under complete information, the restriction to pairwise blocking coalitions is innocuous. Any outcome that can be blocked by a coalition can be blocked by a pairwise coalition. Under incomplete information, this is no longer obviously the case. Expanding the analysis beyond pairwise blocking coalitions would first require taking a stand on what inferences agents can draw from the hypothesis that the entire coalition is willing to participate. This would presumably proceed along the lines off our analysis, though the details are nontrivial. Allowing larger blocking coalitions stands high on our list of prospective extensions (along with moving beyond one-to-one matching).

Many of our results will clearly carry over to models that allowed larger deviating coalitions. If the set of allowed coalitions is increased, more outcomes will be blocked, and consequently the set of unblocked outcomes will be (weakly) smaller. However, any plausible stability notion will leave complete-information stable outcomes unblocked. Thus, our results that incomplete-information stable outcomes exist and are a superset of complete-information stable outcomes, as well as that under quite general conditions incomplete-information stable outcomes are efficient, will continue to hold when more coalitions are allowed. Similarly, the equal treatment of equals may still fail, and the continuity of payoffs when there is little asymmetry of information will hold.

Finally, one might think that an auction-like process could mediate the matching in our environment, since auctions are a common mechanism for matching buyers to sellers in one-sided asymmetric information environments. For example, consider the following setting and second-price auction mechanism.

Let $(w_1,\ w_2,\ ...,w_n)$ and $(f_1,\ f_2,\ ...,f_n)$ be vectors of worker and firm

types to be matched, with the firm types being common knowledge and increasing in index, and the worker types being private information. The premuneration value for worker type w_i matched with firm type f_j is $w_i f_j$, as is the firm's premuneration value. Consider a direct revelation mechanism defined as follows. Let $(\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_n)$ be the announced worker types. Denote the k^{th} order statistic of the reports by $\hat{w}_{(k)}$. The direct mechanism matches the lowest announced worker type with the lowest firm, and charges the worker a price of $p_1 = 0$. The second lowest announced worker type is matched with the second lowest firm type and charged $p_2 = \hat{w}_{(1)} \cdot (f_2 - f_1)$, the increase that the lowest worker would have had for matching with this firm. The k^{th} worker is matched with the k^{th} firm and is charged $p_k = \hat{w}_{(k-1)}(f_k - f_{k-1}) + p_{k-1}$, that is, the increase in the payoff to the worker just "beneath" the k^{th}-worker from being matched with firm k rather than firm $k-1$.

It is straightforward to show that it is a dominant strategy for workers to announce their types truthfully. The difficulty is that the this process need not generate stable outcomes. Consider the case in which workers' types are $(1, 2, 3)$, and firms' types are $(1, 1, 1)$. All firms will then be matched with workers at price 0. But notice that the values to the three firms will be $(1, 2, 3)$, since the firms' premuneration values depend on the type of the worker. The combination of firm type 1 and worker type 2 can then form a blocking pair, as can firm type 1 and worker type 3 as well as firm type 2 and worker type 3.

It is an important component of this instability result that sellers' premuneration values are nontrivial functions of buyers' characteristics. There would be no problem with stability if sellers did not care with which buyer they were matched.[7] Interestingly, Google auctions places on webpages to advertisers, thus operating a mechanism that matches buyers (advertisers) and sellers (webpage owners). This auction would generate stable outcomes if sellers received a flat fee for their spots, rendering them indifferent over the buyers with whom they are matched. However, the sellers' total revenue depends on the number of times an ad generates a click, ensuring that the sellers have premuneration values that are nontrivial functions of buyers' characteristics. Presumably, this fee structure reflects buyers' uncertainty about the quality of the web pages over which they are bidding, protecting them from paying high prices for sites the generate little traffic. In the

[7]For example, Edelman, Ostrovsky, and Schwarz (2007) analyze a generalized second price auction, but essentially assume that sellers' premuneration values are the same for all buyers, eliminating the reason why auctions in our framework will often generate outcomes that are not stable.

process, however, the fee structure opens the possibility that the resulting outcome will not be stable.

A Appendix: Proofs

A.1 Proof of Lemma 1

Only statement 3 of the Lemma requires proof, the others being obvious from the definition. Suppose en route to a contradiction that E is self-stabilizing, but its closure, $\overline{E}$, is not. There is then an outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \overline{E}$, a pair of unmatched agents (i, j), and a transfer p' such that

$$\nu_{\mathbf{w}(i),\mathbf{f}(j)} + p' > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \tag{A.1}$$

and

$$\phi_{\mathbf{w}'(i),\mathbf{f}(j)} - p' > \phi_{\mathbf{w}'(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j} \tag{A.2}$$

for all $\mathbf{w}' \in \Omega$ satisfying

$$\left(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}\right) \in \overline{E}, \tag{A.3}$$

$$\mathbf{w}'(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j)), \quad \text{and} \tag{A.4}$$

$$\nu_{\mathbf{w}'(i),\mathbf{f}(j)} + p' > \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}. \tag{A.5}$$

Since Ω is finite, the set of worker-type assignments that satisfy condition (A.2)–(A.5) is unchanged for $p < p'$ but arbitrarily close. Thus, there is a $p'' < p'$ such that for all $p \in (p'', p')$, the lower transfer p also satisfies (A.1) and (A.2)–(A.5) for (i, j).

Let $\mathbf{p}^n \to \mathbf{p}$ be a sequence satisfying $(\mu, \mathbf{p}^n, \mathbf{w}', \mathbf{f}) \in E$ (recall footnote 3). It is then immediate from (A.1) that there exists an N such that, for all $n > N$ and all $p \in (p'', p')$, $\nu_{\mathbf{w}(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}^n_{i,\mu(i)}$.

