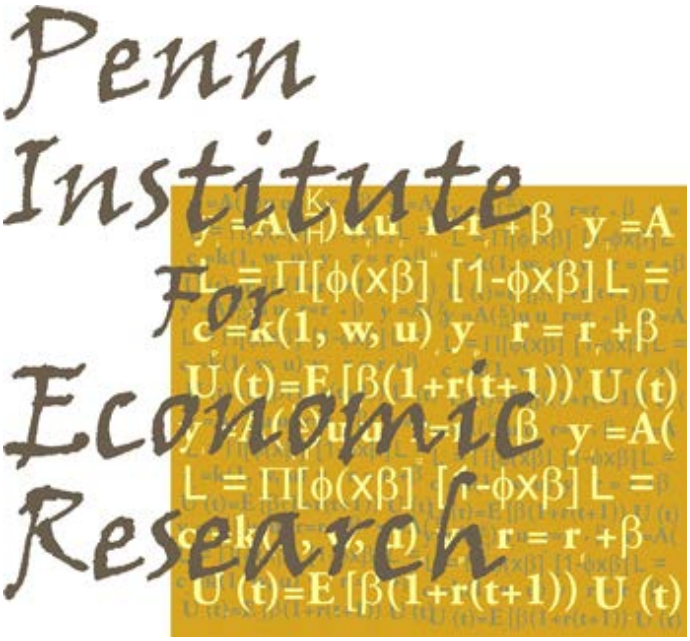

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 13-028

"Stable Matching with Incomplete Information"
Second Version

by

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

http://ssrn.com/abstract=2283156

Stable 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

June 17, 2013

Abstract 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. 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 then provide sufficient conditions for incomplete-information stable matchings to be efficient. Lastly, we define a notion of price-sustainable allocations and show that the set of incomplete-information stable matchings is a subset of the set of such allocations.

*We thank Matt Jackson, three referees, Yeon-Koo Che, Prajit Dutta, Nicole Immorlica, Fuhito Kojima, Dilip Mookherjee, Andrea Prat, Bernard Salanie, Roberto Serrano, and Rajiv Vohra for helpful comments and suggestions.

†We thank the National Science Foundation (grants SES-0961540 and SES-1153893) for financial support.

Contents

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 undergraduates to universities, 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 payment from the firm to the worker, is *stable* if there is no unmatched worker-firm pair that could increase both their payoffs by matching with each other and making an appropriate payment.

The assumption of complete information makes the analysis tractable but is stringent.[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, addressing the following questions. What does it mean for an outcome to be stable under incomplete information? What are the properties of stable outcomes? To what extent does the introduction of asymmetric information in a matching problem alter equilibrium outcomes?

1.1 Beliefs

Our first order of business is to formulate an appropriate modification of stability for problems in which there is asymmetric information. The key to our stability notion is a specification of the beliefs of the agents who might block a candidate stable allocation.

Consider a worker/firm matching problem in which each worker and each firm has a type that is their "quality," and any matched worker-firm pair can generate a surplus that is increasing in both qualities. Suppose that firms' qualities are commonly known, but workers' qualities are not. Each firm knows the quality of the worker she is matched with, and knows the payments 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 how does the firm estimate her payoff when deviating to match with a worker whose quality is unknown?

[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.

What beliefs should she use in calculating her expected payoffs?

We begin by identifying the beliefs the firm can *exclude*, given her knowledge of the allocation and the hypothesis that this allocation is not blocked. In particular, the firm may make inferences about workers' types from the lack of worker-firm pairs wishing to block. These inferences may lead to yet further inferences. We construct an iterative belief-formation process, reminiscent of rationalizability, that captures all such inferences the firm can make. This in general gives rise to a set of "reasonable" beliefs for the firm. We then say that an allocation fails to be stable if some worker-firm pair has a deviation that is profitable, under *any* reasonable belief the firm might have.

In motivating this final step we must distinguish between the viewpoint of the firm and that of the analyst. The firm has some particular belief, drawn from the set of reasonable beliefs, and will participate in a blocking match if it gives her an expected payoff gain, given those beliefs. However, nothing in the structure of the economy or the candidate stable allocation gives the analyst any clue as to what the firm's (reasonable) belief might be. Our goal is to identify the necessary conditions for stability that follow only from the structure of the economy and the hypothesis of stability, and we accordingly reject an allocation only if we are certain there is a successful block. One might seek sharper predictions by augmenting our model with a theory of how firms form beliefs, just as one might impose additional structure to choose between multiple core or complete-information stable outcomes in other circumstances, but we view this as a subsequent exercise.

When is our analysis applicable, or equivalently, how does an allocation and its immunity to blocking come to be commonly known? Our view is that a stable allocation is one that we should expect to persist, and we thus think of the agents in the model repeatedly observing this allocation. Each time they observe the allocation,[3] they can draw further inferences about its properties—first that it is individually rational, then that everyone knows it is individually rational and (given this knowledge) there are no blocking pairs, then that everyone knows that ..., and so on. Each observation corresponds to a step in our iterative belief process, with successive beliefs pushing the agents closer to common knowledge of an allocation's immunity to blocking. A finite number of rounds suffices to determine which allocations are commonly known to be immune to blocking.

[3]We assume, as do Chakraborty, Citanna, and Ostrovsky (2010) do, that the entire allocation is observable.

1.2 Necessary Conditions

We do not address *how* stable outcomes might arise. We view our incomplete-information stability concept as being applied to identify the set of possible incomplete-information stable outcomes, regardless of how they might arise, in much the same way that one studies direct mechanisms to identify the set of implementable outcomes in a mechanism design context. Identifying which stable outcome will appear requires additional institutional information, just as identifying which outcome will be implemented typically requires information about the actual indirect mechanism.

This approach to incomplete-information stability is reminiscent of the study of the core, which (following its formalization by Gillies (1959)) was long used to identify candidates for stability before processes were identified that would reliably lead to core outcomes (e.g., Perry and Reny (1994)). In contrast, the notion of stability and a centralized algorithm for computing stable allocations appeared simultaneously in the study of complete-information matching (Gale and Shapley (1962)). One can imagine a decentralized process that would seemingly lead to stable outcomes under complete information—unmatched agents randomly meet each other and make proposals, with the process stopping when no unmatched pair can improve on their situation by matching—but Lauermann and Nöldeke (2012) demonstrate that only under restrictive assumptions do the obvious such processes lead to stable outcomes (see Section 6.2). Under incomplete information, the outcome of such a process is even less obvious, 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 decentralized foundations for both complete and incomplete-information stable matchings is an open and obviously interesting problem. We return to this issue in Section 6.

Our notion of stability precludes profitable pairwise deviations, but does not consider deviations by larger groups of agents. Under complete information, pairs can block any outcome blocked by larger coalitions (i.e., the set of pairwise stable outcomes coincides with the core), and hence restricting attention to pairwise stability sacrifices no generality. This 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 at their disposal than does any single pair—"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. In many circumstances, we view it as reasonable that the obvious potential blocking coalitions are pairs. We readily imagine a worker seeking a new job or a firm trying to poach a worker, but less readily imagine a set of firms entering into an agreement to reallocate their workers among themselves. Moreover, allowing only pairs to deviate avoids the information-sharing difficulties that would arise with larger coalitions.

1.3 Preview

Sections 2 and 3 develop our stability concept for matching problems with incomplete information. Under general conditions, incomplete-information stable outcomes exist in incomplete-information environments.

Section 4 explores the implications of our notion of incomplete-information stability. Under intuitive sufficient conditions, these outcomes are efficient (in the sense of maximizing total surplus), but in general can fail equal treatment of equals. Incomplete-information stable outcomes are a superset of complete-information stable outcomes. In particular, incomplete-information stable outcomes may be efficient, and hence yield the same matching as would complete information, but involve different payments. These payments are important when considering settings in which firms or workers (or both) invest in their characteristics (types) before they enter the matching market.[4] The payments then determine investment incentives, and so the efficiency of the investment decisions.

Finally, we establish continuity results. Agents' payoffs in stable incomplete-information problems with "little" asymmetry of information are close to the payoffs to those with no asymmetry of information. This provides the robustness result that one need not literally believe in complete information, instead being confident of a complete-information analysis as long as there is not too much one-sided uncertainty in the economy.

Section 5 introduces a notion of price-sustainable allocations and shows that the set of stable outcomes is a (in general, strict) subset of the set of such allocations.

We are not the first to study these kinds of questions, and we discuss the related literature in Section 6.

[4] Cole, Mailath, and Postlewaite (2001a,b) study this question in complete information environments, while Mailath, Postlewaite, and Samuelson (2012, 2013), discussed in Section 6.3, study a competitive model with incomplete information.

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 $W \subset \mathbb{R}$ being the finite set of possible worker types and $F \subset \mathbb{R}$ being the finite set of possible firm types. The function mapping each firm to her type is denoted by $\mathbf{f} : J \to F$. The function mapping each worker to his type is denoted by $\mathbf{w} : I \to W$.

Value is generated by matches. We take as primitive the aggregate match value each agent receives in the absence of any payments between the agents. Following Mailath, Postlewaite, and Samuelson (2012, 2013), 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 (possibly negative) payment made to worker i by firm j.

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 *payment scheme* $\mathbf{p}$ associated with a matching function μ is a vector that specifies a payment $\mathbf{p}_{i,\mu(i)} \in \mathbb{R}$ 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 payments 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 payment 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 notion. There are three workers and firms ($I = J = \{a, b, c\}$). 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 $\mathbf{f}(a) = 2$, $\mathbf{f}(b) = 4$, and $\mathbf{f}(c) = 5$. 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 $\mathbf{w}(a) = 1$, $\mathbf{w}(b) = 3$, and $\mathbf{w}(c) = 2$, and that this is commonly known. The notion of stability for this complete-information setting is familiar from Gale and Shapley (1962) and Shapley and Shubik (1971). Because the surplus function is supermodular, the only stable matching must be positive assortative in type, which is the efficient

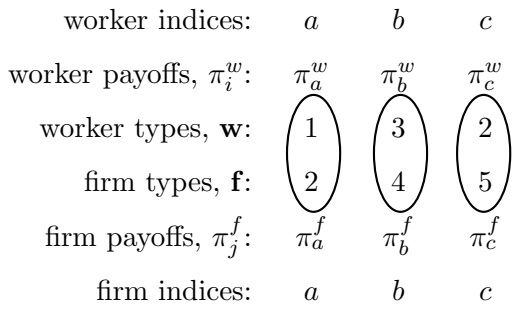

Figure 1: A matching that cannot be complete-information stable. Workers and firms are indexed by column. The matching of types is indicated by the ovals: $\mu(i) = i$, for $i \in \{a, b, c\}$.

matching in the sense of maximizing total surplus (Shapley and Shubik, 1971).

To illustrate the reasoning behind this result, consider the matching shown in Figure 1, which is not assortative. Since the matching of worker b (who has type 3) with firm b (who has type 4) generates a surplus of 24, we have $\pi_b^w + \pi_b^f = 24$, and similarly $\pi_c^w + \pi_c^f = 20$. But the surplus generated by a positive assortative matching by type of the top two workers and firms is 46. In the candidate match of Figure 1, either $\pi_b^w + \pi_c^f < 30$ or $\pi_c^w + \pi_b^f < 16$, and hence either worker b and firm c, or worker c and firm b, can form a blocking coalition (i.e., can match and make a payment 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 to indices. Suppose the realized types and the matching of firms to workers are as in Figure 1, with the payments and payoffs shown in Figure 2. Firms believe the set Ω of possible vectors $(\mathbf{w}(a), \mathbf{w}(b), \mathbf{w}(c))$ 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.

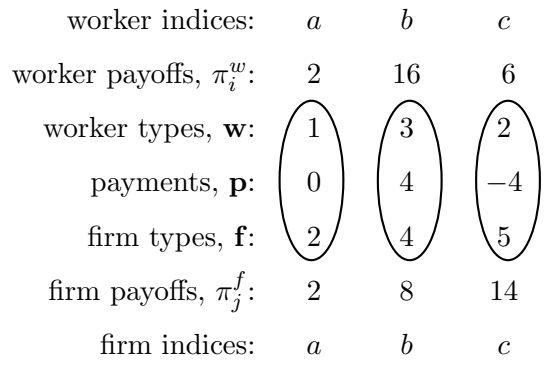

Figure 2: A possible outcome of the worker type assignment under incomplete information, with a matching outcome, payments, and payoffs. Types and premuneration values match those of Figure 1; workers and firms are indexed by column, and the matching is by index (indicated by the ovals).

