Stable Matching under Forward-Induction Reasoning

Luciano Pomatto*

February 2018

Abstract

A standing question in the theory of matching markets is how to define stability under incomplete information. The crucial obstacle is that a notion of stability must include a theory of how beliefs are updated in a blocking pair. This paper proposes a novel epistemic approach. Agents negotiate through offers. Offers are interpreted according to the highest possible degree of rationality that can be ascribed to their proponents, in line with the principle of forward-induction reasoning.

This approach leads to a new definition of stability. The main result shows an equivalence between this notion and "incomplete-information stability", a cooperative solution concept recently put forward by Liu, Mailath, Postlewaite and Samuelson (2014), for markets with one-sided incomplete information.

The result implies that forward-induction reasoning leads to efficient matchings under standard supermodularity conditions. In addition, it provides an epistemic foundation for incomplete-information stability. The paper also shows new connections and distinctions between the cooperative and the epistemic approaches in matching markets.

1 Introduction

Over the past decades, models of matching markets have been applied to the design of college admissions, the analysis of housing markets, and the study of labor and marriage markets. In addition, a vast literature has substantially broadened our conceptual understanding of matching markets (see Roth (2002,2008) and Roth and Sotomayor (1990) for surveys on two-sided matching and their applications).

Much of the existing literature assumes complete information, i.e., that the value of a matching is entirely known to the relevant parties. However, incomplete information is arguably commonplace in most environments.

*California Institute of Technology. Email: luciano@caltech.edu. The paper is a substantial revision of an earlier paper of the same title. I am indebted to Alvaro Sandroni and Marciano Siniscalchi for many useful discussions. I thank Pierpaolo Battigalli, Yi-Chun Chen, George Mailath, Willemien Kets, Larry Samuelson, Rakesh Vohra and several seminar audiences for helpful comments on earlier drafts. All errors are my own.

The crucial difficulty in the study of matching markets with incomplete information lies in the notion of stability. Consider a job market where workers and firms are matched. Under complete information, a matching is *stable* if no pair of workers and firms are willing to reject the existing match to form more profitable partnerships. Consider now a market where there is uncertainty about the profitability of partnerships. Whether or not to leave the existing match is now a complex decision. This is true even if there are well specified ex-ante probabilities over the profitability of each partnership. One reason is that the actions taken to exit the default allocation (starting a negotiation, proposing an agreement, etc.) will typically reveal something about the parties involved. Another reason is that if the matching is to be deemed "stable", then such actions should be unexpected. Hence, agents must revise their beliefs based on zero probability events. So, under incomplete information, a theory of stability must also incorporate a novel theory of beliefs. This makes stability difficult to define, and raises fundamental methodological questions: given that some theory of belief revision is necessary, what type of assumptions on players' beliefs and thought processes can lead to stable matchings? What assumptions can lead to efficient allocations?

This paper considers a novel epistemic approach to two-sided matching markets with incomplete information. We study a class of markets where each agent on one side of the market (e.g., workers) has private information about characteristics of its members (for instance, their skills), that are payoff-relevant for both sides. Each worker is assumed to know his payoff-type and each firm knows the type of the worker it is matched to. Notably, agents are not required to share a common prior. Players' beliefs are assumed to satisfy a simpler "grain of the truth" assumption, which postulates that agents assign at least positive probability to the actual profile of payoff-types.

A default allocation is given. It specifies how workers are matched to firms and at what wages. Utility is transferable. Firms know the characteristics of the workers they are matched to in the default matching, and have the opportunity to negotiate away from the default. Negotiation is modelled as a noncooperative game and occurs through take-it-or-leave-it offers. If no offers are made, or all offers are rejected, then the default allocation is implemented.

The approach taken in this paper is deliberately in between cooperative and noncooperative. As in the classical study of stability and the core, we abstract way from the process by which a certain allocation is formed. At the same time, in order to formalize players' beliefs and thought processes, we model deviations from a given allocation through a noncooperative game.

Consider an agent, named Ann, who receives an offer from another agent, named Bob. Ann cannot know with certainty whether accepting the offer is profitable. She must reach this decision by updating her belief about Bob's characteristics from the fact that he made her an offer. Intuitively, Ann faces questions such as: what must be true about Bob for him to make this offer? What can we infer about Bob from the fact he is the only one who made me an offer? And so forth.