Since E is self-stabilizing, for all $n > N$, and all $p \in (p'', p')$, there exists $\mathbf{w}' \in \Omega$, such that

$$\phi_{\mathbf{w}'(i),\mathbf{f}(j)} - p \leq \phi_{\mathbf{w}'(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}, \tag{A.6}$$

$$\left(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}\right) \in E, \tag{A.7}$$

$$\mathbf{w}'(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j)), \quad \text{and} \tag{A.8}$$

$$\nu_{\mathbf{w}'(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}. \tag{A.9}$$

Since Ω is finite, there exists $\mathbf{w}'$ such that the above holds for infinitely many $n > N$ and for two values $p_1 < p_2 \in (p'', p')$. This yields the desired

contradiction, since the $\mathbf{w}'$ obtained violates condition (A.2)–(A.5): Taking limits along the implied subsequence, (A.6) implies that the inequality in (A.2) is reversed, while (A.7) and (A.8) replicate (A.3) and (A.4), and the strict inequality in (A.5) holds at p_2.

A.2 Proof of Proposition 2

(1) We first show Σ^∞ contains every self-stabilizing set E. By definition $E \subset \Sigma^0$. Suppose $E \subset \Sigma^{k-1}$, for $k \geq 1$, and $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in E$. Since E is self-stabilizing, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is E-stable, and so is Σ^{k-1}-stable (because Σ^{k-1} is a larger set), and so is in Σ^k by the definition of Σ^k. Induction shows that $E \subset \Sigma^\infty$.

(2) We next argue that Σ^∞ is a self-stabilizing set. Suppose not. By construction, $\Sigma^\infty \subset \Sigma^0$ and so every outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^\infty$ is individually rational. Then, there is an outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^\infty$ that is not Σ^∞-stable. In particular, there is an unmatched pair (i, j) and transfer $p \in \mathbb{R}$ such that (1) and condition (2)–(5) hold for $\Sigma = \Sigma^\infty$. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^k$ is Σ^k-stable for each $k \geq 0$, and $((i, j), p)$ satisfies (1), for $\Sigma = \Sigma^k$ condition (2)–(5) must fail. That is, for each k, $\phi_{\mathbf{w}^k(i),\mathbf{f}(j)} - p \leq \phi_{\mathbf{w}^k(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}$ for some $\mathbf{w}^k$ such that (a) $\left(\mu, \mathbf{p}, \mathbf{w}^k, \mathbf{f}\right) \in \Sigma^k$, (b) $\mathbf{w}^k(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j))$, and (c) $\nu_{\mathbf{w}^k(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}^k(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}$. Since $\mathbf{w}^k$ is drawn from a *finite* set of type vectors, there is a $\mathbf{w}^*$ that appears infinitely often in the sequence $\{\mathbf{w}^k\}_k$. Since Σ^k is a decreasing sequence of sets, and $(\mu, \mathbf{p}, \mathbf{w}^*, \mathbf{f}) \in \Sigma^k$ for infinitely many k, $(\mu, \mathbf{p}, \mathbf{w}^*, \mathbf{f}) \in \cap_{k=1}^\infty \Sigma^k = \Sigma^\infty$. Hence, we conclude that $\phi_{\mathbf{w}^*(i),\mathbf{f}(j)} - p \leq \phi_{\mathbf{w}^*(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}$ where $\mathbf{w}^*$ satisfies (a) $(\mu, \mathbf{p}, \mathbf{w}^*, \mathbf{f}) \in \Sigma^\infty$, (b) $\mathbf{w}^*(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j))$, and (c) $\nu_{\mathbf{w}^*(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}^*(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}$. Thus, condition (2)–(5) fails for $\Sigma = \Sigma^\infty$, the desired contradiction.

(3) We have established that Σ^∞ is the largest self-stabilizing set. Meanwhile, the closure of a self-stabilizing set is self-stabilizing. Hence $\Sigma^\infty = \overline{\Sigma^\infty}$.

A.3 Proof of Lemma 2

Define

$$p^\varepsilon := \nu_{w^* f^*} + p^* - \nu_{w^* f} + \varepsilon, \tag{A.10}$$

where $\varepsilon > 0$ will be determined later. The first required inequality (6) with $p = p^\varepsilon$ is

$$\nu_{wf} + \nu_{w^* f^*} + \varepsilon > \nu_{wf^*} + \nu_{w^* f} \quad \text{for any } w \geq w^*,$$

which is immediate when $w = w^*$. When $w > w^*$, it follows from the assumption of strict supermodularity (since $f > f^*$). Since $(\mu, \mathbf{p})$ is an individually rational matching, $\nu_{w^*f^*} + p^* \geq 0$. Hence for any $w \geq w^*$, $f > f^*$, and p^ε defined in (A.10),

$$\nu_{wf} + p^\varepsilon > \nu_{w^*f} + p^\varepsilon > \nu_{w^*f^*} + p^*,$$

proving (7).

After substituting for $p = p^\varepsilon$ defined in (A.10), the inequality (8) becomes

$$\nu_{wf} + \nu_{w^*f^*} + \varepsilon \leq \nu_{wf^*} + \nu_{w^*f}, \quad \text{for any } w < w^*.$$

For ε sufficiently small, this inequality follows from the assumption of strict supermodularity (since $f^* < f$). Inequalities (6–8) immediately hold for $p \in (\nu_{w^*f^*} + p^* - \nu_{w^*f}, p^\varepsilon]$. The proof for the case that w^* is unmatched is similar. ■

A.4 Proof of Proposition 3

A.4.1 Preliminaries: An Inductive Notion of Assortativity

We first formulate an inductive notion of assortativity. We write the finite set of possible worker and firm types as $W = \{w^1, w^2, ..., w^K\}$ and $F = \{f^1, f^2, ..., f^L\}$, with both w^k and f^ℓ increasing in their indices. To deal with unmatched agents, we introduce the notation $\mathbf{f}(\varnothing) = \mathbf{w}(\varnothing) = \varnothing$, with the conventions $\varnothing < w^k$ and $\varnothing < f^\ell$ for any k and ℓ. The function $\mathbf{f} \circ \mu$ *is weakly comonotone with* $\mathbf{w}$ *on* I' if $\mathbf{f}(\mu(i)) \geq \mathbf{f}(\mu(i'))$ for all $i, i' \in I'$ satisfying $\mathbf{w}(i) > \mathbf{w}(i')$.