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 c, firm b, and some payment $\tilde{p} \in (-2, 0)$. 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 b does not know whether worker c is of type 1 or type 2. The proposed deal is advantageous for firm b 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 c to be type 1? It initially appears that this is the case, since firm b knows only that worker c is not of type 3. However, the firm may be able to refine her beliefs on the strength of the fact that worker c is willing to participate in the block. To pursue this, notice that if worker c 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 b then cannot be sure whether the proposal involves a worker of type 1 or type 2. Hence, the firm might 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 known not to be blocked. In this case, firm b can reason as follows: Suppose worker c *were* of type 1. Then firm c would receive a payoff of 9, worker a would be of type 2 and receive payoff 4, and firm c would know that worker a was of type at least 2. Worker a and firm c could then match at payment of 0 (for example), giving each a higher payoff than the candidate stable allocation and thus constituting a blocking pair. But firm b's working hypothesis is that the proposed allocation is not blocked, and hence that there is no such blocking pair. Then worker c cannot be of type 1, and hence must be of type 2. This ensures that the originally proposed block is profitable for firm c, and hence that we indeed have a successful block. 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 b'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 $\mathbf{f}(a) = 2$, $\mathbf{f}(b) = 4$, and $\mathbf{f}(c) = 5$. 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.

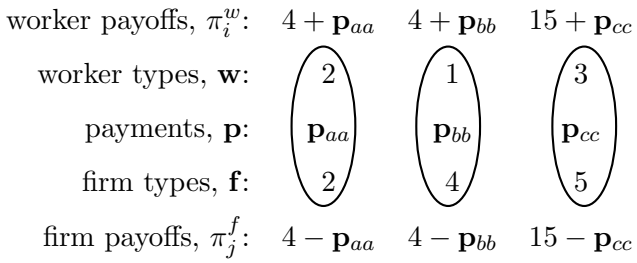

Figure 3: A matching in which the lowest type worker does not match with the lowest type firm. Workers and firms are indexed by column. Types and premuneration values are from Figure 1. The matching of types is indicated by the ovals: $\mu(i) = i$, for $i \in \{a, b, c\}$. The worker type assignment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

Suppose first that

$$\mathbf{p}_{aa} > 4 + \mathbf{p}_{bb},$$

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

Suppose instead that

$$\mathbf{p}_{aa} \leq 4 + \mathbf{p}_{bb},$$

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

This ensures that the lowest types of firm and worker must be matched. As we show in Section 4.1.2, 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 when premuneration values are supermodular, only positive 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*

$$\begin{aligned} \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} &\geq 0 \quad \textit{for all } i \in I \textit{ and} \\ \phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j} &\geq 0 \quad \textit{for all } j \in J. \end{aligned}$$

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 payments.

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 payment* $p \in \mathbb{R}$ *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 payments). 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 immunity of the matching to blocking be common knowledge.

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 payment. 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 and his payment.

We model the firms' inferences via a procedure of iterated elimination of blocked 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).

Consider a firm contemplating a blocking match with a worker, knowing that the realization of worker types is consistent with a set of matching outcomes Σ, and suppose the firm has a probability distribution over those consistent worker types. The firm would agree to a contemplated change in partner only if the expected payoff from doing so was positive. Typically, a variety of probability beliefs over worker types will be consistent with Σ, with a contemplated change having a positive expected value for some beliefs and a negative one for others. If the expected payoff of the change is positive for *every* belief the firm might have, we do not need to know the firm's beliefs to be sure the firm will agree to the change. Our notion of blocking is designed to only exclude outcomes that we can be certain will give rise to objection:

Definition 4 *Fix a non-empty set of individually rational matching out-*

comes, Σ*. A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma$ *is* Σ-blocked *if there is a worker-firm pair* (i, j) *and payment* $p \in \mathbb{R}$ *satisfying*

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

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*

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

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

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

A matching outcome $(\mu, \mathbf{p}, \mathbf{w}) \in \Sigma$ *is* Σ-stable *if it is not* Σ*-blocked.*

Inequality (1) requires that worker i receive a higher payoff in the potential block than under the match. Inequality (2) requires that firm j expect a higher payoff in the proposed block than under the match, 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 potential block 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 Σ^0-blocked, where Σ^0 is the set of all individually rational matching outcomes, irrespective of the level of payments. 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 B.3). In other cases, the precise nature of the payments determines whether the matching outcome is Σ^0-blocked. For example, the matching outcome in Figure 4 may or may not be Σ^0-blocked, depending on $\mathbf{p}$. Suppose first that $\mathbf{p}_{cc} = -2$, and consider a candidate blocking pair consisting of worker b

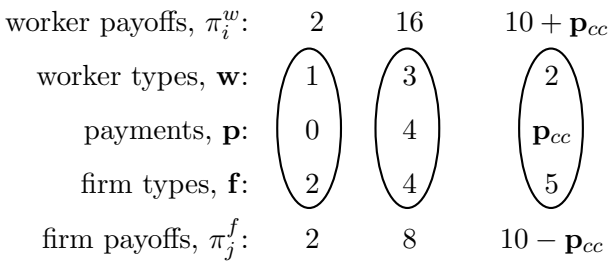

Figure 4: A matching outcome that is not Σ^0-stable (where Σ^0 is the set of individually rational matching outcomes) for the payment $\mathbf{p}_{cc} = -2$, but is Σ^0-stable for the payment $\mathbf{p}_{cc} = -4$ (the outcome from Figure 2). Types and premuneration values are from Figure 1.

(who has type 3) and firm c, with payment $p \in (1, 2)$. Worker b prefers this resulting match to the proposed equilibrium outcome. Moreover, firm c can calculate that a worker matched with firm b would prefer such an alternative match if and only if the worker is of type 3, ensuring that firm c also strictly prefers the candidate blocking match and hence that the candidate outcome is Σ^0-blocked. In contrast, the outcome with $\mathbf{p}_{cc} = -4$ (the outcome from Figure 2) is Σ^0-stable: Note first that worker b and firm c 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 c and firm b 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 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. Of course, if a firm-worker pair blocks a particular match, this does not mean the resulting match is stable. The new match may itself be blocked, and the fact that a worker-pair blocked the initial match may change some firms' information. We are interested in understanding the set of potential final outcomes of such a process, i.e., the set of outcomes that are immune to further such changes, without assuming too much about the nature of the protocol (extensive form) of firm-worker interactions and without assuming anything about the firms' beliefs beyond that implied by the common knowledge of the matching. Toward, this end, we exclude an allocation only if we can identify a block for which we are

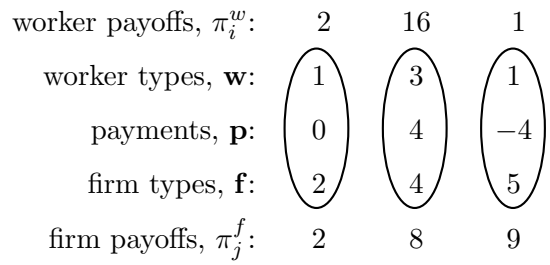

Figure 5: The payments and matching from Figure 2 with a different worker type realization.

confident the firm 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 Σ^1-blocked and hence is not contained in Σ^2, because outcomes with $\mathbf{w}'(c) = 1$ (such as the one displayed in Figure 5) are not contained in Σ^1 (for Figure 5, there is a successful block at the payment $p = -\frac{1}{2}$ of worker c with firm a).

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 individually rational matching outcomes E is* self-stabilizing *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). Note that the first claim trivially yields existence, since complete-information stable outcomes always exist in our setting.

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.*[5]
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.*

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 $E \subset \Sigma^\infty$.*

[5]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 $(\mu, \mathbf{p}^n, \mathbf{w}, \mathbf{f}) \in E$.

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 $\mathbf{w}$ stabilizing the allocation $(\mu, \mathbf{p})$.

4 Implications of Incomplete-Information Stability

4.1 Allocative Efficiency

4.1.1 Payoff Assumptions

While our notion of incomplete-information stability is based upon a demanding notion of blocking and hence is relatively permissive, under natural assumptions on premuneration values, stable matchings maximize total surplus. We consider the following assumptions:

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

Assumption 2 (Supermodularity) *The worker premuneration value ν_{wf} and the match surplus $\nu_{wf} + \phi_{wf}$ are strictly supermodular in w and f.*

Assumption 3 (Submodularity) *The worker premuneration value ν_{wf} and the match surplus $\nu_{wf} + \phi_{wf}$ are strictly submodular in w and f.*

We focus the discussion on the case in which Assumptions 1–2 hold. 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/submodularity assumptions are imposed on the worker premuneration values and on total surplus, but not separately on firm premuneration values.

4.1.2 Efficiency Under Supermodularity

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 a blocking coalition at the associated payment. 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 B.1).

Lemma 2 *Suppose Assumptions 1 and 2 (supermodularity) hold, 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 payment* p^**, then for any firm with type* $f > f^*$*, there exists* $\varepsilon > 0$ *such that for any* $p \in (\nu_{w^*f^*} + p^* - \nu_{w^*f},\ \nu_{w^*f^*} + p^* - \nu_{w^*f} + \varepsilon]$*,*

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

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

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

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

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

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

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 payment of p just above $\nu_{w^*f^*} + p^* - \nu_{w^*f}$. The type f firm understands that the worker benefits from participating 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 payment 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.

Under supermodularity, an outcome is efficient (i.e., maximizes total surplus) only if it features positive assortative matching. In addition, efficiency requires that pairs producing negative surpluses are not matched. Incomplete-information stability guarantees both properties, and so all stable outcomes are efficient. Appendix B proves the following.

Proposition 3 *Under Assumptions 1 (monotonicity) and 2 (supermodularity), every incomplete-information stable outcome is efficient.*

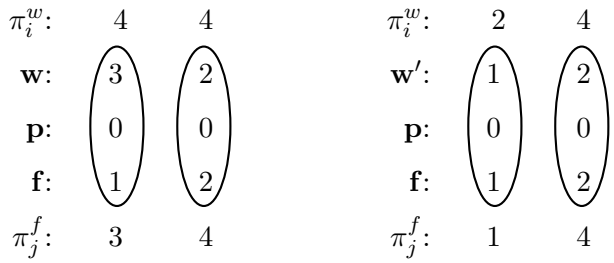

Figure 6: A failure of efficiency in the absence of strict supermodularity. The first matching is incomplete-information stable, stabilized by the second complete-information stable outcome. In this example, $W = \{1, 2, 3\}$, $F = \{1, 2\}$, $\nu_{wf} = w + f$, and $\phi_{wf} = wf$.

We now describe an example demonstrating that without strict supermodularity stable outcomes may be inefficient. There are two workers and two firms. It is commonly known that $\mathbf{f}(a) = 1$, $\mathbf{f}(b) = 2$, and $\mathbf{w}(b) = 2$. We suppose that worker a's type could be either 1 or 3, and the realized value is 3. The worker's premuneration value is $w + f$ (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-stabilizing as a singleton set. This outcome stabilizes $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ as follows. The only potential blocking pair at type assignment $(\mathbf{w}, \mathbf{f})$ involves worker a and firm b. However, workers of type 1 and type 3 would both be willing to participate in such a block at any price greater than -1, and hence there is no way for firm b to exclude the possibility that worker a is type 1. The allocation then cannot be blocked, and hence we have incomplete-information stability.

4.1.3 Efficiency Under Submodularity

Under submodularity, an outcome is efficient only if it features negative assortative matching. In addition, efficiency may require certain agents to

		Firm types		
		1	2	3
	1	−7	0	4
	2	−3	3	5
Worker	3	1	6	7.5
types	4	5	8	8.5
	5	8	8.5	8.75

Figure 7: Submodular worker and firm premuneration values for the example illustrating "too much" matching under incomplete-information stability.

be unmatched. Just as with supermodular values, firms can draw relatively sharp inferences about the type of worker from the worker's willingness to participate in a blocking coalition at the associated payment. Using Lemma 2's analogue (see Online Appendix 1.1), we show that incomplete-information stability guarantees negative assortative matching.

However, incomplete-information stability may support "too much" matching. Consider the following example with three types of firms and five types of workers. Workers and firms receive the same premuneration values in a match; these values are described in Figure 7. These premuneration values are submodular.[6] As usual, we have normalized names so that worker a matches with firm a, worker b with firm b, and worker c with firm c. The firm type assignment is $\mathbf{f}(a) = 3$, $\mathbf{f}(b) = 2$, and $\mathbf{f}(c) = 1$. Suppose $\Omega = W^I$, and consider the pair of worker type assignments, $\mathbf{w} = (1, 2, 5)$ and $\mathbf{w}' = (1, 3, 4)$. A self-stabilizing set is given in Figure 8.