An additional consideration must also be made. For the matching to be stable, the default matching should be such that no offers are expected to be made. Hence, Bob's offer should be unexpected by everyone except him. Assume that Bob, under the default allocation, is matched to an agent, named Adam, who knows Bob's characteristics. This consideration leads Ann to an additional question: What inference should be made about Bob considering that Adam expected Bob to make no offers? Thus, in choosing her action, Ann should take into account that Adam did not expect Bob to make an offer to her.

The approach taken in this paper is to follow the idea that offers are interpreted according to the highest degree of sophistication that can be ascribed to those who make them. This is formalized by assuming that players behave accordingly to a notion of extensive-form rationalizability (Pearce, 1984) that builds upon the work of Battigalli and Siniscalchi (2002, 2007) on forward-induction reasoning. Stability is defined by imposing three requirements on players' actions and beliefs. Informally,

1. Agents are rational and abstain from making offers;
2. Players expect no offer to be made by other agents; and
3. In case a player deviates and makes an offer, the offer is interpreted according to the highest degree of strategic sophistication that can be ascribed to its proponent.

If all three requirements are satisfied, then the default allocation is said to be *stable under forward induction.* Rationality is defined by requiring players' actions to be optimal (given their beliefs) at every history they act. Requirement (2) is formalized by the assumption that players assign probability 1, at the beginning of the game, to the event that other players will not make offers.

The third requirement is crucial and expresses forward-induction reasoning. It is formalized through an iterative definition. Each player expects others to be rational and also expects others to believe, ex-ante, that no offer will be made. This belief is held at the beginning of the game and conditional on any offer, provided that the offer does not provide decisive proof against it. As a further step in their thought process, agents expect other players to believe in their opponents rationality and their surprise upon observing an offer. This more sophisticated belief is held at the beginning of the game and conditional on any history that does not contradict it. This iteration progresses through higher orders. Each step leads players to rationalize the observed behavior according to a higher degree of sophistication. Requirement (3) is formalized by taking the limit of this iteration.

The main result of this paper, Theorem 1, characterize the set of matching outcomes that are stable under forward induction. Perhaps surprisingly, this characterization leads to a solution concept that can be made operational and tractable. A matching outcome is stable under forward-induction if and only if it is *incomplete-information stable*, a cooperative notion recently introduced by Liu, Mailath, Postlewaite and Samuelson (2014). This notion satisfies two fundamental properties: existence and efficiency under standard supermodularity conditions.

The results of this paper show that stability under forward induction can be applied through a simple algorithm. They also provide a new connection between cooperative and noncooperative approaches in matching markets. In addition, the noncooperative approach taken in this paper allows for a specific understanding of what thought processes can lead to stability and efficiency in matching markets. By formalizing how players negotiate, it makes possible to provide explicit epistemic foundations for incomplete-information stability.

At the same time, the paper highlights important differences between the cooperative approach and the current approach based on forward-induction reasoning. One difference lies in the type of informational assumptions. Liu, Mailath, Postlewaite and Samuelson (2014) assume that the matching and the profile of wages are common knowledge. In this paper, we make the weaker hypothesis that workers' beliefs about other agents' payoff-types, matches and wages assign positive probability to the actual realization.

A second important difference is in the criterion by which firms evaluate the risk involved in matching with an agent of unknown type. A strict interpretation of incomplete-informations stability suggests firms choose whether or not to participate to a blocking pair by considering their worst-case payoff with respect to the uncertain payoff-type of the worker they are matching to. In this paper, agents are assumed to be expected utility maximizers.

Finally, while the main result of the paper shows an equivalence between the set of outcomes that are stable under the two notions of stability, the two solution concepts are remarkably different. Like stability under stability-under forward induction, incomplete-information stability is defined through an iterative elimination procedure. However, Liu, Mailath, Postlewaite and Samuelson (2014) do not provide an explicit epistemic characterization of incomplete-information stability. In addition, their solution concept is defined without reference to a noncooperative game.