Definition A.1 *For* $1 \leq k < K$, *a matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* k^{th}-order worker-assortative *if, for all* $w > w^k$, $\mathbf{f} \circ \mu$ *is weakly comonotone with* $\mathbf{w}$ *on* $I' = \{i : \mathbf{w}(i) \in \{w^1, \ldots, w^k, w\}\}$. *For* $1 \leq \ell < L$, *a matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* ℓ^{th}-order firm-assortative *if, for all* $f > f^\ell$, $\mathbf{w} \circ \mu^{-1}$ *is weakly comonotone with* $\mathbf{f}$ *on* $J' = \{j : \mathbf{f}(j) \in \{f^1, \ldots, f^\ell, f\}\}$. *A matching* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* worker-assortative *if it is* $(K-1)^{th}$*-order worker-assortative; it is* firm-assortative *if it is* $(L-1)^{th}$*-order firm-assortative. A matching* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* assortative *if it is both worker-assortative and firm-assortative.*

Note that the worker-assortativity order is defined in terms of the grand set of all worker types W, not the ex post realized types; similarly for firm-assortativity. For example, if $\mathbf{w}(i) \neq w^1$ for all i, i.e., no worker has

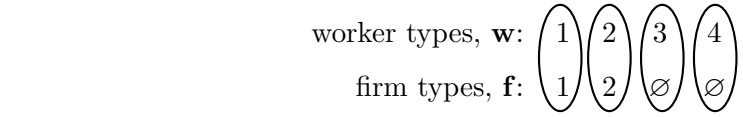

Figure 9: A matching that is 1st-order, but is not 2nd-order worker assortative. There are 4 workers and 2 firms, $W = \{1, 2, 3, 4\}$ and $F = \{1, 2\}$, and workers and firms have different types.

Figure 10: The two nontrivial 2nd-order worker-assortative matchings for the environment of Figure 9. The first is worker assortative, while the second is not.

the lowest possible type w^1, then $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is trivially first-order worker-assortative by definition. In addition, k^{th}-order worker-assortativity requires not only that the k lowest worker types $\{w^1, ..., w^k\}$ are matched with firms assortatively, but also that any workers with a higher type $w > w^k$ are matched with (weakly) higher firm types. For example, the matching in Figure 9 is *not* 2nd-order worker-assortative even though the lowest two worker types are matched assortatively. Figure 10 displays the two nontrivial 2nd-order worker-assortative matchings.

In addition to the first matching displayed in Figure 10, there is a trivial worker-assortative matching in which no worker or firm is matched and the two worker-assortative matchings displayed in Figure 11.

The first matching in Figure 11 is not firm-assortative and hence not assortative. The first matching in Figure 10, the second matching in Figure 11 and the trivial matching in which no worker or firm is matched are both worker- and firm-assortative. Note that our definition of assortative matching does not exclude the case that all agents remain unmatched. By Definition A.1, this matching is 3rd-order worker-assortative and 1st-order firm assortative, and hence assortative. This case is important because if, for example, $\nu_{wf} = \phi_{wf} = -1$, everyone staying unmatched is the only individually rational, and hence the only efficient, matching. But, if $\nu_{wf} = \phi_{wf} = 1$, this assortative matching is not efficient. In fact, it is easy to see that

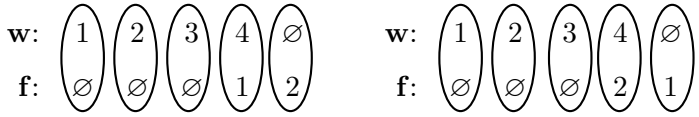

Figure 11: The two other nontrivial worker-assortative matchings for the environment of Figure 9.

any assortative matching can be efficient for the appropriate specification of premuneration values.

The following straightforward observation delineates the difference between assortativity and efficiency (we omit the proof).

Lemma A.1 *Under Assumptions 1 and 2: (a) An efficient matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is assortative. (b) If an assortative matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is not efficient, then either there exists a matched worker-firm pair that generates a negative surplus, i.e., there exists* $i \in I$ *such that* $\mu(i) \in J$ *and* $\nu_{\mathbf{w}(i),\mathbf{f}(\mu^{-1}(i))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu^{-1}(i))} < 0$*; or there exist an unmatched worker and an unmatched firm who could have generated a positive surplus by matching together, i.e., there exist a worker* $i \in I$ *and a firm* $j \in J$ *such that* $\mu(i) = \mu^{-1}(j) = \varnothing$ *and* $\nu_{\mathbf{w}(i),\mathbf{f}(j)} + \phi_{\mathbf{w}(i),\mathbf{f}(j)} > 0$.

The following observation is useful in our inductive proofs.

Lemma A.2 *A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* $(k+1)^{th}$*-order worker assortative if and only if it is* k^{th}*-order worker assortative and for all* $w > w^{k+1}$*,* $\mathbf{f}(\mu)$ *is weakly comonotone with* $\mathbf{w}$ *on* $\{i : \mathbf{w}(i) = w^{k+1}, w\}$*. A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* $(\ell+1)^{th}$*-order firm assortative if and only if it is* ℓ^{th}*-order worker assortative and for all* $f > f^{\ell+1}$*,* $\mathbf{w}(\mu^{-1})$ *is weakly comonotone with* $\mathbf{f}$ *on* $\{j : \mathbf{f}(j) = f^{\ell+1}, f\}$.