The matching outcome on the right of Figure 8, though incomplete information stable, is inefficient (see Figure 9). This inefficiency arises from two aspects: the efficient outcome involves some unmatched agents, and the two worker type assignment functions in the self-stabilizing set "cross." In particular, at $\mathbf{w}'$, $\mathbf{w}$ is the firm 3 pessimistic worker type assignment, and

[6]Since the matches of worker types 1 and 2 with a firm type 1 yield negative surpluses in Figure 7, we are effectively assuming agents in a match cannot simply ignore their partners and guarantee a value of 0. However, the relevant features of the example are unchanged if we replace the negative values with zeroes. While the resulting premuneration values are not globally submodular, they are submodular on the restricted domain where premuneration values are strictly positive.

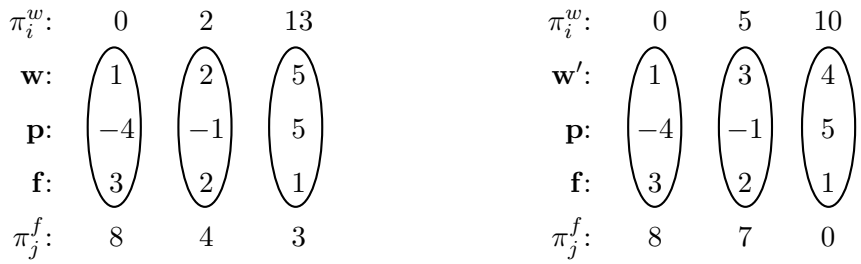

Figure 8: A self-stabilizing set for the premuneration values given in Figure 7. The matching outcome on the left is complete information stable (and efficient).

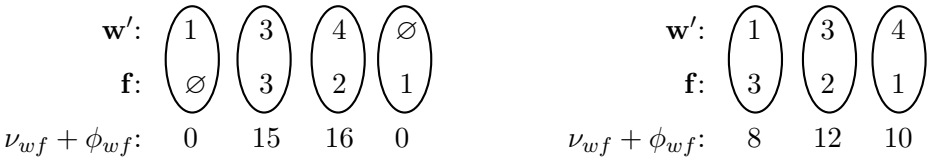

Figure 9: The matching outcome on the right of Figure 8 (reproduced here on the right) is inefficient, being dominated by the matching on the left in this figure.

yet under that type assignment, the worker matched with firm 1 has a higher type than under $\mathbf{w}'$.

Eliminating the possibility of unmatched pairs in an efficient matching is sufficient to guarantee the efficiency of incomplete information outcomes. Online Appendix 1 proves the following proposition. Say that an outcome is *negative assortative* if for all $i, i' \in I$ such that $\mu(i), \mu(i') \in J$, if $\mathbf{w}(i) < \mathbf{w}(i')$, then $\mathbf{f}(\mu(i)) \geq \mathbf{f}(\mu(i'))$. Note that this notion of negative assortativity does not impose any restrictions on the type of unmatched agents.

Proposition 4 *Under Assumptions 1 (monotonicity) and 3 (submodularity), every incomplete-information stable outcome is negative assortative. Moreover, if in addition $\phi_{wf} + \nu_{wf} > 0$ for all pairs $wf \in W \times F$, then every incomplete-information stable outcome is efficient.*

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 10. 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 10. 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 a and firm b cannot block $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$. This follows

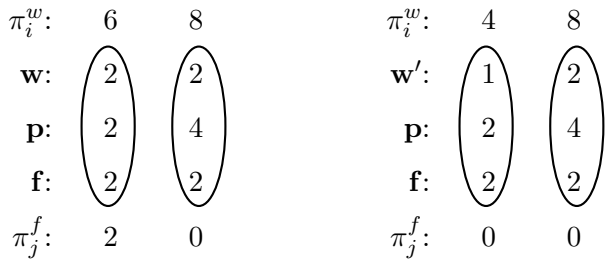

Figure 10: A failure of equal treatment. The first matching is incomplete-information stable, stabilized by the second complete-information stable outcome. In this example, $W = \{1, 2\}$, $F = \{2\}$ and $\nu_{wf} = \phi_{wf} = wf$.

from the possibility that worker a is of type 1 rather than type 2, since any payment 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 4.1.2 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 are close or coincide.

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 that it requires the supports to be close.

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

Assumption 4 (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 $\pi(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \mathbb{R}^{I \times J}$ be the vector of payoffs that workers and firms receive in the matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$, and denote by $\pi(\Sigma^\infty(\Omega)) \subset \mathbb{R}^{I \times J}$ 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 (the proof is in Appendix C.1).

Proposition 5 *Suppose Assumption 4 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 4.1.2 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 restriction on the set of possible worker-type assignments:

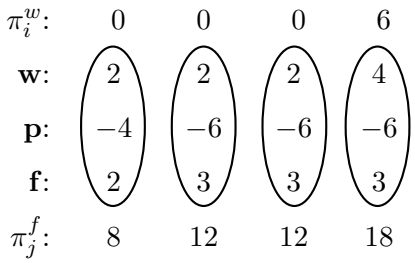

Figure 11: An incomplete-information stable matching outcome (when Ω is a set of permutations) that is not complete information stable.

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 11 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 and 2 types of workers. Premuneration values are given by $\nu_{wf} = \phi_{wf} = wf$. This is not complete-information stable, as worker d and firm c can form a blocking pair. In the incomplete-information setting, any payment at which worker d is willing to match with firm c also makes worker b willing to match with firm c. Firm c 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.

More formally, let E be the set of allocations in which μ, $\mathbf{p}$, and $\mathbf{f}$ are as shown in Figure 11, worker a is known to be of type 2, and the types of workers b, c, and d are drawn from the set of permutations of $(2, 2, 4)$. Then E is a self-stabilizing set. Notice that this self-stabilizing set contains no complete-information stable outcome—while we often find it convenient to

show that an allocation is incomplete-information stable by pairing it in a self-stabilizing set with a complete-information stable outcome, the presence of the latter is not necessary. Indeed, taking Σ to be the support of our self-stabilizing set E, no complete-information stable allocation can give rise to the price function $\mathbf{p}$.

The difficulty in this example is that the observables, namely firms' types and payments, 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 C.2):

Proposition 6 *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), when different workers have different types, the payment $\mathbf{p}$, which is observable, is fully informative about worker type regardless of firm types. The sufficient conditions given in this proposition are not necessary, and can be slightly weakened at the cost of somewhat more complicated statements. For example, it suffices that under every assortative matching, there is no overlap between strings of identical worker types and strings of identical firm types.

5 Stability and Pricing

In this section, we examine the connection between the set of stable outcomes and allocations that we might see in a market environment. Section 4.1.2 established conditions under which an inefficient matching is not incomplete-information stable. The instability will arise because there is a payment for some unmatched pair at which both will be sure they gain. We now examine whether one can rely on a price system to ensure that inefficient outcomes will similarly not persist. We introduce a notion of

price-sustainable outcomes in order to answer this question. The basic idea is to formulate a notion analogous to the stability notion for incomplete information problems above, but requiring the decisions of both workers and firms be mediated through market prices rather than direct contact.

We might expect price sustainability to be either more or less demanding than incomplete-information stability. Objections to a candidate allocation under price sustainability must be made at the existing prices, constraining the ability to convey information and making it more difficult to block an allocation. However, blocking an allocation under price sustainability requires only a market imbalance, arising from a single agent's preference for a different match, rather than the double coincidence of wants required under incomplete-information stability.

As with our notion of incomplete-information stability, our focus is on matches that have already been formed. We do not address how such matches and the transfers within the match arose. This distinguishes our notion from notions of competitive equilibrium, since there is no privileged outcome in a typical model of competitive equilibria.

5.1 The Economy

The "commodities" in a two-sided matching market are "partnerships" of the form (i, j), denoting a match between worker i and firm j, with $(i, \varnothing)$ denoting an unmatched worker i and $(\varnothing, j)$ denoting unmatched firm j. Each firm j' can demand at most one partnership, of the form (i, j') for some i, and each worker i' can similarly supply at most one partnership, of the form (i', j). Let $\mathbf{P}$ be a price matrix $\mathbf{P} : I \times J \to \mathbb{R}$ associating a price with each match of the form (i, j), while defining the prices for staying unmatched as $\mathbf{P}_{i\varnothing} = \mathbf{P}_{\varnothing j} = 0$ for any $i \in I$ and $j \in J$.

A *price-taking matching outcome* specifies the partnerships that are traded, i.e., a matching function $\mu : I \to J \cup \{\varnothing\}$, and a price matrix $\mathbf{P}$.

Definition 8 *A price-taking matching outcome* $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ *is* individually rational *if* $\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{P}_{i,\mu(i)} \geq 0$ *for any* $i \in I$ *and* $\phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{P}_{\mu^{-1}(j),j} \geq 0$ *for any* $j \in J$.

Note that the definition of individual rationality presumes that each firm knows the type of the worker she is matched with, consistent with the idea that we are considering an existing match.

5.2 Price-Sustainable Matching

Intuitively, a price-sustainable matching outcome requires that workers and firms choose their partnerships optimally, given fixed prices, and market clearing. Worker i must find it optimal to match with $\mu(i)$ at a price $\mathbf{P}_{i,\mu(i)}$ instead of matching with a firm $j \neq \mu(i)$ at a price $\mathbf{P}_{ij}$ or staying alone at a price 0, with a similar requirement for firms (taking into account each firm's incomplete information about workers with whom the firm is not matched). Market clearing is explicit in the definition of $\mu : I \to J \cup \{\varnothing\}$.

Definition 9 *Fix a non-empty set of individually rational price-taking matching outcomes Ψ. A price-taking matching outcome $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \in \Psi$ is* Ψ-price-sustainable *(or simply* Ψ-sustainable*) if there is no $i \in I$ and $j \in J$ for which*

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

or

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

for all $\mathbf{w}' \in \Omega$ and $\mathbf{P}' : I \times J \to \mathbb{R}$ satisfying

$$\begin{gathered}\left(\mu, \mathbf{P}', \mathbf{w}', \mathbf{f}\right) \in \Psi, \\ \mathbf{w}'(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j)), \quad \textit{and} \\ \mathbf{P}'_{i',\mu(i')} = \mathbf{P}_{i',\mu(i')} \textit{ and } \mathbf{P}'_{i'j} = \mathbf{P}_{i',j}, \quad \forall i' \in I.\end{gathered}$$

Each firm knows the type of her own worker and every agent knows the price of any partnership $(i, \mu(i))$, which is to say that agents know the prices of the goods that are traded. We also assume that each worker i' knows the prices of each partnership (i', j) and each firm j' knows the price of each partnership (i, j'). Hence, each agent knows the prices of all of the goods in his or her consumption set. We do not assume that firm j' knows the prices of partnerships (i, j'') for which $j'' \neq \mu(i)$, so that j' does not know the prices of partnerships that are not traded and that j' *could not* trade.

A price-taking matching outcome fails to be Ψ-sustainable if some matched agent prefers to stay unmatched or if some agent wants to deviate to a different match at the equilibrium price for that match, regardless of whether the other side wants to accept the agent or not.

Since firm j does not observe workers' types (other than j's current match), we must again consider firm j's beliefs about worker types. For sustainability to fail because a firm has a superior alternative transaction, this alternative must be superior for every type assignment $\mathbf{w}'$ and price matrix

$\mathbf{P}'$ satisfying the three criteria given in Definition 9: (i) the type assignment must be consistent with matching outcomes in the set Ψ, a restriction that will become operational in the iterative argument we construct next; (ii) the type assignment must not contradict what the firm j already knows at the stage, i.e., must assign to firm j the type $\mu^{-1}(j)$ of worker with whom firm j is matched; (iii) $\mathbf{P}'$ must be consistent with the prices the firm knows.

Definition 10 *Let Ψ^0 be the set of all individually rational price-taking matching outcomes. For $k \geq 1$, define*

$$\Psi^k := \left\{(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \in \Psi^{k-1} : (\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \text{ is } \Psi^{k-1}\text{-sustainable}\right\}.$$

The set of price-sustainable outcomes *is given by*

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

If $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ is a price-sustainable outcome, the outcome $(\mu, \mathbf{P})$ is a price-sustainable allocation at $(\mathbf{w}, \mathbf{f})$.

Many rounds of iteration may be required before the process introduced in Definition 10 reaches the set of price-sustainable outcomes (we provide an illustration in Online Appendix 2.1).