To study more in detail the relation between the different notions of stability, in Theorems 2 and 3 we analyze the "finite order" implications of the two solution concepts. That is, we compare the set of matching outcomes that survive n steps of each elimination procedure. We show that the two solution concepts are not logically nested. In particular, the set of matching outcomes that survive n steps of one elimination procedure do not necessarily include (nor are necessarily included in) the set of outcomes that survive n steps of the other elimination procedure. This suggests that the two solution concepts, while leading to the same set of stable allocations, are motivated by different assumptions on agents' thought processes.

1.1 Related Literature

This paper is linked to several strands of the literature. Starting with Wilson (1978), notions of core under incomplete information have been introduced by Vohra (1999), Dutta and Vohra (2005), Serrano and Vohra (2007), de Clippel (2007), Myerson (2007), and Peivandi (2013), among others. The current paper shares some similarities with Serrano and Vohra (2007), where blocking coalitions are formed noncooperatively, as equilibrium outcomes of a voting game.

Matching under incomplete information has been studied by Roth (1989), Chade (2006), Ehlres and Masso (2007), Hoppe, Moldovanu, and Sela (2009), Chakraborty, Citanna, and Ostrovsky (2010), and Chade, Lewis, and Smith (2011), among others. Bikhchandani (2014) extends the analysis of Liu, Mailath, Postlewaite, and Samuelson (2014) to markets without transferable utility and presents a notion of stability for markets with two-sided incomplete information. Chen and Hu (2017) provide an alternative foundation for incomplete-information stability. They establish that any dynamic process that allows randomly chosen blocking pairs to rematch will converge almost surely to an allocation that is incomplete-information stable. Essential for their analysis is the assumption that uninformed agents evaluate blocking pairs according to a max-min criterion. In this paper, agents are assumed to be Bayesian expected utility maximizers. Finally, Liu (2017) introduces a definition of stability under incomplete information for a class of matching markets where agents on the uninformed side of the market do not know the types of the agents they are matched with, but instead share a common belief over types.

This paper builds upon the literature on forward-induction reasoning. Extensive form rationalizability was first introduced in Pearce (1984), while the best rationalization principle was first formalized in Battigalli (1996). Common strong belief in rationality was defined and characterized in Battigalli and Siniscalchi (2002), and in Battigalli and Siniscalchi (2003,2007) for games with payoff uncertainty. The implications of common strong belief in rationality are also studied in Battigalli and Friedenberg (2012) and Battigalli and Prestipino (2013). Battigalli and Siniscalchi (2003) show that in two-players signaling games common strong belief in rationality and a fixed distribution over messages and types provides a characterization of the set of self-confirming equilibria satisfying the iterated intuitive criterion (the result shares some similarities with the analysis in Sobel, Stole and Zapater, 1990). There are several significant differences between this paper and Battigalli and Siniscalchi (2003). In this paper, players do not share a common belief over the payoff-types of the informed players. In addition, unlike signaling games, the information structure of the blocking game we consider in this paper does not have a product structure. Finally, the main result of the paper, the characterization of Theorem 2, does not share similarities with other results in the literature.

The idea that players may rationalize past behavior has a long history in game theory.

The idea of forward induction goes back to Kohlberg (1981). Solution concepts expressing different forms of forward induction were introduced in Pearce, Kohlberg and Mertens (1986), Cho and Kreps (1987), Van Damme (1989), Reny (1992), Govindan and Wilson (2009), and Man (2012), among others.

2 Two-Sided Matching Markets

We consider a two-sided matching environment with transferable utility, following Crawford and Knoer (1981) and Liu, Mailath, Postlewaite and Samuelson (2014). A set of *agents* is divided in two groups, denoted by I and J. For concreteness, I is referred to as the set of *workers* and J as the set of *firms*. We assume $|I| \geq 2$. Each worker is endowed with a *payoff-type* belonging to a finite set W. Each firm $j \in J$ is also endowed with a payoff-type belonging to a finite set F. We denote by $\mathbf{w} \in W^I$ and $\mathbf{f} \in F^J$ the corresponding *profiles* of attributes.