Proof. The "only if" parts are immediate by definition. "If": since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is k^{th}-order worker assortative, $\mathbf{f}(\mu)$ is weakly comonotone with $\mathbf{w}$ on $\{i : \mathbf{w}(i) \in \{w^1, \ldots, w^k, w^{k+1}\}\}$. Moreover, for all $w > w^{k+1}$, $\mathbf{f}(\mu)$ is weakly comonotone with $\mathbf{w}$ on $\{i : \mathbf{w}(i) = w^{k+1}, w\}$. If there is a worker i satisfying $\mathbf{w}(i) = w^{k+1}$, then it is immediate that $\mathbf{f}(\mu)$ is weakly comonotone with $\mathbf{w}$ on $\{i : \mathbf{w}(i) \in \{w^1, ..., w^k, w^{k+1}, w\}\}$. Suppose then that $\mathbf{w}(i) \neq w^{k+1}$ for all $i \in I$. Then, for all $w > w^{k+1}$, $\mathbf{f}(\mu)$ is trivially weakly comonotone with $\mathbf{w}$ on $\{i : \mathbf{w}(i) = w^{k+1}, w\}$ for any $\mathbf{f}$. Nonetheless, since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is k^{th}-order worker assortative, we immediately have that

$(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is $(k+1)^{\text{th}}$-order worker assortative, since for all $w > w^{k+1}$, the sets $\{i : \mathbf{w}(i) \in \{w^1, \ldots, w^k, w\}\}$ and $\{i : \mathbf{w}(i) \in \{w^1, \ldots, w^k, w^{k+1}, w\}\}$ agree. The proof for firm assortativity is identical. ■

A.4.2 Completion of The Proof of Proposition 3

Worker-Assortativity. Without loss of generality, assume the true type assignment $(\mathbf{w}, \mathbf{f})$ is such that $\mathbf{w} : \{1, ..., n\} \to W$ and $\mathbf{f} : \{1, ..., n\} \to F$ are weakly increasing. Thus players with lower identities have lower types. (We still need to keep in mind that the firms do not know $\mathbf{w}$.)

We use an induction argument, based on the following two lemmas.

Lemma A.3 *If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, then $(\mu, \mathbf{p})$ is first-order worker assortative under $(\mathbf{w}, \mathbf{f})$.*

Proof. Suppose to the contrary that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$ yet it is not first-order worker assortative. Then by definition, $\mathbf{f}(\mu)$ is not comonotone with $\mathbf{w}$ on $\{i : \mathbf{w}(i) \in \{w^1, w\}\}$ for some $w > w^1$. That is, there exist two workers, say 1 and 2, such that $\mathbf{w}(2) > \mathbf{w}(1) = w^1$ but $\mathbf{f}(\mu((2)) < \mathbf{f}(\mu((1)) \neq \varnothing$.

Claim A.1 *If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, then $\mu(2) \neq \varnothing$.*

> **Proof.** Suppose not, i.e., $\mu(2) = \varnothing$. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^0$, worker 1's individual rationality in the matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ implies that the payoff of firm $\mu(1)$ in this matching outcome is bounded above by the total surplus generated, $\pi^f_{\mu(1)} \leq \nu_{w^1, \mathbf{f}(\mu(1))} + \phi_{w^1, \mathbf{f}(\mu(1))}$.
>
> Consider the worker-firm pair $(2, \mu(1))$ with a transfer p. By Lemma 2 (taking $w^* = \mathbf{w}(2)$ and $f = \mathbf{f}(\mu)(1)$), there exists $\varepsilon > 0$ such that for $-\nu_{\mathbf{w}(2), \mathbf{f}(\mu(1))} < p \leq -\nu_{\mathbf{w}(2), \mathbf{f}(\mu(1))} + \varepsilon$,
>
> $$\nu_{w, \mathbf{f}(\mu)(1)} + p > 0, \qquad \text{for any } w \geq \mathbf{w}(2), \text{ and} \tag{A.11}$$
> $$\nu_{w, \mathbf{f}(\mu)(1)} + p \leq 0, \qquad \text{for any } w < \mathbf{w}(2). \tag{A.12}$$
>
> Worker 2 is better off because he gets a positive payoff, from (A.11); firm $\mu(1)$ will assign probability 1 that the deviating worker's type is at least $\mathbf{w}(2)$ because of (A.11) and (A.12). Hence, the expected payoff of firm $\mu(1)$ in this deviation is bounded below by $\phi_{\mathbf{w}(2), \mathbf{f}(\mu(1))} - p$. By taking p close to $-\nu_{\mathbf{w}(2), \mathbf{f}(\mu)(1)}$, this lower bound can be made arbitrarily close to $\phi_{\mathbf{w}(2), \mathbf{f}(\mu(1))} + \nu_{\mathbf{w}(2), \mathbf{f}(\mu(1))}$. Since $\mathbf{w}(2) > w^1$, strict supermodularity implies that $\phi_{\mathbf{w}(2), \mathbf{f}(\mu(1))} + \nu_{\mathbf{w}(2), \mathbf{f}(\mu(1))} > \nu_{w^1, \mathbf{f}(\mu(1))} +$

$\phi_{w^1,\mathbf{f}(\mu(1))} \geq \pi^f_{\mu(1)}$. Hence $(2, \mu(1))$ forms a blocking pair with price p. This contradicts the assumption that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$. □

Claim A.2 *If* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, $\mathbf{w}(2) > \mathbf{w}(1) = w^1$, *and* $\mathbf{f}(\mu(2)) < \mathbf{f}(\mu(1)) \neq \varnothing$, *then*

$$\phi_{\mathbf{w}(2),\mathbf{f}(\mu(1))} + \nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))} \leq \nu_{\mathbf{w}(2),\mathbf{f}(\mu(2))} + \mathbf{p}_{2,\mu(2)} + \phi_{w^1,\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)}. \tag{A.13}$$

Proof. By Lemma 2, worker 2 can credibly reveal his type being at least $\mathbf{w}(2)$ by approaching firm $\mu(1)$ with a transfer

$$p = \nu_{\mathbf{w}(2),\mathbf{f}(\mu(2))} + \mathbf{p}_{2,\mu(2)} - \nu_{\mathbf{w}(2),\mathbf{f}(\mu)(1)} + \varepsilon,$$

for some small $\varepsilon > 0$.

Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, firm $\mu(1)$ rejects worker 2 and p. That is, the firm would be worse off in the new match. Hence,

$$\phi_{w,\mathbf{f}(\mu(1))} - p \leq \phi_{w^1,\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)} \text{ for some } w \geq \mathbf{w}(2). \tag{A.14}$$

Since ϕ_{wf} is increasing in w and f, the statement in (A.14) holds if and only if

$$\phi_{\mathbf{w}(2),\mathbf{f}(\mu(1))} - p \leq \phi_{w^1,\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)}.$$

Substituting for p,

$$\phi_{\mathbf{w}(2),\mathbf{f}(\mu(1))} - \left(\nu_{\mathbf{w}(2),\mathbf{f}(\mu(2))} + \mathbf{p}_{2,\mu(2)} - \nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))} + \varepsilon\right) \leq \phi_{w^1,\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)},$$

implying (A.13). □

Claim A.3 *If* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, $\mathbf{w}(2) > \mathbf{w}(1) = w^1$, *and* $\mathbf{f}(\mu((2)) < \mathbf{f}(\mu(1)) \neq \varnothing$, *then*

$$\nu_{w^1,\mathbf{f}(\mu(2))} + \phi_{w^1,\mathbf{f}(\mu(2))} \leq \left(\nu_{w^1,\mathbf{f}(\mu(1))} + \mathbf{p}_{1,\mu(1)}\right) + \left(\phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}\right). \tag{A.15}$$

Proof. If the inequality in (A.15) did not hold, we can find $q \in \mathbb{R}$ such that

$$\nu_{w^1,\mathbf{f}(\mu(2))} + q > \nu_{w^1,\mathbf{f}(\mu(1))} + \mathbf{p}_{1,\mu(1)} \quad \text{and} \tag{A.16}$$

$$\phi_{w^1,\mathbf{f}(\mu(2))} - q > \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}. \tag{A.17}$$

Since b is increasing and w^1 is the smallest type, (A.17) implies that

$$\min_{w \in W} \phi_{w,\mathbf{f}(\mu(2))} - q > \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}. \tag{A.18}$$

Hence, (A.16) and (A.18) imply $(1, \mu(2))$ is a blocking pair, contradicting $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$. □

Finally, we combine Claims A.2 and A.3. Adding the two inequalities, we obtain

$$\begin{aligned}\left(\nu_{w^1,\mathbf{f}(\mu(2))} + \nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))}\right) + \left(\phi_{w^1,\mathbf{f}(\mu(2))} + \phi_{\mathbf{w}(2),\mathbf{f}(\mu(1))}\right) \\ \leq \left(\nu_{w^1,\mathbf{f}(\mu(1))} + \nu_{\mathbf{w}(2),\mathbf{f}(\mu(2))}\right) + \left(\phi_{w^1,\mathbf{f}(\mu(1))} + \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))}\right).\end{aligned}$$

Recalling that $w^1 < \mathbf{w}(2)$ and $\mathbf{f}(\mu(2)) < \mathbf{f}(\mu(1))$, this inequality contradicts strict supermodularity. ■

Lemma A.4 *For any $k \geq 1$, if $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^k$, then $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is k^{th}-order worker assortative.*

Proof. We proceed by induction. Suppose the claim holds for $k \geq 1$ (from Lemma A.3, the claim holds for $k = 1$). Suppose to the contrary that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1}$, and $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is not $(k+1)^{\text{th}}$-order worker assortative. There then exist two workers $i < i'$ such that worker i's type is $\mathbf{w}(i) = w^{k+1} < \mathbf{w}(i')$ and $\mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i'))$. The proof of Claim A.1 shows (with obvious modifications) that $\mu(i') \neq \varnothing$.

Claim A.4 *If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1}$, $\mathbf{w}(i) = w^{k+1} < \mathbf{w}(i')$ and $\mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i'))$, then*

$$\nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} + \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} \leq \nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} + \mathbf{p}_{i',\mu(i')} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)}. \tag{A.19}$$

Proof. By Lemma 2, worker i' can credibly and profitably reveal his type as being at least $\mathbf{w}(i')$ by approaching firm $\mu(i)$ with a transfer

$$p = \nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} + \mathbf{p}_{i',\mu(i')} - \nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} + \varepsilon,$$

for some small $\varepsilon > 0$. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1}$, $(i', \mu(i))$ together with p cannot make firm $\mu(i)$ better off for any consistent belief. Hence, there exists $w \geq \mathbf{w}(i')$ such that

$$\phi_{w,\mathbf{f}(\mu(i))} - p \leq \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)}.$$

By monotonicity of b and $\mathbf{w}(i) < \mathbf{w}(i') \leq w$, we have

$$\phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} - p \leq \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)}.$$

Substituting for p, we get (A.19). □

Claim A.5 *If* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1}$, $\mathbf{w}(i) = w^{k+1} < \mathbf{w}(i')$ *and* $\mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i'))$, *then*

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} \leq \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} + \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \tag{A.20}$$

Proof. Suppose to the contrary that the claimed inequality does not hold. We can then find $q \in \mathbb{R}$ such that

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} + q > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \quad \text{and} \tag{A.21}$$

$$\phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \tag{A.22}$$

By monotonicity of b, (A.22) implies

$$\phi_{w,\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')} \text{ for all } w \geq \mathbf{w}(i) = w^{k+1}. \tag{A.23}$$

By the induction hypothesis, Σ^k only contains outcomes that are k^{th}-order worker assortative. Consider the following set of worker type assignments:

$$\Omega' = \Big\{ \mathbf{w}' \in \Omega : \left(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}\right) \in \Sigma^k, \mathbf{w}'\left(i'\right) = \mathbf{w}\left(i'\right), \\ \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i'))} + q > \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \Big\}.$$

For any $\mathbf{w}' \in \Omega'$, we have $\mathbf{w}'(i) \geq w^{k+1}$. To see this, suppose to the contrary that $\mathbf{w}'(i) \leq w^k$. By assumption, $\mathbf{w}'(i') = \mathbf{w}(i') > w^{k+1} > w^k$ and $\mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i'))$. But then $\mathbf{w}'(i') > \mathbf{w}'(i)$, while $\mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i'))$, and so $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f})$ is not k^{th}-order worker assortative, contradicting the assumption that $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \Sigma^k$.