We can compare our notion of a price-sustainable matching to a variety of formulations of competitive equilibrium with incomplete information. Radner (1979) introduces a notion of competitive equilibrium for economies with incomplete information, showing that (generically) competitive equilibrium prices reveal all asymmetric information. Every agent consumes every good in Radner's model, making it reasonable to assume that the prices of all goods are common knowledge, whereas most of the goods are untraded in our case, and we do not assume that the agents have common knowledge of the prices of untraded goods. Hatfield, Kominers, Nichifor, Ostrovsky, and Westkamp (2013) examine a notion of competitive equilibrium for a complete-information economy in which it is possible (but not necessarily the case) that every agent consumes every good, and in which all prices are known, whether the goods involved are traded or not.

As in the case of stability, there is a convenient fixed-point characterization of price-sustainable outcomes.

Definition 11 *A nonempty set of individually rational price-taking matching outcomes C is* self-sustaining *if every $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \in C$ is C-sustainable.*

The following result (proven in Online Appendix 2.2) is analogous to its stability counterpart.

Lemma 3 *The set of price-sustainable outcomes Ψ^{∞} is self-sustaining. If C is self-sustaining, then $C \subset \Psi^{\infty}$.*

By virtue of this lemma, to show that $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ is price sustainable it suffices to find a set C that is self-sustaining and contains $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$.

5.3 Stable and Price-Sustainable Matching Outcomes

In a price-sustainable matching outcome, firms draw inferences about workers' types from prices. However, the assumption that all transactions must occur at the candidate prices limits the inferences firms can draw. In a stable matching outcome, there are no restrictions on the payments that might be involved in a candidate blocking pair. This allows more information to be revealed. Greater information revelation makes it easier for pairs to identify beneficial deviations, and hence the stability requirement is more demanding than that of price sustainability. As a result, the set of stable matching outcomes is a subset of the set of price-sustainable outcomes. Online Appendix 2.3 proves:

Proposition 7 *If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is incomplete-information stable, then there exists $\mathbf{P} : I \times J \to \mathbb{R}$ extending $\mathbf{p}$ (so that $\mathbf{P}_{i\mu(i)} = \mathbf{p}_{i\mu(i)}$) such that $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ is price sustainable.*

We now illustrate that stability can strictly refine price-sustainable outcomes. In particular, there are price-sustainable outcomes that are not incomplete-information stable. Suppose the premuneration values are given by $\nu_{wf} = wf$ and $\phi_{wf} = 2 + wf$. Suppose moreover that Ω contains two type vectors: $\mathbf{w} = (3, 2)$ and $\mathbf{w}' = (1, 2)$. Consider the matching outcomes $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ and $(\mu, \mathbf{P}, \mathbf{w}', \mathbf{f})$ given in Figure 12.

Note first that the singleton set $\{(\mu, \mathbf{P}, \mathbf{w}', \mathbf{f})\}$ is self-sustainable: For example, if firm b takes worker a instead, firm b gets payoff $\pi_b^{fd} = 2 + wf - \mathbf{P}_{ab} = 2 + (1 \cdot 2) - 0 = 4$. In the outcome $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$, firm b is uncertain of the type of worker a. We use $\mathbf{w}'$ to enforce firm b's optimization. Hence by deviating, the firm cannot rule out a payoff of 4.

Note that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ cannot be incomplete-information stable, where $\mathbf{p}$ is the restriction of $\mathbf{P}$, because the matching outcome is not efficient. Consider a proposed blocking pair consisting of worker a and firm b, and a payment of 1. Note that the price is $\mathbf{P}_{ab} = 0$. This offer differentiates worker

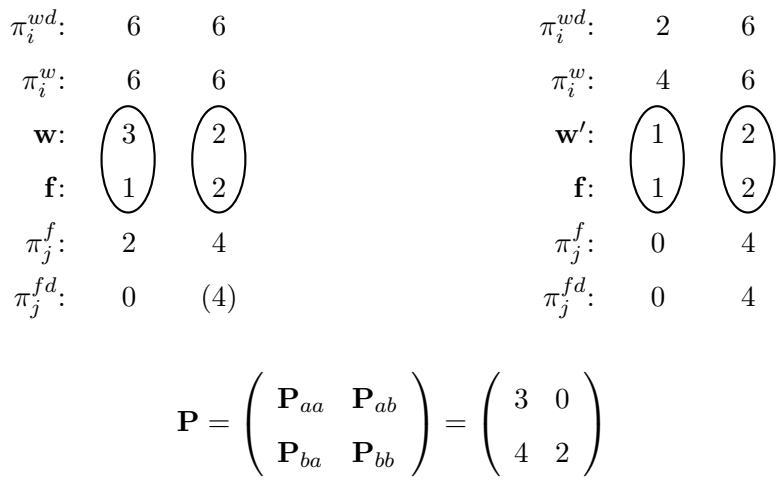

Figure 12: Example of a price-sustainable matching outcome that is not incomplete-information stable. There is a single price matrix. Worker i's payoff from matching with firm $j \neq i$ is denoted π_i^{wd}; firm j's payoff from matching with worker $i \neq j$ is denoted π_j^{fd}. Since $\Omega = \{\mathbf{w}, \mathbf{w}'\} = \{(3,2),(1,2)\}$, at $\mathbf{w}$, firm b cannot rule out a payoff of 4 (implied by $\mathbf{w}'$) from matching with worker a.

type 3 from worker type 1. To see this, note that the former's payoff by this deviation is 7, larger than 6, the payoff from the candidate matching; while the latter's payoff is 3, less than 4, the payoff from the candidate matching. Since worker type 3 can reveal its type, firm type 2 obtains a payoff of 7 by matching with this worker at a price 1.

6 Discussion

6.1 Necessary Conditions

Our analysis begins 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 can infer from the common knowledge that the matching is unblocked. 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 our notion of stability as capturing necessary conditions for an outcome to be the potential product of a matching process. In this sense, our work is most closely related to the literature on the core, particularly 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 by coalitions of agents, making 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. 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.[7] While our analysis is quite similar in spirit, there are important differences. 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

[7] See Serrano and Vohra (2007) and Myerson (2007) for similar models, and see Yenmez (2013) for similar stability notions in a matching environment.

inference and is consistent with existence.

The process of drawing iterated inferences could also appear in other contexts, leading to analogous notions of incomplete-information stability, even if the details of the inferences would differ. For example, a marriage model is a special case of our model in which there are no transfers. The absence of transfers would make it more difficult for an agent to convey information through a proposed block, precluding a result analogous to Lemma 2, and hence we would expect the set of incomplete-information stable outcomes to be relatively large, but the structure of the analysis carries through unchanged.

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)).

6.2 Origins

In a context 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.[8] However, as Roth and Vande Vate (1990, p. 1475) note, many matching markets do not make use of centralized mechanisms. Roth and Vande Vate (1990) analyze a process that allows randomly chosen blocking pairs to match and show that the process converges to a stable matching, though they do not model the incentives facing the agents throughout this process.[9] 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 the market) or returning to the unmatched pool, and with agents choosing throughout so as to maximize

[8] 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). However, every Nash equilibrium outcome of this deferred-acceptance-based mechanism in which proposers follow their weakly dominant strategy of announcing truthfully is stable with respect to the agents' true preferences (Roth, 1984).

[9] Kojima and Ünver (2008) analyze a similar model with many-to-many matching.

their expected payoffs.[10] Lauermann and Nöldeke (2012) show that the equilibria of this process converge to the set of stable outcomes, if, but only if, there is a unique stable outcome. Even under complete information, we cannot be assured of convergence to a stable outcome when there are multiple such outcomes.

Under incomplete information, the connection between stable matches and the process by which stable matches are formed is yet less obvious. In the process of encountering others 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 at the beginning. 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 centralized mechanisms. 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. Dizdar and Moldovanu (2012) identify conditions under which, given incomplete information and *nontransferable* utility, a mechanism exists that invariably yields complete-information stable outcomes.

[10] Adachi (2003) analyzes a model that is similar, but one in which agents who match leave the market but are replaced by "clones," and in which agents are restricted to pure strategies.

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.[11]

6.3 Premuneration Values

Why do we work with premuneration values and prices, rather than simply abstract divisions of the surplus? Indeed, given that prices are simply transfers and efficiency depends only on the matching pattern, why not simply

[11] 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.

ignore prices?

The importance of premuneration values is stressed by Mailath, Postlewaite, and Samuelson (2012, 2013), who 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 payment. While there are significant modelling differences between the models in that paper and here, the two models share the property that the attributes of all the agents on on one side of the matching market are public, while those of all the agents on the other side are private. An important feature of Mailath, Postlewaite, and Samuelson (2012, 2013) is that premuneration values, which are typically irrelevant in complete-information environments, become important in the presence of incomplete information. In particular, premuneration values play a critical role in determining whether the post-investment matching outcomes creates the incentives required for agents to undertake efficient investments. The modeling assumptions in the current paper reflect a belief that people often undertake investments before entering matching markets, and that premuneration values and payments affect investment incentives.

Premuneration values will also play in important role in studying how stable outcomes might arise. For example, 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. 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 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.[12] Interestingly, Google auctions locations 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 process, however, the fee structure opens the possibility that the resulting outcome will not be stable.

6.4 Extensions

There are two obvious directions for extending our analysis. First, 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, but this is no longer the case under incomplete information. Expanding the analysis beyond pairwise blocking coalitions will require taking a stand on what inferences agents can draw from the hypothesis that the entire coalition is willing to participate. While the details are nontrivial, many of our results will clearly carry

[12] 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.

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.

Second, we have examined one-to-one matching. The analysis could be readily extended to simple cases of many-to-one matching. For example, suppose that firms can hire more than one worker, but that each worker cares only about the firm with whom he is matched (and not about the characteristics of the other workers matched with that firm). Suppose further that the firm's payoff is an additively separable function of the types of workers it hires. We could then find generalizations of our monotonicity and supermodularity assumptions ensuring an efficient match. Extensions to richer many-to-one matching models is an obvious next step.

A Appendix: Proofs for Section 3

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 payment 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 payment 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 5). 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 Σ^∞-blocked. In particular, there is an unmatched pair (i, j) and payment $p \in \mathbb{R}$ 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 $\Sigma^\infty = \overline{\Sigma^\infty}$.

B Appendix: Proofs for Section 4.1.2

B.1 Proof of Lemma 2

Define

$$p^\varepsilon := \nu_{w^*f^*} + p^* - \nu_{w^*f} + \varepsilon, \tag{B.1}$$

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 (B.1),

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

proving (7).

After substituting for $p = p^\varepsilon$ defined in (B.1), 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. ■

B.2 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 $\mathbf{w}(i) > \mathbf{w}(i')$.

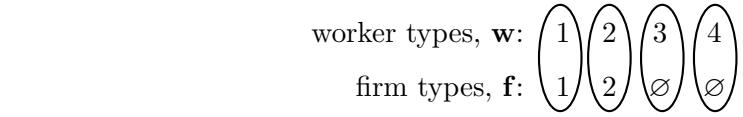

Figure B.1: A matching that is 1st-order, but is not 2nd-order worker assortative. There are 4 workers and 2 firms, $W = \{1, 2, 3, 4\}$ and $F = \{1, 2\}$, and workers and firms have different types.

Figure B.2: The two nontrivial 2nd-order worker-assortative matchings for the environment of Figure B.1. The first is worker assortative, while the second is not.

Definition B.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 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 B.1 is *not* 2nd-order worker-assortative even though the lowest two worker types are matched assortatively. Figure B.2 displays the two nontrivial 2nd-order worker-assortative matchings.

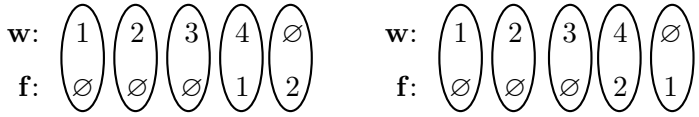

Figure B.3: The two other nontrivial worker-assortative matchings for the environment of Figure B.1.

In addition to the first matching displayed in Figure B.2, there is a trivial worker-assortative matching in which no worker or firm is matched and the two worker-assortative matchings displayed in Figure B.3.