A *matching function* is a map $\mu : I \to J \cup \{\emptyset\}$ that is injective on $\mu^{-1}(J)$. If $\mu(i) = j$ then worker i is hired by firm j. If $\mu(i) = \emptyset$ then worker i is unemployed. Similarly, if $\mu^{-1}(j) = \emptyset$ then no worker is hired by firm j. A worker is assigned to at most one firm and a firm can hire at most one worker.

A match between a worker of type w and a firm of type f gives rise, in the absence of monetary transfers, to a payoff of $\nu(w, f)$ for the worker and $\phi(w, f)$ for the firm. Following Mailath, Postlewaite and Samuelson (2013), we refer to ν and ϕ as *premuneration values*. The premuneration values of an unmatched worker or firm is equal to 0. In order to have a unified notation for both matched and unmatched agents, let $\nu(w, f_\emptyset) = 0$ for every $w \in W$ and $\phi(w_\emptyset, f) = 0$ for every $f \in F$.

Associated to a matching function is a payment scheme $\mathbf{p}$ specifying for each pair $(i, \mu(i))$ of matched agents a transfer $\mathbf{p}_{i,\mu(i)} \in \mathbb{R}$ from firm $\mu(i)$ to worker i. Unmatched workers receive no payments. We use the notation $\mathbf{p}_{i,\emptyset} = \mathbf{p}_{\emptyset,j} = 0$ for every i and j. Under the matching μ and payment scheme $\mathbf{p}$, the utility of worker i and firm j is given by

$$\nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)} \text{ and } \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$$

respectively.

A *matching outcome* is a tuple $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ specifying workers' and firms' payoff-types and an *allocation* $(\mu, \mathbf{p})$ consisting of a matching function and a payment scheme. A matching outcome is individually rational if it provides nonnegative payoff to all workers and firms.

2.1 Stability under Complete Information

An allocation $(\mu, \mathbf{p})$ is given. It will be referred to as the *default allocation*, or *status quo*. Agents have the opportunity to negotiate and abandon the status quo in favor of new partnerships, but if no agreement is reached, then the default allocation remains in place. If the matching outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is common knowledge, then this is the setting studied by Shapley and Shubik (1971) and Crawford and Knoer (1981).

Definition 1 A matching outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is *complete-information stable* if it is individually rational and there is no worker i, firm j and payment q such that

$$\begin{aligned} \nu(\mathbf{w}_i, \mathbf{f}_j) + q &> \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)} \text{ and} \\ \phi(\mathbf{w}_i, \mathbf{f}_j) - q &> \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}. \end{aligned}$$

Under complete information, an individually rational matching outcome fails to be stable if it is possible to find a worker i and firm j (i.e. a *blocking pair*) who can improve upon the status quo by forming a different and more profitable match at a wage q. As shown by Shapley and Shubik (1971), for any profiles $\mathbf{w}$ and $\mathbf{f}$ there always exists an allocation with the property that the resulting matching outcome is complete-information stable, and every stable outcome is efficient.

2.2 Incomplete Information

The standard framework is now altered by relaxing the assumption of complete information. We consider markets where agents have only partial information regarding other agents' types as well as the current allocation. We study markets with one-sided, interim, incomplete information.

We are given a finite set $\mathbf{M}$ of possible matching outcomes. We refer to $\mathbf{M}$ as the *market*. For simplicity, each $\mathbf{m} \in \mathbf{M}$ is assumed to be individually rational. Players' information about the matching outcome is modelled as a profile $(\mathcal{P}_k)_{k \in I \cup J}$ of information partition on $\mathbf{M}$. For every $\mathbf{m} \in \mathbf{M}$ and player k we denote by $\mathcal{P}_k(\mathbf{m}) \subseteq \mathbf{M}$ the information available to k when the actual outcome is $\mathbf{m}$.