It then follows from (A.23) that

$$\min_{\mathbf{w}' \in \Omega'} \phi_{\mathbf{w}'(i),\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \tag{A.24}$$

Hence, from (A.21) and (A.24), the unmatched pair $(i, \mu(i'))$ at transfer q can form a blocking pair. A contradiction. □

Summing (A.19) and (A.20), we have

$$\begin{aligned}&\left(\nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i))}+\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))}\right)+\left(\phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i))}+\phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))}\right)\\&\qquad\leq\left(\nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))}+\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))}\right)+\left(\phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))}+\phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))}\right),\end{aligned}$$

contradicting strict supermodularity. ■

Assortativity.

Lemma A.5 *If* $(\mu,\mathbf{p},\mathbf{w},\mathbf{f})\in\Sigma^{\infty}$, *then it is assortative.*

Proof. From Lemmas A.3 and A.4, we have the worker assortativity of $(\mu,\mathbf{p},\mathbf{w},\mathbf{f})$. If it is not firm assortative, then we can find two firms with different types, say firms j and j' with $\mathbf{f}(j)<\mathbf{f}(j')$, such that $\mathbf{w}(\mu^{-1}(j))>\mathbf{w}(\mu^{-1}(j'))$. If $\mu^{-1}(j')\neq\varnothing$, worker assortativity is violated. Hence, $\mu^{-1}(j')=\varnothing$.

From Lemma 2 (taking $w^*=\mathbf{w}(\mu^{-1}(j))$, $f^*=\mathbf{f}(j)$, $f=\mathbf{f}(j')$, and $p^*=\mathbf{p}_{\mu^{-1}(j),j}$), if worker $\mu^{-1}(j)$ proposes a price $p=\nu_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}+\mathbf{p}_{\mu^{-1}(j),j}-\nu_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')}+\varepsilon$ for small $\varepsilon>0$ to firm j', and if firm j' accepts the proposal, then worker $\mu^{-1}(j)$ is better off and can prove that his type is at least $\mathbf{w}(\mu^{-1}(j))$. That is,

$$\nu_{w,\mathbf{f}(j')}+p>\nu_{w,\mathbf{f}(j)}+\mathbf{p}_{\mu^{-1}(j),j},\qquad\text{for any } w\geq\mathbf{w}(\mu^{-1}(j)),\tag{A.25}$$

$$\nu_{w,\mathbf{f}(j')}+p\geq 0,\qquad\text{for any } w\geq\mathbf{w}(\mu^{-1}(j)),\tag{A.26}$$

$$\nu_{w,\mathbf{f}(j')}+p\leq\nu_{w,\mathbf{f}(j)}+\mathbf{p}_{\mu^{-1}(j),j},\qquad\text{for any } w<\mathbf{w}(\mu^{-1}(j)).\tag{A.27}$$

It remains to argue that firm j' indeed finds it profitable to accept this proposal (so that $\left(\mu^{-1}(j),j'\right)$ can form a blocking pair). Since ϕ_{wf} is strictly increasing in f, we can choose ε such that $0<\varepsilon<\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')}-\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}$. As worker $\mu^{-1}(j)$ can credibly signal that his type is at least $\mathbf{w}(\mu^{-1}(j))$, the payoff to firm j' from this proposal is bounded below by

$$\begin{aligned}&\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')}-p\\&\quad=\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')}-\nu_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}-\mathbf{p}_{\mu^{-1}(j),j}+\nu_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')}-\varepsilon\\&\quad>\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}-\nu_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}-\mathbf{p}_{\mu^{-1}(j),j}+\nu_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')}\\&\quad>\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}-\mathbf{p}_{\mu^{-1}(j),j}\\&\quad\geq 0,\end{aligned}$$

where the three inequalities follow from substituting the upper bound of ε, the monotonicity of ν_{wf}, and the individual rationality of the candidate matching.

Efficiency. Suppose $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{\infty}$ is not efficient. Then by Lemma A.1, there are two cases.

(1) There exists $i \in I$ such that $\mu(i) \in J$ and $\nu_{\mathbf{w}(i),\mathbf{f}(\mu^{-1}(i))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu^{-1}(i))} < 0$. This clearly violates individual rationality.

(2) There exist a worker $i \in I$ and a firm $j \in J$ such that $\mu(i) = \mu^{-1}(j) = \varnothing$ and $\nu_{\mathbf{w}(i),\mathbf{f}(j)} + \phi_{\mathbf{w}(i),\mathbf{f}(j)} > 0$. In this case, suppose worker i proposes a transfer of $p = -\nu_{\mathbf{w}(i),\mathbf{f}(j)} + \varepsilon$ to firm j, where $\varepsilon > 0$ is to be determined later. Hence $\nu_{\mathbf{w}(i),\mathbf{f}(j)} + p = \varepsilon > 0$. If $\mathbf{w}(i)$ is the lowest worker type among W, then this transfer will make both the worker and firm unambiguously better off if $\varepsilon < \nu_{\mathbf{w}(i),\mathbf{f}(j)} + \phi_{\mathbf{w}(i),\mathbf{f}(j)}$. If $\mathbf{w}(i)$ is not the lowest type, take $\varepsilon < \min\left\{\phi_{\mathbf{w}(i),\mathbf{f}(j)} + \nu_{\mathbf{w}(i),\mathbf{f}(j)}, \nu_{\mathbf{w}(i),\mathbf{f}(j)} - \max_{w<\mathbf{w}(i)} \nu_{w,\mathbf{f}(j)}\right\}$. By monotonicity, for any $w < \mathbf{w}(i)$,

$$\begin{aligned}
\nu_{w,\mathbf{f}(j)} + p &= \nu_{w,\mathbf{f}(j)} - \nu_{\mathbf{w}(i),\mathbf{f}(j)} + \varepsilon \\
&< \nu_{w,\mathbf{f}(j)} - \nu_{\mathbf{w}(i),\mathbf{f}(j)} + \nu_{\mathbf{w}(i),\mathbf{f}(j)} - \max_{w<\mathbf{w}(i)} \nu_{w,\mathbf{f}(j)} \\
&= \nu_{w,\mathbf{f}(j)} - \max_{w<\mathbf{w}(i)} \nu_{w,\mathbf{f}(j)} \\
&\leq 0.
\end{aligned}$$

So firm j will believe the worker has a type of at least $\mathbf{w}(i)$, and will expect a payoff bounded below by

$$\phi_{\mathbf{w}(i),\mathbf{f}(j)} - p = \phi_{\mathbf{w}(i),\mathbf{f}(j)} + \nu_{\mathbf{w}(i),\mathbf{f}(j)} - \varepsilon > 0.$$

Hence, (i, j) form a blocking pair.