The first matching in Figure B.3 is not firm-assortative and hence not assortative. The first matching in Figure B.2, the second matching in Figure B.3 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 B.1, this matching is 3^{rd}-order worker-assortative and 1^{st}-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 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 B.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 B.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}, $\mathbf{f}\circ\mu$ *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 firm assortative and for all* $f>f^{\ell+1}$, $\mathbf{w}\circ\mu^{-1}$ *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, $\mathbf{f}\circ\mu$ is weakly comonotone with $\mathbf{w}$ on $\{i:\mathbf{w}(i)\in\{w^1,\ldots,w^k,w^{k+1}\}\}$. Moreover, for all $w>w^{k+1}$, $\mathbf{f}\circ\mu$ 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 $\mathbf{f}\circ\mu$ 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 $w>w^{k+1}$, $\mathbf{f}\circ\mu$ 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. ■

B.3 The Proof of Proposition 3

Worker-Assortativity. Without loss of generality, assume worker and firm indices are positive integers and the true type assignment $(\mathbf{w},\mathbf{f})$ is such that $\mathbf{w}:I\to W$ and $\mathbf{f}:J\to F$ 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 B.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 there is some $(\mu,\mathbf{p},\mathbf{w},\mathbf{f})\in\Sigma^1$ not first-order worker assortative. Then by definition, $\mathbf{f}\circ\mu$ is not weakly 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 B.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 payment p. By Lemma 2 (taking $w^* = \mathbf{w}(2)$ and $f = \mathbf{f}(\mu(1))$, there exists $\varepsilon > 0$ such that for $-\nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))} < p \leq -\nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))} + \varepsilon$,

$$\nu_{w,\mathbf{f}(\mu(1))} + p > 0, \qquad \text{for any } w \geq \mathbf{w}(2), \text{ and} \tag{B.2}$$

$$\nu_{w,\mathbf{f}(\mu(1))} + p \leq 0, \qquad \text{for any } w < \mathbf{w}(2). \tag{B.3}$$

Worker 2 is better off because he gets a positive payoff, from (B.2); firm $\mu(1)$ will assign probability 1 that the deviating worker's type is at least $\mathbf{w}(2)$ because of (B.2) and (B.3). 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 B.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*

$$\phi_{\mathbf{w}(2),\mathbf{f}(\mu(1))} + \nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))} \leq \nu_{\mathbf{w}(2),\mathbf{f}(\mu(2))} + \mathbf{p}_{2,\mu(2)} + \phi_{w^1,\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)}. \tag{B.4}$$

Proof. Consider the worker-firm pair $(2, \mu(1))$ with payment p given by.

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

For sufficiently small $\varepsilon > 0$, by Lemma 2, every worker with type strictly below $\mathbf{w}(2)$ prefers his current match.

Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, firm $\mu(1)$ rejects worker 2 and p. That is, the firm must not be strictly better off in the new match. Hence,

$$\phi_{w,\mathbf{f}(\mu(1))} - p \leq \phi_{w^1,\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)} \text{ for some } w \geq \mathbf{w}(2). \tag{B.5}$$

Since ϕ_{wf} is increasing in w and f, the statement in (B.5) 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 (B.4). □

Claim B.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*

$$\nu_{w^1,\mathbf{f}(\mu(2))} + \phi_{w^1,\mathbf{f}(\mu(2))} \leq \left(\nu_{w^1,\mathbf{f}(\mu(1))} + \mathbf{p}_{1,\mu(1)}\right) + \left(\phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}\right). \quad \text{(B.6)}$$

Proof. If the inequality in (B.6) did not hold, we can find $q \in \mathbb{R}$ such that

$$\nu_{w^1,\mathbf{f}(\mu(2))} + q > \nu_{w^1,\mathbf{f}(\mu(1))} + \mathbf{p}_{1,\mu(1)} \quad \text{and} \quad \text{(B.7)}$$

$$\phi_{w^1,\mathbf{f}(\mu(2))} - q > \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}. \quad \text{(B.8)}$$

Since ϕ is increasing and w^1 is the smallest type, (B.8) implies that

$$\min_{w \in W} \phi_{w,\mathbf{f}(\mu(2))} - q > \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}. \quad \text{(B.9)}$$

Hence, (B.7) and (B.9) imply $(1, \mu(2))$ is a blocking pair, contradicting $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$. □

Finally, we combine Claims B.2 and B.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 B.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 B.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 B.1 shows (with obvious modifications) that $\mu(i') \neq \varnothing$.

Claim B.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*

$$\nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} + \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} \leq \nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} + \mathbf{p}_{i',\mu(i')} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)}. \tag{B.10}$$

Proof. Worker i' strictly prefers a block with firm $\mu(i)$ at a payment

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

for some small $\varepsilon > 0$, if and, by Lemma 2, only if, his type is at least $\mathbf{w}(i')$. 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 ϕ 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 (B.10). □

Claim B.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*

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} \leq \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} + \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \tag{B.11}$$

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

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} + q > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \quad \text{and} \tag{B.12}$$

$$\phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \tag{B.13}$$

By monotonicity of ϕ, (B.13) implies

$$\phi_{w,\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')} \text{ for all } w \geq \mathbf{w}(i) = w^{k+1}. \tag{B.14}$$

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 (B.14) that

$$\min_{\mathbf{w}' \in \Omega'} \phi_{\mathbf{w}'(i), \mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'), \mathbf{f}(\mu(i'))} - \mathbf{p}_{i', \mu(i')}. \tag{B.15}$$

Hence, from (B.12) and (B.15), the unmatched pair $(i, \mu(i'))$ at payment q can form a blocking pair. A contradiction. □

Summing (B.10) and (B.11), we have

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

contradicting strict supermodularity. ■

Assortativity.

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

Proof. From Lemmas B.3 and B.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, $\mu^{-1}(j') = \varnothing$.

Consider the potential blocking match of worker $\mu^{-1}(j)$ and firm j' with a payment $p = \nu_{\mathbf{w}(\mu^{-1}(j)), \mathbf{f}(j)} + \mathbf{p}_{\mu^{-1}(j), j} - \nu_{\mathbf{w}(\mu^{-1}(j)), \mathbf{f}(j')} + \varepsilon$. For $\varepsilon > 0$, worker $\mu^{-1}(j)$ strictly prefers the block to his current match if and, by Lemma 2 (taking $w^* = \mathbf{w}(\mu^{-1}(j))$, $f^* = \mathbf{f}(j)$, $f = \mathbf{f}(j')$, and $p^* = \mathbf{p}_{\mu^{-1}(j), j}$) only if, his type is at least $\mathbf{w}(\mu^{-1}(j))$. That is,

$$\begin{aligned}&\nu_{w, \mathbf{f}(j')} + p > \nu_{w, \mathbf{f}(j)} + \mathbf{p}_{\mu^{-1}(j), j}, && \text{for any } w \geq \mathbf{w}(\mu^{-1}(j)), \\ &\nu_{w, \mathbf{f}(j')} + p \geq 0, && \text{for any } w \geq \mathbf{w}(\mu^{-1}(j)), \\ &\nu_{w, \mathbf{f}(j')} + p \leq \nu_{w, \mathbf{f}(j)} + \mathbf{p}_{\mu^{-1}(j), j}, && \text{for any } w < \mathbf{w}(\mu^{-1}(j)).\end{aligned}$$

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) for ε small. Since ϕ_{wf}

is strictly increasing in f, we can choose ε such that $0 < \varepsilon < \phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j')} - \phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)}$. The payoff to firm j' in this block 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 B.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, consider the potential blocking match of worker i and firm j with a payment $p = -\nu_{\mathbf{w}(i),\mathbf{f}(j)} + \varepsilon$, 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 payment 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

$$\varepsilon < \min\{\phi_{\mathbf{w}(i),\mathbf{f}(j)} + \nu_{\mathbf{w}(i),\mathbf{f}(j)}, \nu_{\mathbf{w}(i),\mathbf{f}(j)} - \max_{w<\mathbf{w}(i)} \nu_{w,\mathbf{f}(j)}\}.$$

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 type 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. ■

C Appendix: Proofs for Section 4.3

C.1 Proof of Proposition 5

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 $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f}) \in \Sigma^\infty(\Omega^n)$ such that $\pi(\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f}) \notin \xi_\varepsilon(\pi(\Sigma^\infty(\{\mathbf{w}\})))$.

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 $(\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f}) \in \Sigma^\infty(\Omega^n)$ imply that the sequence $\{\mathbf{p}^n\}$ is bounded. Notice also that $\|(\mu^n, \mathbf{p}^n, \mathbf{w}, \mathbf{f}) - (\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f})\| \to 0$ as $n \to \infty$. Since $\pi(\mu^n, \mathbf{p}^n, \mathbf{w}^n, \mathbf{f}) \notin \xi_\varepsilon(\pi(\Sigma^\infty(\{\mathbf{w}\})))$, it follows that for sufficiently large n,

$$\pi(\mu^n, \mathbf{p}^n, \mathbf{w}, \mathbf{f}) \notin \xi_{\frac{\varepsilon}{2}}(\pi(\Sigma^\infty(\{\mathbf{w}\}))). \tag{C.1}$$

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 (C.1) 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,

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

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

C.2 Proof of Proposition 6

If different firms have different types, then from Lemma B.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 $1 < k \leq \left|\mathbf{f}^{-1}(f)\right|$,

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

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 payment $\mathbf{p}_{\mu^{-1}(j_k),j_k} + \varepsilon$, 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' payments. 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 payment $\mathbf{p}_{\mu^{-1}(j_k),j_k} + \varepsilon$ 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 payments 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

Adachi, H. (2003): "A Search Model of Two-Sided Matching Under Nontransferable Utility," *Journal of Economic Theory*, 113(2), 182–198. 34

Becker, G. S. (1973): "A Theory of Marriage; Part I," *Journal of Political Economy*, 81(4), 813–846. 17

Bernheim, B. D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52(4), 1007–1028. 12

Chade, H. (2006): "Matching with Noise and the Acceptance Curse," *Journal of Economic Theory*, 129(1), 81–113. 35

Chade, H., G. Lewis, and L. Smith (2011): "Student Portfolios and the College Admissions Problem," Aizona State University, Harvard University, and University of Wisconsin. 35

Chakraborty, A., A. Citanna, and M. Ostrovsky (2010): "Two-Sided Matching with Interdependent Values," *Journal of Economic Theory*, 145(1), 85–105. 2, 35

Cole, H. L., G. J. Mailath, and A. Postlewaite (2001a): "Efficient Non-Contractible Investments in Finite Economies," *Advances in Theoretical Economics*, 1(1), Article 2, http://www.bepress.com/bejte/advances/vol1/iss1/art2. 4

——— (2001b): "Efficient Non-Contractible Investments in Large Economies," *Journal of Economic Theory*, 101(2), 333–373. 4

Crawford, V. P., and E. M. Knoer (1981): "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, 49(2), 437–450. 5, 11

Dizdar, D., and B. Moldovanu (2012): "Surplus Division and Efficient Matching," University of Bonn. 34

Dutta, B., and R. Vohra (2005): "Incomplete Information, Credibility and the Core," *Mathematical Social Sciences*, 50(2), 148–165. 32

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. 37

Ehlers, L., and J. Masso (2007): "Incomplete Information and Singleton Cores in Matching Markets," *Journal of Economic Theory*, 1(36), 587–600. 35

Forges, F. (1994): "Posterior Efficiency," *Games and Economic Behavior*, 6(2), 238–261. 32

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69(1), 9–15. 1, 3, 6, 33

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. 12

Gillies, D. B. (1959): "Solutions to General Non-Zero-Sum Games," in *Contributions to the Theory of Games IV. (Annals of Mathematics Studies 40)*, ed. by R. D. Luce, and A. W. Tucker, pp. 47–85. Princeton University Press, Princeton. 3

Hatfield, J. W., S. D. Kominers, A. Nichifor, M. Ostrovsky, and A. Westkamp (2013): "Stability and Competitive Equilibrium in Tading Networks," Stanford University, University of Chicago, University of St. Andrews, Stanford University, and University of Bonn. 29

Holmström, B., and R. B. Myerson (1983): "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51(6), 1799–1819. 32

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. 35

Kojima, F., and M. U. Ünver (2008): "Random Paths to Pairwise Stability in Many-to-Many Matching Markets: A Study on Market Equilibration," *International Journal of Game Theory*, 36(3–4), 473–488. 33

Lauermann, S., and G. Nöldeke (2012): "Stable Marriages and Search Frictions," Mimeo. 3, 33, 34

Lee, S.-H. (2004): "Early Admission Program: Does It Hurt Efficiency?," Ph.D. thesis, University of Pennsylvania, Ch. 1. 35

Mailath, G. J., A. Postlewaite, and L. Samuelson (2012): "Premuneration Values and Investments in Matching Markets," PIER Working Paper No. 12-008, University of Pennsylvania. 4, 5, 36

——— (2013): "Pricing and Investments in Matching Markets," *Theoretical Economics*, forthcoming. 4, 5, 36

Milgrom, P. R., and N. Stokey (1982): "Information, Trade, and Common Knowledge," *Journal of Economic Theory*, 26(1), 17–27. 12

Myerson, R. B. (1995): "Sustainable Matching Plans with Adverse Selection," *Games and Economic Behavior*, 9(1), 35–65. 33