Fix a matching outcome $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \mathbf{M}$. For every firm j, we assume

$$\mathcal{P}_j(\mathbf{m}) = \left\{(\tilde{\mathbf{w}}, \mathbf{f}, \mu, \mathbf{p}) \in \mathbf{M} : \tilde{\mathbf{w}}_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}\right\}.$$

Hence, each firm knows the current profile $\mathbf{f}$ of firms' types, the allocation $(\mu, \mathbf{p})$ and the type of the worker it is matched to, if any. Workers, on the other hand, are only required to possess minimal information about the environment. For every worker i define

$$\mathcal{P}_i^*(\mathbf{m}) = \left\{(\tilde{\mathbf{w}}, \mathbf{f}, \tilde{\mu}, \tilde{\mathbf{p}}) \in \mathbf{M} : \tilde{\mathbf{w}}_i = \mathbf{w}_i, \tilde{\mu}(i) = \mu(i) \text{ and } \tilde{\mathbf{p}}_{i,\mu(i)} = \mathbf{p}_{i,\mu(i)}\right\}.$$

That is, under the information partition $\mathcal{P}_i^*$, each worker i knows the profile $\mathbf{f}$, her payoff-type $\mathbf{w}_i$, her match $\mu(i)$, and her wage. We assume that for each worker, her information partition $\mathcal{P}_i$ satisfies $\mathcal{P}_i(\mathbf{m}) \subseteq \mathcal{P}_i^*(\mathbf{m})$ for every $\mathbf{m}$. That is, $\mathcal{P}_i^*$ is a *lower bound* on the amount of information available to i. This allows for a fairly general formulation.

In addition to the information specified by the partitions, agents entertain probabilistic beliefs about what they do not know. Beliefs will be described in Section 4.

3 The Blocking Game

This section introduces a simple noncooperative game by which players negotiate over new partnerships in order to abandon the status quo allocation. Negotiation occurs through take-it-or-leave-it offers.

3.1 Model

In this noncooperative game the set of players is $I \cup J$. We are given a matching outcome $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ belonging to $\mathbf{M}$. The game is played in two stages. In each stage, actions are played simultaneously and are observed by all players.

In the first stage, each worker i can *abstain* ("a") or *make an offer* (j, q), where j is a firm other than $\mu(i)$ and q belongs to Q, a fixed finite subset of $\mathbb{R}$. Informally, an offer (j, q) means that worker i is willing to break the status quo and form a new partnership with firm j at a wage q. The assumption of a discrete currency $Q \subseteq \mathbb{R}$ simplifies the exposition by avoiding measurability considerations. We assume Q to be a sufficiently fine grid.[1]

In the second stage, each firm that has received at least one offer chooses between rejecting all offers ("r") or accepting one offer of her choice.

Payoffs are defined as follows. For every offer (j, q) by worker i that has been accepted, call the resulting combination (i, j, q) a *blocking offer*. For every blocking offer (i, j, q), worker i is matched to firm j at a wage q and the two agents receive payoffs $\nu(\mathbf{w}_i, \mathbf{f}_j) + q$ and $\phi(\mathbf{w}_i, \mathbf{f}_j) - q$, respectively. If worker i is not part of a blocking offer but $\mu(i)$ is, then i receives a payoff of 0 (i.e., i becomes unmatched). Similarly, if firm j is not part of a blocking offer but $\mu^{-1}(j)$ is, then j receives a payoff of 0. All the other agents remain matched according to the original allocation $(\mu, \mathbf{p})$ and obtain the corresponding payoffs.

3.2 Discussion

The game has two features that play an important role in the analysis. The first is that offers are binding: an offer that is accepted is immediately implemented. The second is that inaction preserves the status quo. That is, if no offers are made then the original allocation $(\mu, \mathbf{p})$ is applied. Both features make the game close in spirit to assumptions that are implicit in the interpretation of the core under complete information (see, for instance, the discussion in Myerson, 1991).

It should be emphasized that in this game incomplete information is analyzed at the interim stage. In particular, there is no ex-ante stage at which workers plan their actions conditional on every realized matching outcome.

[1] See section A.1 for a formal statement of this assumption. The results are not sensitive to the particular specification of Q.