This completes the proof of Proposition 3. ■

A.5 Proof of Proposition 4

Suppose to the contrary, that there exists $\varepsilon > 0$ such that for any integer $n > 0$, there exists $\Omega^n \subset \xi_{\frac{1}{n}}(\mathbf{w})$ and $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, f) \in \Sigma^{\infty}(\Omega^n)$ such that $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, f) \notin \xi_{\varepsilon}(\pi(\Sigma^{\infty}(\{\mathbf{w}\})))$.

We denote by $||\cdot||$ the Euclidean metric. Notice that $\|\mathbf{w}^n - \mathbf{w}\| \to 0$ as $n \to \infty$. Hence, the boundedness of $\{\mathbf{w}^n\}$ and the individual rationality of $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, f) \in \Sigma^{\infty}(\Omega^n)$ imply that the sequence $\{\mathbf{p}^n\}$ is bounded. Notice also that $\|(\mu^n, \mathbf{p}^n, \mathbf{w}, f) - (\mu^n, \mathbf{p}^n, \mathbf{w}^n, f)\| \to 0$ as $n \to \infty$. Since $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, f) \notin \xi_{\varepsilon}(\pi(\Sigma^{\infty}(\{\mathbf{w}\})))$, it follows that for sufficiently large n,

$$\pi(\mu^n, \mathbf{p}^n, \mathbf{w}, \mathbf{f}) \notin \xi_{\frac{\varepsilon}{2}}(\pi(\Sigma^{\infty}(\{\mathbf{w}\}))). \tag{A.28}$$

Since there is a finite number of possible matchings, at least one (denoted μ) appears infinitely often in the sequence. Taking a subsequence if necessary, we may assume μ^n is constant, equal to μ, and $\mathbf{p}^n$ converges to some limit, denoted $\mathbf{p}$. We then have from (A.28) that $\pi(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \notin \pi(\Sigma^\infty(\{\mathbf{w}\}))$, that is, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is not complete information stable.

Since individual rationality is satisfied along the sequence, it is trivially satisfied in the limit. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is not complete information stable, there is a pair (i, j) together with a price $p \in \mathbb{R}$ such that

$$\begin{aligned} \nu_{\mathbf{w}(i),\mathbf{f}(j)} + p &> \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}, \\ \phi_{\mathbf{w}(i),\mathbf{f}(j)} - p &> \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)}. \end{aligned}$$

Then by continuity there exists integer $N_1 > 0$ such that

$$\begin{aligned} \nu_{\mathbf{w}(i),\mathbf{f}(j)} + p &> \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}, \\ \phi_{\mathbf{w}'(i),\mathbf{f}(j)} - p &> \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)} \text{ for any } \mathbf{w}' \in \xi_{\frac{1}{N_1}}(\mathbf{w}). \end{aligned}$$

Further by continuity, there exists integer $N_2 > 0$ such that if $n > N_2$,

$$\begin{aligned} \nu_{\mathbf{w}^n(i),\mathbf{f}(j)} + p &> \nu_{\mathbf{w}^n(i),\mathbf{f}(\mu^n(i))} + \mathbf{p}^n_{i,\mu^n(i)}, \\ \phi_{\mathbf{w}'(i),\mathbf{f}(j)} - p &> \phi_{\mathbf{w}^n(i),\mathbf{f}(\mu^n(i))} - \mathbf{p}^n_{i,\mu^n(i)} \text{ for any } \mathbf{w}' \in \xi_{\frac{1}{N_1}}(\mathbf{w}). \end{aligned}$$

Takes $n > \max\{N_1, N_2\}$. Then,

$$\begin{aligned} \nu_{\mathbf{w}^n(i),\mathbf{f}(j)} + p &> \nu_{\mathbf{w}^n(i),\mathbf{f}(\mu^n(i))} + \mathbf{p}^n_{i,\mu^n(i)}, \\ \phi_{\mathbf{w}'(i),\mathbf{f}(j)} - p &> \phi_{\mathbf{w}^n(i),\mathbf{f}(\mu^n(i))} - \mathbf{p}^n_{i,\mu^n(i)} \text{ for any } \mathbf{w}' \in \Omega^n \subset \xi_{\frac{1}{N_1}}(\mathbf{w}). \end{aligned}$$

Therefore, $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, f) \notin \Sigma^\infty(\Omega^n)$. A contradiction. ■

A.6 Proof of Proposition 5

If different firms have different types, then from Lemma A.5, the worker type assignment $\mathbf{w}$ is common knowledge, and incomplete information stability trivially coincides with complete information stability.

Suppose different workers have different types and several firms share the same type f. Write $\mathbf{f}^{-1}(f)$ as this set of firms. We claim that $\mathbf{p}_{\mu^{-1}(j),j}$ is different for each $j \in \mathbf{f}^{-1}(f)$.

Define

$$j_1 = \arg\min_{j \in \mathbf{f}^{-1}(f)} \mathbf{w}(\mu^{-1}(j)),$$

and for $1 < k \leq \left|\mathbf{f}^{-1}(f)\right|$,

$$j_k = \underset{j \in \mathbf{f}^{-1}(f) \setminus \{j_1, \ldots, j_{k-1}\}}{\arg\min} \mathbf{w}(\mu^{-1}(j)).$$

Note that because no two workers have the same type, firm j_k knows the ranking of worker $\mu^{-1}(j_k)$: the worker $\mu^{-1}(j_k)$ is the k^{th} worst among those who match with some firm in the set $\mathbf{f}^{-1}(f)$. Firm j_k's profit is

$$\pi_{j_k} = \phi_{\mathbf{w}(\mu^{-1}(j_k)),\mathbf{f}(j_k)} - \mathbf{p}_{\mu^{-1}(j_k),j_k}.$$

We proceed by induction.