——— (2007): "Virtual Utility and the Core for Games with Incomplete Information," *Journal of Economic Theory*, 136(1), 260–285. 32

Nagypal, E. (2004): "Optimal Application Behavior with Incomplete Information," Mimeo. 35

Pearce, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52(4), 1029–50. 12

Perry, M., and P. J. Reny (1994): "A Noncooperative View of Coalition Formation and the Core," *Econometrica*, 62(4), 795–817. 3

Radner, R. (1979): "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica*, 45(3), 655–678. 29

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7(4), 617–628. 1, 33

——— (1984): "Misrepresentation and Stability in the Marriage Problem," *Journal of Economic Theory*, 34(2), 383–387. 33

——— (1989): "Two-Sided Matching with Incomplete Information about Others' Preferences," *Games and Economic Behavior*, 1(2), 191–209. 34

Roth, A. E., and M. A. O. Sotomayer (1990): *Two-Sided Matching*. Cambridge University Press, Cambridge. 1

Roth, A. E., and J. H. Vande Vate (1990): "Random Paths to Stability in Two-Sided Markets," *Econometrica*, 58(6), 1475–1480. 33

Serrano, R., and R. Vohra (2007): "Information Transmission in Coalitional Voting Games," *Journal of Economic Theory*, 134(1), 117–137. 32

Shapley, L. S., and M. Shubik (1971): "The Assignment Game I: The Core," *International Journal of Game Theory*, 1(1), 111–130. 1, 5, 6, 7, 11

Wilson, R. (1978): "Information, Efficiency, and the Core of an Economy," *Econometrica*, 46(4), 807–816. 32

Yenmez, M. B. (2013): "Incentive Compatible Matching Mechanisms: Consistency with Various Stability Notions," *American Economic Journal: Microeconomics*, forthcoming. 32

Online Appendix for Stable Matching with Incomplete Information

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

June 17, 2013

Contents

1 Proofs for Section 4.1.3

1.1 Information-Revealing Prices

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.

Lemma O.1 *Suppose Assumptions 1 and 3 hold, 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*

payment p^*, then for any firm with type $f < f^*$, there exists $\varepsilon > 0$ such that for any $p \in (\nu_{w^*f^*} + p^* - \nu_{w^*f},\ \nu_{w^*f^*} + p^* - \nu_{w^*f} + \varepsilon]$,

$$\nu_{wf} + p > \nu_{wf^*} + p^*, \qquad \textit{for all } w \geq w^*, \tag{O.1}$$
$$\nu_{wf} + p \geq 0, \qquad \textit{for all } w \geq w^*, \textit{ and} \tag{O.2}$$
$$\nu_{wf} + p \leq \nu_{wf^*} + p^*, \qquad \textit{for all } w < w^*. \tag{O.3}$$

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

$$\nu_{wf} + p > 0, \qquad \textit{for all } w \geq w^*, \textit{ and}$$
$$\nu_{wf} + p \leq 0, \qquad \textit{for all } w < w^*.$$

Proof. Define

$$p^\varepsilon := \nu_{w^*f^*} + p^* - \nu_{w^*f} + \varepsilon, \tag{O.4}$$

where $\varepsilon > 0$ will be determined later. The first required inequality (O.1) 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 submodularity (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 (O.4),

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

proving (O.2).

After substituting for $p = p^\varepsilon$ defined in (O.4), the inequality (O.3) 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 submodularity (since $f < f^*$). Inequalities (O.1)–(O.3) 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. ■

1.2 Proof of Proposition 4

1.2.1 Preliminaries

Constrained Efficiency We begin by formulating an inductive notion of efficiency. As before, we write the finite set of possible worker and firm types as $W = \{w^1, w^2, \ldots, w^K\}$ and $F = \{f^1, f^2, \ldots, f^L\}$, 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. For any matching function μ, denote by I_μ the set of matched workers and by J_μ $(= \mu(I_\mu))$ the set of matched firms. By definition, $|I_\mu| = |J_\mu|$.

Definition O.1 *A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* constrained efficient on $W' \subset W$ *if*

$$\sum\nolimits_{i \in \mathbf{w}^{-1}(W') \cap I_\mu} \left[\nu_{\mathbf{w}(i)\mathbf{f}(\mu(i))} + \phi_{\mathbf{w}(i)\mathbf{f}(\mu(i))}\right] = \max_{\mu' \in M} \sum\nolimits_{i \in \mathbf{w}^{-1}(W') \cap I_\mu} \left[\nu_{\mathbf{w}(i)\mathbf{f}(\mu'(i))} + \phi_{\mathbf{w}(i)\mathbf{f}(\mu'(i))}\right],$$

where M *is the set of one-to-one functions from* $\mathbf{w}^{-1}(W') \cap I_\mu$ *onto* $\mu\left(\mathbf{w}^{-1}(W')\right) \cap J_\mu$.[1]

In this definition, M consists of all possible matching functions between $\mathbf{w}^{-1}(W') \cap I_\mu$ and $\mu\left(\mathbf{w}^{-1}(W') \cap I_\mu\right) = \mu(\mathbf{w}^{-1}(W')) \cap J_\mu$ with no agent in these two sets left unmatched. Hence, constrained efficiency might violate individual rationality; e.g., it could be that a matched worker-firm pair generates a negative surplus. The following observation follows immediately from the definition of submodularity.

Lemma O.2 *A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is constrained efficient on* $W' \subset W$ *if and only if the matching outcome is negative assortative (i.e., for all* $i, i' \in I$ *such that* $\mu(i), \mu(i') \in J$*, if* $\mathbf{w}(i) < \mathbf{w}(i')$*, then* $\mathbf{f}(\mu(i)) \geq \mathbf{f}(\mu(i'))$*).*

Definition O.2 *For* $1 \leq k < K$*, a matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* k^{th}-order constrained efficient *if, for all* $w > w^k$ *and* $w \in W$*,* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is constrained efficient on $\{w^1, \ldots, w^k, w\}$.

The following observation explores the submodularity assumption and is useful in our inductive proofs.

[1] We adopt the convention that M is empty if $\mathbf{w}^{-1}(W')$ is empty, and that a summation over a empty set equals to 0.

Lemma O.3 *A matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is* $(k+1)^{th}$*-order constrained efficient if and only if it is* k^{th}*-order constrained efficient and for all* $w > w^{k+1}$*, it is constrained efficient on* $\{w^{k+1}, w\}$.

Proof. The "only if" parts are immediate by definition. "If": suppose $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is k^{th}-order constrained efficient. Consider any $w > w^{k+1}$. If $\mathbf{w}(i) \neq w^{k+1}$ and $\mathbf{w}(i) \neq w$ for all $i \in I_\mu$, then trivially, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is constrained efficient on $\{w^1, \ldots, w^k, w^{k+1}, w\}$. Suppose $\mathbf{w}(i) = w^{k+1}$ and $\mathbf{w}(i') = w$ for some $i, i' \in I_\mu$. By assumption, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is constrained efficient on $\{w^{k+1}, w\}$, and hence it follows from Lemma O.2 that $\mathbf{f}(\mu(i)) \geq \mathbf{f}(\mu(i'))$. By Lemma O.2 again, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is constrained efficient on $\{w^1, \ldots, w^k, w^{k+1}, w\}$. Hence, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is $(k+1)^{\text{th}}$-order constrained efficient. ■

Unmatched Agents

Lemma O.4 *Suppose* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is incomplete information stable.*

1. *There does not exist* $i, i' \in I$ *such that* $\mu(i) = \varnothing \neq \mu(i')$ *and* $\mathbf{w}(i) > \mathbf{w}(i')$*,*
2. *there does not exist* $j, j' \in J$ *such that* $\mu^{-1}(j) = \varnothing \neq \mu^{-1}(j')$ *and* $\mathbf{f}(j) > \mathbf{f}(j')$*, and*
3. *there does not exist* $i \in I$ *and* $j \in J$ *such that* $\mu(i) = \varnothing = \mu^{-1}(j)$ *and* $\nu_{\mathbf{w}(i)\mathbf{f}(j)} + \phi_{\mathbf{w}(i)\mathbf{f}(j)} > 0$.

Proof. Statement 1: Suppose there exist $i, i' \in I$ such that $\mu(i) = \varnothing \neq \mu(i')$ and $\mathbf{w}(i) > \mathbf{w}(i')$. We claim that $(i, \mu(i'))$ can form a blocking pair with a payment $p = -\nu_{\mathbf{w}(i)\mathbf{f}(\mu(i'))} + \varepsilon$ for some small enough $\varepsilon > 0$. To see this, note that by Lemma O.1, worker i receives a positive payoff and reveals that his type is at least $\mathbf{w}(i)$ if ε is small enough. The expected payoff of firm $\mu(i')$ from this deviation is at least

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

Since $\mathbf{w}(i) > \mathbf{w}(i')$, Assumption 1 implies that $\phi_{\mathbf{w}(i)\mathbf{f}(\mu(i'))} + \nu_{\mathbf{w}(i)\mathbf{f}(\mu(i'))} - \varepsilon > \phi_{\mathbf{w}(i')\mathbf{f}(\mu(i'))} + \nu_{\mathbf{w}(i')\mathbf{f}(\mu(i'))}$ for a small ε. But $\phi_{\mathbf{w}(i')\mathbf{f}(\mu(i'))} + \nu_{\mathbf{w}(i')\mathbf{f}(\mu(i'))}$ is the total surplus in a match $(i', \mu(i'))$. So firm $\mu(i')$ finds this deviation profitable.

Statement 2: Suppose there exist $j, j' \in J$ such that $\mu^{-1}(j) = \varnothing \neq \mu^{-1}(j')$ and $\mathbf{f}(j) > \mathbf{f}(j')$. We claim $\left(\mu^{-1}(j'), j\right)$ form a blocking pair with

payment $p = \mathbf{p}_{\mu^{-1}(j'),j'} + \varepsilon$ for some $\varepsilon > 0$. Observe first that worker $\mu^{-1}(j')$ receives a strictly higher payoff in this block, since ν is weakly monotonic in f and so

$$\nu_{\mathbf{w}(\mu^{-1}(j'))\mathbf{f}(j)} + \mathbf{p}_{\mu^{-1}(j'),j'} + \varepsilon > \nu_{\mathbf{w}(\mu^{-1}(j'))\mathbf{f}(j')} + \mathbf{p}_{\mu^{-1}(j'),j'}.$$

Although firm j may be uncertain about the type of worker $\mu^{-1}(j')$, the firm knows that the individual rationality for firm j' and the strict monotonicity of ϕ in f imply

$$0 \leq \phi_{\mathbf{w}(\mu^{-1}(j'))\mathbf{f}(\mu(j'))} - \mathbf{p}_{\mu^{-1}(j'),j'} < \phi_{\mathbf{w}(\mu^{-1}(j'))\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j'),j'}.$$

Therefore for ε sufficiently small, firm j gets a strictly positive payoff in the matching with worker $\mu^{-1}(j')$ at price $p = \mathbf{p}_{\mu^{-1}(j'),j'} + \varepsilon$.

Statement 3: Suppose there exist $i \in I$ and $j \in J$ such that $\mu(i) = \varnothing \neq \mu^{-1}(j)$ and $\nu_{\mathbf{w}(i)\mathbf{f}(j)} + \phi_{\mathbf{w}(i)\mathbf{f}(j)} > 0$. We claim (i,j) form a blocking pair with payment $p = -\nu_{\mathbf{w}(i)\mathbf{f}(j)} + \varepsilon$ for small enough ε. Worker i's payoff is positive from this deviation, and by Lemma O.1, firm j knows the block is only possible if worker i's type is at least $\mathbf{w}(i)$. Assumption 1 implies that firm j's payoff is at least $\nu_{\mathbf{w}(i)\mathbf{f}(j)} + \phi_{\mathbf{w}(i)\mathbf{f}(j)} - \varepsilon$ which is strictly positive for ε sufficiently small. ■

1.2.2 Completion of The Proof of Proposition 4

Constrained Efficiency Lemma O.5 and Lemma O.6 inductively show that every incomplete information stable matching outcome is constrained efficient.

Lemma O.5 *If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, then $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is first-order constrained efficient.*

Proof. Suppose to the contrary that some $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$ is not first-order constrained efficient. Then by Lemma O.2, there exist two workers, say 1 and 2, such that $\mathbf{w}(2) > \mathbf{w}(1) = w^1$ and $\mathbf{f}(\mu(2)) > \mathbf{f}(\mu((1)) \neq \varnothing$.