We now introduce some auxiliary notation that will be useful in what follows. Let H denote the set of all non-terminal histories and denote by $\varnothing$ the empty (or *initial*) history. Each history $h \in H$ other than $\varnothing$ describes what offers, if any, have been made and to what firms. For each firm, denote by $H_j \subseteq D$ the set of histories where j has received at least one offer. A strategy of worker i is an element $s_i \in \{a\} \cup (J \backslash \mu(i)) \times Q$. A strategy of firm j is represented by a function $s_j : H_j \to I \cup \{r\}$ with the property that if $s_j(h) \neq r$ then $s_j(h)$ belongs to the set of workers who made an offer to j at history h. The set of strategies of each player k is denoted by S_k. For every history h and player k we denote by $S_{-k}(h)$ the set of all strategies in S_{-k} that lead to history h for some $s_k \in S_k$.

4 Rationalizability

At the core of our analysis is a notion of rationalizability based on the following requirements: (i) players are rational, (ii) they do not expect others to break away from the status quo, and (iii) informally, at every history they mantain the highest possible degree of belief in hypotheses (i) and (ii). We first collect some preliminary definitions:

Conditional Beliefs. A *conditional probability system* for player k is a collection of conditional probabilities[2]

$$b_k = (b_k(\cdot|h))_{h \in H} \in \prod_{h \in H} \Delta(\mathbf{M} \times S_{-k}(h))$$

with the property that for every history h such that $\mathbf{M} \times S_{-k}(h)$ has strictly positive probability under the initial belief $b_k(\cdot|\varnothing)$, the conditional probability $b_k(\cdot|h)$ is derived from $b_k(\cdot|\varnothing)$ by applying Bayes' rule (recall that $\varnothing$ denotes the empty history).

Beliefs and Information. Players' beliefs are required to conform to the information agents possess about the current matching outcome. Formally, given $\mathbf{m} \in \mathbf{M}$ and a player k, a conditional probability system b_k *is consistent with k's information* if $b_k(\mathcal{P}_k(\mathbf{m}) \times S_{-k}(h) | h) = 1$ for all $h \in H$. So, under this assumption, players are certain, at every history, of the information described by their partition.

Given outcome $\mathbf{m} \in \mathbf{M}$, a conditional probability system b_k satisfies the *grain of truth assumption* if $b_k(\{\mathbf{m}\} \times S_{-k}|\varnothing) > 0$. The assumption requires player k to assign strictly positive probability, at the beginning of the game, to the actual matching outcome. It will be sufficient to require *workers'* beliefs to satisfy the grain of truth assumption. Formally, given $\mathbf{m} \in \mathbf{M}$ and a conditional probability system b_k we say that b_k is *consistent* if it consistent with k's information and, in case $k \in I$, it satisfies the grain of truth assumption.

Stability. Given a player k, a conditional probability system b_k *believes in no competing offers* if the initial probability $b_k(\cdot|\varnothing)$ assigns probability 1 to each worker $i \neq k$ not making

[2] for every finite set S, we denote by $\Delta(S)$ the set of probability measures on S.

offers. The assumption expresses the idea that if a current matching is deemed to be stable then players will not expect others to initiate a negotiation in order to deviate from the match.

Optimality. Given a player k, a strategy s_k and a pair $(\mathbf{m}, s_{-k})$ in $\mathbf{M} \times S_{-k}$, let $U_k(s_k, s_{-k}, \mathbf{m})$ denote the resulting payoff for player k. A strategy s_k is *sequentially optimal under* b_k if at every history h where k is asked to act, the action specified by s_k maximizes the expectation of U_k with respect to $b_k(\cdot|h)$.

In addition to sequential rationality we assume that, given a conditional probability system b_i, a worker i makes offers only if she is not indifferent between making offers and abstaining. This tie-breaking assumption rules out cases where a worker makes offers that she believes will be rejected with probability 1.

To simplify the language, we call a strategy s_k *optimal under* b_k if it sequentially optimal under b_k and, in the case k is a worker, it satisfies the tie-breaking assumption described above.

4.1 Rationalizability and Stability

We now define our main solution concept. For the next definition, given a subset $\Psi \subseteq \mathbf{M} \times S$ we denote by Ψ_k and Ψ_{-k} the projection of Ψ on, respectively, $\mathbf{M} \times S_k$ and $\mathbf{M} \times S_{-k}$.