Step 1. $\mathbf{p}_{\mu^{-1}(j_1),j_1} < \mathbf{p}_{\mu^{-1}(j_k),j_k}$ for any $k > 1$.

Suppose to the contrary $\mathbf{p}_{\mu^{-1}(j_1),j_1} \geq \mathbf{p}_{\mu^{-1}(j_k),j_k}$ for some $k > 1$. Then $\pi_{j_1} < \pi_{j_k}$ because b is strictly supermodular and firm j_1 is matched with a strictly worse worker type than firm j_k. Then $\left(\mu^{-1}(j_k), j_1\right)$ can form a blocking pair with a transfer $\mathbf{p}_{\mu^{-1}(j_k),j_k} + \varepsilon$, a contradiction.

Step 2. Fix k and assume for the purpose of induction that for some $\ell' < k-1$, $\mathbf{p}_{\mu^{-1}(j_\ell),j_\ell} < \mathbf{p}_{\mu^{-1}(j_k),j_k}$ for any $1 \leq \ell \leq \ell'$. Therefore, everyone knows that the subset of firms in $\mathbf{f}^{-1}(f)$ who are matched with the worst ℓ' workers have the lowest ℓ' transfers. Suppose $\mathbf{p}_{\mu^{-1}\left(j_{\ell'+1}\right),j_{\ell'+1}} \geq \mathbf{p}_{\mu^{-1}(j_k),j_k}$. Then $\left(\mu^{-1}(j_k), j_{\ell'+1}\right)$ with transfer $\mathbf{p}_{\mu^{-1}(j_k),j_k} + \varepsilon$ form a blocking pair. Therefore, $\mathbf{p}_{\mu^{-1}\left(j_{\ell'+1}\right),j_{\ell'+1}} < \mathbf{p}_{\mu^{-1}(j_k),j_k}$.

Step 3. The induction argument in the first two steps establishes that $\mathbf{p}_{\mu^{-1}(j_k),j_k}$ is strictly increasing in k. Therefore, firms know that high type workers get strictly higher transfers from the set $\mathbf{f}^{-1}(f)$ in an incomplete information stable matching, and hence there is no uncertainty about the types of workers employed by $\mathbf{f}^{-1}(f)$.

Hence once again worker type assignments are common knowledge. ■

References

Becker, G. S. (1973): "A Theory of Marriage; Part I," *Journal of Political Economy*, 81(4), 813–846.

Bernheim, B. D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

Chade, H. (2006): "Matching with Noise and the Acceptance Curse," *Journal of Economic Theory*, 129(1), 81–113.

Chade, H., G. Lewis, and L. Smith (2011): "Student Portfolios and the College Admissions Problem," Aizona State University, Harvard University, and University of Wisconsin.

Chakraborty, A., A. Citanna, and M. Ostrovsky (2010): "Two-Sided Matching with Interdependent Values," *Journal of Economic Theory*, 145(1), 85–105.

Crawford, V. P., and E. M. Knoer (1981): "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, 49(2), 437–450.

Dutta, B., and R. Vohra (2005): "Incomplete Information, Credibility and the Core," *Mathematical Social Sciences*, 50(2), 148–165.

Edelman, B., M. Ostrovsky, and M. Schwarz (2007): "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 1(97), 242–259.

Ehlers, L., and J. Masso (2007): "Incomplete Information and Singleton Cores in Matching Markets," *Journal of Economic Theory*, 1(36), 587–600.

Forges, F. (1994): "Posterior Efficiency," *Games and Economic Behavior*, 6(2), 238–261.

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69(1), 9–15.

Geanakoplos, J. (1994): "Common Knowledge," in *Handbook of Game Theory with Economic Applications, Volume 2*, ed. by R. J. Aumann, and S. Hart, pp. 1437–1496. North Holland.

Holmström, B., and R. B. Myerson (1983): "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51(6), 1799–1819.

Hoppe, H. C., B. Moldovanu, and A. Sela (2009): "The Theory of Assortative Matching based on Costly Signals," *Review of Economic Studies*, 76(1), 253–281.

Lauermann, S., and G. Nöldeke (2012): "Stable Marriages and Search Frictions," Mimeo.

Lee, S.-H. (2004): "Early Admission Program: Does It Hurt Efficiency?," Ph.D. thesis, University of Pennsylvania, Ch. 1.

Mailath, G. J., A. Postlewaite, and L. Samuelson (2012a): "Premuneration Values and Investments in Matching Markets," PIER Working Paper No. 12-008, University of Pennsylvania.

——— (2012b): "Pricing and Investments in Matching Markets," *Theoretical Economics*, forthcoming.

Milgrom, P. R., and N. Stokey (1982): "Information, Trade, and Common Knowledge," *Journal of Economic Theory*, 26, 17–27.

Myerson, R. B. (1995): "Sustainable Matching Plans with Adverse Selection," *Games and Economic Behavior*, 9(1), 35–65.

——— (2007): "Virtual Utility and the Core for Games with Incomplete Information," *Journal of Economic Theory*, 136(1), 260–285.

Nagypal, E. (2004): "Optimal Application Behavior with Incomplete Information," Mimeo.

Pearce, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–50.

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, (4), 617–628.

——— (1984): "Misrepresentation and Stability in the Marriage Problem," *Journal of Economic Theory*, 34(2), 383–387.

——— (1989): "Two-Sided Matching with Incomplete Information about Others' Preferences," *Games and Economic Behavior*, 1(2), 191–209.

Roth, A. E., and M. A. O. Sotomayer (1990): *Two-Sided Matching*. Cambridge University Press, Cambridge.

Serrano, R., and R. Vohra (2007): "Information Transmission in Coalitional Voting Games," *Journal of Economic Theory*, 134(1), 117–137.

Shapley, L. S., and M. Shubik (1971): "The Assignment Game I: The Core," *International Journal of Game Theory*, 1(1), 111–130.

Wilson, R. (1978): "Information, Efficiency, and the Core of an Economy," *Econometrica*, 46(4), 807–816.