Claim O.1 *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*

$$\phi_{\mathbf{w}(2),\mathbf{f}(\mu(1))} + \nu_{\mathbf{w}(2),\mathbf{f}(\mu(1))} \leq \nu_{\mathbf{w}(2),\mathbf{f}(\mu(2))} + \mathbf{p}_{2,\mu(2)} + \phi_{\mathbf{w}(1),\mathbf{f}(\mu(1))} - \mathbf{p}_{1,\mu(1)}. \quad \text{(O.5)}$$

Proof. Consider a match by worker 2 and firm $\mu(1)$ with payment

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

for small $\varepsilon > 0$. By Lemma O.1, this match is only attractive to worker 2 if his type is $\mathbf{w}(2)$ or higher. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$, firm $\mu(1)$ must not be better off in this match. Hence,

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

Substituting for p,

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

implying (O.5). □

Claim O.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*

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

Proof. If the inequality in (B.6) did not hold, we can find $q \in \mathbb{R}$ such that

$$\nu_{\mathbf{w}(1),\mathbf{f}(\mu(2))} + q > \nu_{\mathbf{w}(1),\mathbf{f}(\mu(1))} + \mathbf{p}_{1,\mu(1)} \quad \text{and} \tag{O.6}$$

$$\phi_{\mathbf{w}(1),\mathbf{f}(\mu(2))} - q > \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}. \tag{O.7}$$

Since ϕ is weakly increasing and $\mathbf{w}(1) = w^1$ is the smallest type, (O.7) implies

$$\min_{w \in W} \phi_{w,\mathbf{f}(\mu(2))} - q > \phi_{\mathbf{w}(2),\mathbf{f}(\mu(2))} - \mathbf{p}_{2,\mu(2)}. \tag{O.8}$$

Hence, (O.6) and (O.8) imply $(1, \mu(2))$ is a blocking pair, contradicting $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^1$. □

Finally, we combine Claims O.1 and O.2. Adding the two inequalities, we obtain

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

Since $\mathbf{w}(1) < \mathbf{w}(2)$ and $\mathbf{f}(\mu(2)) > \mathbf{f}(\mu(1))$, this inequality contradicts the strict submodularity of $\nu + \phi$. ■

Lemma O.6 *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 constrained efficient.*

Proof. We proceed by induction. Suppose the claim holds for some $k \geq 1$ (from Lemma O.5, 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 constrained efficient. There then exist two workers i and i' such that worker i's type is $\mathbf{w}(i) = w^{k+1} < \mathbf{w}(i')$ and $\varnothing \neq \mathbf{f}(\mu(i)) < \mathbf{f}(\mu(i'))$.

Claim O.3 *If $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1}$, $\mathbf{w}(i) = w^{k+1} < \mathbf{w}(i')$ and $\varnothing \neq \mathbf{f}(\mu(i)) < \mathbf{f}(\mu(i'))$, then*

$$\nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} + \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i))} \leq \nu_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} + \mathbf{p}_{i',\mu(i')} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i))} - \mathbf{p}_{i,\mu(i)}. \quad \text{(O.9)}$$

Proof. Consider a match by worker i' and firm $\mu(i)$ with payment

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

for small $\varepsilon > 0$. By Lemma O.1, this match is only attractive to worker i' if his type is at least $\mathbf{w}(i')$. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1}$, the match $(i', \mu(i))$ 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 ϕ and $\mathbf{w}(i') \leq w$, 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 (O.9). □

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

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} \leq \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} + \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \quad \text{(O.10)}$$

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

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} + q > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \quad \text{and} \quad \text{(O.11)}$$

$$\phi_{\mathbf{w}(i),\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \quad \text{(O.12)}$$

By monotonicity of ϕ, (O.12) implies

$$\phi_{w,\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}, \quad \text{for all } w \geq \mathbf{w}(i) = w^{k+1}. \quad \text{(O.13)}$$

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

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

We claim that for any $\mathbf{w}' \in \Omega'$, $\mathbf{w}'(i) \geq w^{k+1}$. 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 constrained efficient, contradicting the assumption that $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \Sigma^k$.

It then follows from (O.13) that

$$\min_{\mathbf{w}' \in \Omega'} \phi_{\mathbf{w}'(i),\mathbf{f}(\mu(i'))} - q > \phi_{\mathbf{w}(i'),\mathbf{f}(\mu(i'))} - \mathbf{p}_{i',\mu(i')}. \quad \text{(O.14)}$$

Hence, from (O.11) and (O.14), the unmatched pair $(i, \mu(i'))$ at payment q can form a blocking pair. A contradiction. □

Summing (O.9) and (O.10), 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 submodularity of $\nu + \phi$.

■

Efficiency and Constrained Efficiency By definition, efficiency implies constrained efficiency, but the converse is not true without further assumptions. As shown in the example, leaving some agents unmatched, or creating more matched pairs, could improve efficiency. It follows from Lemma O.4 that in a stable matching outcome, all the unmatched workers (firms, resp.)

Figure O.1: An illustration of the derivation of (O.15).

must have lower realized types than the matched workers (firms, resp.) and no ex post surplus can be generated by unmatched agents.

Without loss, we assume $\mathbf{w}(i)$ is increasing in i. Suppose the incomplete information stable matching outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is constrained efficient but not efficient. Lemma O.4 implies that all unmatched agents have types lower than those in I_μ and J_μ. Let μ^* be an efficient matching. Since efficiency implies constrained efficiency, Lemma O.4 implies that all unmatched agents under μ^* have types lower than those of I_{μ^*} and J_{μ^*}. Therefore, without loss of generality, we can assume either $I_\mu \subsetneq I_{\mu^*}$ and $J_\mu \subsetneq J_{\mu^*}$, or $I_{\mu^*} \subsetneq I_\mu$ and $J_{\mu^*} \subsetneq J_\mu$. In addition, it follows from Lemma O.2 that we can assume that, without loss of generality, $\mu^*(i) \neq \mu(i)$ for each worker $i \in I_{\mu^*}$.

Under the hypothesis that all matches yield a positive surplus, it is immediate that are no unmatched pairs of workers and firms under μ^*, and so we must have $I_\mu \subsetneq I_{\mu^*}$ and $J_\mu \subsetneq J_{\mu^*}$.

Lemma O.7 *Suppose $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is a constrained-efficient incomplete-information stable matching outcome. Suppose $I_\mu \subsetneq I_{\mu^*}$ and $J_\mu \subsetneq J_{\mu^*}$. Then $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is efficient.*

Proof. We claim that for each $i \in I_\mu$, it must be that

$$\mathbf{f}(\mu^*(i)) \leq \mathbf{f}(\mu(i)). \tag{O.15}$$

To see that (O.15) holds, note that by assumption I_{μ^*} is obtained from I_μ by adding some lower worker types. Lemma O.2 implies that those added low types must match with high type firms under μ^*, and hence $i \in I_\mu$ will be rematched to lower firms under μ^* (see Figure O.1).

In the statement of the next claim, (O.16) is well-defined for workers $i \in I_{\mu^*} \setminus I_\mu$ using the convention $\nu_{\mathbf{w}(i),\varnothing} = \mathbf{p}_{i,\varnothing} = 0$.

Claim O.5 *Suppose the matching outcome* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ *is incomplete information stable. If* $I_\mu \subsetneq I_{\mu^*}$, $J_\mu \subsetneq J_{\mu^*}$, *and* $\mathbf{f}(\mu(i)) \neq \mathbf{f}(\mu^*(i))$ *for all* $i \in I_\mu$, *then for each* $i \in I_{\mu^*}$,

$$\nu_{\mathbf{w}(i),\mathbf{f}(\mu^*(i))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu^*(i))} \leq \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} + \phi_{\mathbf{w}(\mu^{-1}(\mu^*(i))),\mathbf{f}(\mu^*(i))} - \mathbf{p}_{\mu^{-1}(\mu^*(i)),\mu^*(i)}. \quad \text{(O.16)}$$

Proof. Since $\mathbf{f}(\mu(i)) \neq \mathbf{f}(\mu^*(i))$ for all $i \in I_\mu$, (O.15) implies $\mathbf{f}(\mu^*(i)) < \mathbf{f}(\mu(i))$ for all $i \in I_\mu$. By Lemma O.1, under μ, for each worker $i \in I_\mu$, the match with firm $\mu^*(i)$ with payment

$$p := \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} - \nu_{\mathbf{w}(i),\mathbf{f}(\mu^*(i))} + \varepsilon, \quad \text{(O.17)}$$

for $\varepsilon > 0$ small is only profitable if his type is at least $\mathbf{w}(i)$.

In addition, under μ, each worker $i \in I_{\mu^*} \setminus I_\mu$ is unmatched, and hence (by Lemma O.1 again) there are matches for such workers i with firm $\mu^*(i) \neq \varnothing$ with payment (O.17) that are only profitably if his type is at least $\mathbf{w}(i)$.

Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is stable, for each $i \in I_{\mu^*}$, firm $\mu^*(i)$ must not find any such match profitable, that is,

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

Substituting p and take $\varepsilon \to 0$, we obtain (O.16). □

Summing (O.16) over all $i \in I_{\mu^*}$, the payments cancel, yielding

$$\sum\nolimits_{i \in I_{\mu^*}} \left(\nu_{\mathbf{w}(i),\mathbf{f}(\mu^*(i))} + \phi_{\mathbf{w}(i),\mathbf{f}(\mu^*(i))}\right) \leq \sum\nolimits_{i \in I_{\mu^*}} \left(\nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} + \phi_{\mathbf{w}(\mu^{-1}(\mu^*(i))),\mathbf{f}(\mu^*(i))}\right),$$

contradicting the hypothesized inefficiency of $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$.

Claim O.5 requires $\mathbf{f}(\mu(i)) \neq \mathbf{f}(\mu^*(i))$ for all $i \in I_\mu$ in order to apply Lemma O.1 for each such i. To illustrate the issue that arises if $\mathbf{f}(\mu(i)) = \mathbf{f}(\mu^*(i))$ for some $i \in I_\mu$, consider the matchings in Figure O.2. In this example, the argument in the proof of Claim O.5 cannot be used because under μ, there is no match for the type 2 worker with another firm whose type is the same as his currently matched firm that is only profitable if his type is at least 2. However, in this case, to compare the efficiency of μ and μ^*, we only need to consider where they differ, that is, we only need to look at the two matchings illustrated in Figure O.3.

Claim O.5 applies to $\hat{\mu}$ and $\hat{\mu}^*$, and the same conclusion holds. We omit the obvious formal argument that requires additional notation. ■

Figure O.2: An illustration of the case ruled out in Claim O.5.

Figure O.3: The relevant part of the matching from Figure O.2.

2 Example and Proofs for Section 5

2.1 The Example Illustrating Multiple Rounds

Consider n firms and n workers: $W = F = \{1, ..., n\}$. Worker types are drawn from the set of all permutations. Worker premuneration values are identically zero, $\nu_{wf} \equiv 0$. Firm premuneration values are given by

$$\phi_{wf} = \begin{cases} wf, & \text{if } w \leq f, \\ f^2, & \text{if } w > f. \end{cases}$$

Consider the price matrix $\mathbf{P} = \mathbf{0}$, which assigns a price of zero to every match. Any price-taking matching outcome is then individually rational. In the matching given in Figure O.4, each firm is matched with the worker of the same index, except for an inversion in the match of the top two firms and workers. We show this outcome is not price sustainable. The only potentially profitable deviation in this matching outcome is for the type n firm to buy the type n worker; all workers and all other firms are getting the most they could obtain under the constant price matrix matrix $\mathbf{P} = \mathbf{0}$. However, under incomplete information, the type-n firm does not know which worker is type n, and hence this matching outcome is Ψ^0-sustainable. In fact, for the same reason, every matching outcome that matches the type 1 worker with

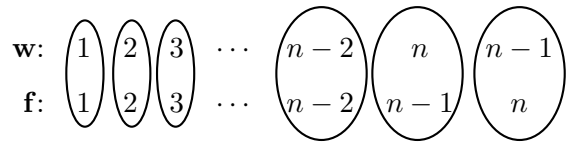

Figure O.4: The matching illustrating the need for many rounds of iteration in Definition 10.

the type 1 firm is Ψ^0-sustainable. Moreover, Ψ^1, the set of Ψ^0-sustainable outcomes, coincides with the set of matching outcomes that match the type 1 worker with the type 1 firm. To see this, it is enough to notice that any type $j > 1$ firm, if matched with the type 1 worker, would profitably deviate to purchase any other worker, whose type (given that types are drawn from the set of permutations) must be larger than 1. Now restricting attention to the set Ψ^1, it must be that the set of Ψ^1-sustainable outcomes coincides with the set of matching outcomes that match type i workers with type i firms for $i = 1, 2$, a subset of Ψ^1. To see this, note that any firm $j > 2$, if matched with the type 2 worker, would profitably deviate to any worker other than the one matched with the type 1 firm (since the latter must be type 1, given that we are looking at Ψ^1). Iterating this argument, we conclude that Ψ^{k+1}, the set of Ψ^k-competitive outcomes, coincides with the set of matching outcomes that match type i workers with type i firms for $i = 1, 2, ..., k + 1$. Hence, the candidate matching outcome we constructed above is in Ψ^{n-2}, but not in Ψ^{n-1}.