Definition 2 Let $\mathfrak{R}^0 = \mathbf{M} \times S$. Inductively, for every $n \geq 1$ define $\mathfrak{R}^n$ to be set of pairs $(\mathbf{m}, s) \in \mathbf{M} \times S$ such that for each player k there exists a consistent conditional probability system b_k such that the following hold:

(P1-n) s_k is optimal under b_k;

(P2-n) b_k believes in no competing offers;

(P3-n) $b_k\left(\mathfrak{R}^{n-1}_{-k}\middle|\varnothing\right) = 1$; and

(P4-n) for all $h \in H$ and $m \in \{0, ..., n-1\}$,

$$\text{if } \left(\mathcal{P}_k(\mathbf{m}) \times S_{-k}(h)\right) \cap \mathfrak{R}^m_{-k} \neq \emptyset \text{ then } b_k\left(\mathfrak{R}^m_{-k}|h\right) = 1. \quad (1)$$

A pair $(\mathbf{m}, s)$ is n*-rationalizable* if it belongs to $\mathfrak{R}^n$. The set of rationalizable outcome-strategy pairs is defined as $\mathfrak{R}^\infty = \bigcap_{n \geq 0} \mathfrak{R}^n$.

Definition 3 An outcome $\mathbf{m} \in \mathbf{M}$ is *stable under forward induction* if $(\mathbf{m}, a) \in \mathfrak{R}^\infty_i$ for every worker i. That is, if it is rationalizable for every worker to abstain from making offers.

Definition 2 is, essentially, an instance of Battigalli and Siniscalchi (2003)'s notion of strong Δ-rationalizability. We now describe the logic underlying the definition.

Consider a pair $(\mathbf{m}, s)$ consisting of a matching outcome and a profile of strategies. The pair is n-rationalizable if for each player k we can find conditional beliefs b_k so that b_k and s_k satisfy four basic conditions. Properties (P1-n) and (P2-n) establish that players are rational and expect others not to engage in negotiation. As n goes to infinity, (P3-n) implies that rationality and belief in no competing offers are almost common belief at the beginning of the game.

Property (P4-n) is crucial and disciplines beliefs conditional upon observing unexpected offers. Consider a history h reached after a worker made an unexpected offer to firm k. Notice that $\mathfrak{R}^1 \supseteq \ldots \supseteq \mathfrak{R}^{n-1}$ constitute increasingly stringent assumptions on players' beliefs and behavior. By (1), conditional on the offer, firm k assigns probability 1 to the strongest assumption $\mathfrak{R}^m$ that, by satisfying $(\mathcal{P}_k(\mathbf{m}) \times S_{-k}(h)) \cap \mathfrak{R}^m_{-k} \neq \emptyset$, has not been refuted by the observed offer *and* k's information $\mathcal{P}_k(\mathbf{m})$ about the market. Hence, (P4-n) captures the idea that players interpret offers according to the highest possible degree of sophistication that can be attached to their proponents and by mantaining, as much as possible, the assumption that such an offer was ex-ante unexpected.

Property (P4-n) expresses forward-induction reasoning. Following Battigalli and Siniscalchi (2003), say that a player "strongly believes" an event if she believes the event at the beginning of the game and at every history where the event is not contradicted by the evidence. Upon observing an offer, when $n = 2$, property (P4-n) requires players to strongly believe the event "other players are rational and did not expect the offer". When $n = 3$, each player strongly believes that "other players are rational, did not expect the offer and strongly believe that others are rational and did not expect the offer." And so on. The results in Battigalli and Prestipino (2013) can be adapted to show that at each n, Definition 2 captures the implications of, informally, (i) rationality, (ii) consistency, (iii) belief in non-competing offers and n orders of strong belief in (i)-(iii).

Finally, a matching outcome $\mathbf{m}$ is deemed to be stable under forward induction if, under $\mathbf{m}$, abstaining is a rationalizable strategy for every worker. It should be remarked that abstaining from making offers is not required to be the *only* rationalizable strategy. This makes stability under forward induction a relatively permissive solution concept. It will also make the results on ex-post efficiency more striking.