2.2 Proof of Lemma 3

Take any set of price-taking matching outcomes, C, that is self-sustaining. Then $C \subset \Psi^0$, where Ψ^0 is the set of individually rational outcomes. Since each $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \in C$ is C-competitive, it follows from Definition 9 that such $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ is Ψ^0-sustainable, i.e. $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \in \Psi^1$. Hence, $C \subset \Psi^1$. Using again the fact that each $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f}) \in C$ is C-sustainable and $C \subset \Psi^1$, we obtain $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ is Ψ^1-sustainable and hence $C \subset \Psi^2$. We conclude that $C \subset \Psi^\infty$ by iterating this argument. That is, Ψ^∞ contains any set that is self-sustainable. Moreover, by definition, Ψ^∞ is Ψ^∞-sustainable and hence is self-sustainable. ■

2.3 Proof of Proposition 7

The proof uses the fixed-point characterizations of stability and price sustainability. Fix an incomplete-information stable outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ and let E be a self-stabilizing set that contains it. In particular, by part 4 of Lemma 1, we can take E such that it contains matching outcomes with the same allocation $(\mu, \mathbf{p})$. Our goal is to extend $\mathbf{p}$ to $\mathbf{P}$ in an appropriate way and then show $(\mu, \mathbf{P}, \mathbf{w}, \mathbf{f})$ is price sustainable. To do so, we extend the entire self-stabilizing set E to a set of price-taking outcomes C and then show C is self-sustainable.

2.3.1 Step 1. Constructing C

Let E be a self-stabilizing set. For each element of E, $(\mu, \mathbf{p}, \tilde{\mathbf{w}}, \mathbf{f})$, we extend $(\mu, \mathbf{p}, \tilde{\mathbf{w}}, \mathbf{f})$ to $(\mu, \tilde{\mathbf{P}}, \tilde{\mathbf{w}}, \mathbf{f})$, and define C as the resulting set of price-taking matching outcomes.

Consider a worker-firm pair $(i, j) \in I \times J$ such that $j \neq \mu(i)$. Since E is self-stabilizing, there does *not* exist $p \in \mathbb{R}$ such that

$$\nu_{\tilde{\mathbf{w}}(i),\mathbf{f}(j)} + p > \nu_{\tilde{\mathbf{w}}(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \tag{O.18}$$

and

$$\left(\min_{\mathbf{w}' \in \Omega(\tilde{\mathbf{w}}(\mu^{-1}(j)),i,j,p)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)} \right) - p > \phi_{\tilde{\mathbf{w}}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}, \tag{O.19}$$

where $\Omega(w, i, j, p)$ is the set of worker type assignments $\mathbf{w}'$ satisfying

$$(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in E, \tag{O.20}$$

$$\mathbf{w}'(\mu^{-1}(j)) = w, \tag{O.21}$$

and

$$\nu_{\mathbf{w}'(i),\mathbf{f}(j)} + p > \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))} + \mathbf{p}_{i,\mu(i)}. \tag{O.22}$$

For each firm j, let $\hat{\Omega}(w, j)$ denote the set of worker type assignments satisfying (O.20) and (O.21). Note that $\Omega(w, i, j, p) \subset \hat{\Omega}(w, j)$ for any $p \in \mathbb{R}$, since $\Omega(w, i, j, p)$ is further restricted by the requirement that worker i is purportedly willing to form a blocking pair at price p (allowing firm j to draw inferences about i's type).

Claim O.6 *For any worker-firm pair (i, j),*

$$\min_{\mathbf{w}' \in \hat{\Omega}(\tilde{\mathbf{w}}(\mu^{-1}(j)),j)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)} + \max_{\mathbf{w}' \in \hat{\Omega}(\tilde{\mathbf{w}}(\mu^{-1}(j)),j)} \left(\nu_{\mathbf{w}'(i),\mathbf{f}(j)} - \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))} \right) \leq \mathbf{p}_{i,\mu(i)} + \phi_{\tilde{\mathbf{w}}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}. \tag{O.23}$$

Proof. Suppose to the contrary that the inequality does not hold. Then there exists $\mathbf{w}^* \in \hat{\Omega}\left(\tilde{\mathbf{w}}(\mu^{-1}(j)), j\right)$ such that

$$\left(\min_{\mathbf{w}' \in \hat{\Omega}(\tilde{\mathbf{w}}(\mu^{-1}(j)),j)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)}\right) + \nu_{\mathbf{w}^*(i),\mathbf{f}(j)} - \nu_{\mathbf{w}^*(i),\mathbf{f}(\mu(i))} > \mathbf{p}_{i,\mu(i)} + \phi_{\tilde{\mathbf{w}}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}.$$

Since $\mathbf{w}^*(\mu^{-1}(j)) = \tilde{\mathbf{w}}(\mu^{-1}(j))$, this inequality is the same as

$$\left(\min_{\mathbf{w}' \in \hat{\Omega}(\mathbf{w}^*(\mu^{-1}(j)),j)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)}\right) + \nu_{\mathbf{w}^*(i),\mathbf{f}(j)} - \nu_{\mathbf{w}^*(i),\mathbf{f}(\mu(i))} > \mathbf{p}_{i,\mu(i)} + \phi_{\mathbf{w}^*(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}. \quad \text{(O.24)}$$

But (O.24) implies that there exists $p^* \in \mathbb{R}$ such that

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

and

$$\left(\min_{\mathbf{w}' \in \hat{\Omega}(\mathbf{w}^*(\mu^{-1}(j)),j)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)}\right) - p^* > \phi_{\mathbf{w}^*(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}. \quad \text{(O.25)}$$

Since $\Omega\left(\mathbf{w}^*(\mu^{-1}(j)), i, j, p^*\right) \subset \hat{\Omega}\left(\mathbf{w}^*(\mu^{-1}(j)), j\right)$, (O.25) implies that

$$\left(\min_{\mathbf{w}' \in \Omega(\mathbf{w}^*(\mu^{-1}(j)),i,j,p^*)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)}\right) - p^* > \phi_{\mathbf{w}^*(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j},$$

contradicting the nonexistence of a price p satisfying (O.18)–(O.19). □

The following is now an immediate implication of (O.23): there exists $\tilde{\mathbf{P}}_{ij}^{\tilde{\mathbf{w}}(\mu^{-1}(j))} \in \mathbb{R}$ such that

$$\left(\min_{\mathbf{w}' \in \hat{\Omega}(\tilde{\mathbf{w}}(\mu^{-1}(j)),j)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)}\right) - \tilde{\mathbf{P}}_{ij}^{\tilde{\mathbf{w}}(\mu^{-1}(j))} \leq \phi_{\tilde{\mathbf{w}}(\mu^{-1}(j)),\mathbf{f}(j)} - \mathbf{p}_{\mu^{-1}(j),j}. \quad \text{(O.26)}$$

and

$$\max_{\mathbf{w}' \in \hat{\Omega}(\tilde{\mathbf{w}}(\mu^{-1}(j)),j)} \left(\nu_{\mathbf{w}'(i),\mathbf{f}(j)} - \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))}\right) + \tilde{\mathbf{P}}_{ij}^{\tilde{\mathbf{w}}(\mu^{-1}(j))} \leq \mathbf{p}_{i,\mu(i)}.$$

The critical feature of $\tilde{\mathbf{P}}_{ij}^{\tilde{\mathbf{w}}(\mu^{-1}(j))}$ is that it only depends on $\tilde{\mathbf{w}}$ through the value of $\tilde{\mathbf{w}}(\mu^{-1}(j))$.

We are now in a position to extend the set of stable outcomes E to a set of price-taking outcomes C as follows. For each $(\mu, \mathbf{p}, \tilde{\mathbf{w}}, \mathbf{f}) \in E$, define an associated price-taking matching outcome $\left(\mu, \tilde{\mathbf{P}}, \tilde{\mathbf{w}}, \mathbf{f}\right)$ by

$$\tilde{\mathbf{P}}_{ij} = \begin{cases} \mathbf{p}_{ij}, & j = \mu(i), \\ \tilde{\mathbf{P}}_{ij}^{\tilde{\mathbf{w}}(\mu^{-1}(j))}, & \text{otherwise.} \end{cases} \tag{O.27}$$

We define C as the set of all price-taking matching outcomes derived from E in this way.

2.3.2 Step 2. The price sustainability of C

Fix a firm j and an outcome $(\mu, \tilde{\mathbf{P}}, \tilde{\mathbf{w}}, \mathbf{f}) \in C$, and consider the set $\Omega'(j)$ of worker type assignments $\mathbf{w}' \in \Omega$ for which there there exists $\mathbf{P}'$ such that

$$\left(\mu, \mathbf{P}', \mathbf{w}', \mathbf{f}\right) \in C,$$
$$\mathbf{w}'(\mu^{-1}(j)) = \tilde{\mathbf{w}}(\mu^{-1}(j)),$$

and

$$\mathbf{P}'_{i',\mu(i')} = \tilde{\mathbf{P}}_{i',\mu(i')} \text{ and } \mathbf{P}'_{i',j} = \tilde{\mathbf{P}}_{i',j}, \quad \forall i' \in I.$$

Observe that by the definition of C, if $(\mu, \tilde{\mathbf{P}}, \tilde{\mathbf{w}}, \mathbf{f})$, $(\mu, \mathbf{P}', \mathbf{w}', \mathbf{f}) \in C$, then $\mathbf{P}'_{i',\mu(i')} = \tilde{\mathbf{P}}_{i',\mu(i')} = \mathbf{p}_{i',\mu(i')}$ for any $i' \in I$; if in addition, $\mathbf{w}'(\mu^{-1}(j)) = \tilde{\mathbf{w}}(\mu^{-1}(j))$, then $\tilde{\mathbf{P}}_{i',j} = \mathbf{P}'_{i',j} = \mathbf{P}_{i',j}^{\tilde{\mathbf{w}}(\mu^{-1}(j))}$ for any $i' \in I$. Therefore,

$$\Omega'(j) = \hat{\Omega}\left(\tilde{\mathbf{w}}(\mu^{-1}(j)), j\right)$$

where $\hat{\Omega}(w, j)$ is defined just after (O.22). Hence, it follows from (O.26)–(O.27) that for any $i \in I$ and $j \in J$,

$$\left(\min_{\mathbf{w}' \in \Omega'(j)} \phi_{\mathbf{w}'(i),\mathbf{f}(j)}\right) - \tilde{\mathbf{P}}_{ij} \leq \phi_{\tilde{\mathbf{w}}(\mu^{-1}(j)),\mathbf{f}(j)} - \tilde{\mathbf{P}}_{\mu^{-1}(j),j}. \tag{O.28}$$

and

$$\max_{\mathbf{w}' \in \Omega'(j)} \left(\nu_{\mathbf{w}'(i),\mathbf{f}(j)} - \nu_{\mathbf{w}'(i),\mathbf{f}(\mu(i))}\right) + \tilde{\mathbf{P}}_{ij} \leq \tilde{\mathbf{P}}_{i,\mu(i)}. \tag{O.29}$$

Inequality (O.28) implies

$$\phi_{\mathbf{w}'(i),\mathbf{f}(j)} - \tilde{\mathbf{P}}_{ij} \leq \phi_{\mathbf{w}(\mu^{-1}(j)),\mathbf{f}(j)} - \tilde{\mathbf{P}}_{\mu^{-1}(j),j}$$

for some $\mathbf{w}' \in \Omega'(j)$. Since $(\mu, \tilde{\mathbf{P}}, \tilde{\mathbf{w}}, \mathbf{f}) \in C$, by definition, $\tilde{\mathbf{w}} \in \hat{\Omega}\left(\tilde{\mathbf{w}}(\mu^{-1}(j)), j\right) = \Omega'(j)$, and so (O.29) implies

$$\nu_{\tilde{\mathbf{w}}(i),\mathbf{f}(j)} + \tilde{\mathbf{P}}_{ij} \leq \nu_{\tilde{\mathbf{w}}(i),\mathbf{f}(\mu(i))} + \tilde{\mathbf{P}}_{i,\mu(i)}.$$

Hence, by Definition 11, C is self-sustainable.