5 Incomplete-Information Stability

A notion of stability under incomplete information was recently introduced by Liu, Mailath, Postlewaite, and Samuelson (2014). Its definition takes the form of an iterative elimination procedure defined over the set of matching outcomes.

Definition 4 Let $\Lambda^0 = \mathbf{M}$. Inductively, for each $\ell \in \mathbb{N}$ define Λ^ℓ as the set of all outcomes $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{\ell-1}$ such that there is no $i \in I$, $j \in J$ and $q \in \mathbb{R}$ such that

$$\nu\left(\mathbf{w}_i, \mathbf{f}_j\right) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)} \tag{2}$$

and

$$\phi\left(\mathbf{w}'_i, \mathbf{f}_j\right) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j} \tag{3}$$

for all $\mathbf{w}' \in \mathbf{W}$ such that $\mathbf{m}' = (\mathbf{w}', \mathbf{f}, \mu, \mathbf{p})$ satisfies

$$\begin{aligned} \mathbf{m}' &\in \Lambda^{\ell-1}, & (4)\\ \mathcal{P}_j\left(\mathbf{m}'\right) &= \mathcal{P}_j\left(\mathbf{m}\right), \text{ and} & (5)\\ \nu\left(\mathbf{w}'_i, \mathbf{f}_j\right) + q &> \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}. & (6) \end{aligned}$$

Λ^ℓ is the set of matching outcomes that are *level ℓ incomplete-information stable*. The set of *incomplete-information stable* matching outcomes is $\Lambda^\infty = \bigcap_{\ell=1}^{\infty} \Lambda^\ell$.

In Liu, Mailath, Postlewaite and Samuelson (2014), the set Λ^0 is set equal to the set of all individually rational outcomes, rather than a finite set $\mathbf{M}$ as in Definition 4. The discretization $\Lambda^0 = \mathbf{M}$ simplifies the statements of our main results and avoids measurability considerations. A matching is stable in the definition of Liu, Mailath, Postlewaite and Samuelson (2014) if and only if it is stable (as defined above) for some market $\mathbf{M}$.[3]

An outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is eliminated in the first iteration if it is possible to find a worker i, a firm j, and a wage q so that the two agents can form a new partnership that is profitable for the worker and gives the firm a higher payoff than the original allocation $(\mu, \mathbf{p})$ for all types $\mathbf{w}'_i$ that satisfy restrictions (4)-(6). When $\ell = 1$, this amounts to considering type profiles $\mathbf{w}'$ that do not contradict the fact that j knows the type of the worker he is matched to, and such that the the partnership, if agreed upon, would be profitable for the worker. Successive iterations shrink the set of types that satisfy (4). In the ℓ-th step of the procedure, the same reasoning is applied to the set of matching outcome that have survived $\ell - 1$ steps of the elimination process.

As shown by Liu, Mailath, Postlewaite, and Samuelson (2014), incomplete-information stability satisfies two surprising properties. First, any complete-information stable matching is also incomplete-information stable.

Proposition 1 *Every complete-information stable matching outcome is incomplete-information stable.*

An outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is (ex-post) *efficient* if it achieves a maximal total surplus across all matching outcomes, keeping $\mathbf{w}$ and $\mathbf{f}$ fixed. Under standard supermodularity assumptions, any stable matching is efficient.

[3]This follows immediately from Lemma 1 and Proposition 2 in Liu, Mailath, Postlewaite and Samuelson (2014).

Proposition 2 *Let $W \subset \mathbb{R}$ and $F \subset \mathbb{R}$. Assume ν and ϕ are strictly increasing and strictly supermodular. Then, every incomplete information stable matching outcome is efficient.*[4]

6 Characterization Theorems

The next theorem characterizes the set of matching outcomes that are stable under forward induction.

Theorem 1 *A matching outcome $\mathbf{m} \in \mathbf{M}$ is stable under forward induction if and only if it is incomplete information stable.*

Theorem 1 provides epistemic foundations for incomplete-information stability, which can be interpreted as the outcome of noncooperative negotiation under the assumption that players revise their beliefs according to forward-induction reasoning. From a different perspective, the result shows, together with Proposition 2, that forward-induction reasoning leads to efficiency under supermodular premuneration values.

6.1 Exact Characterization

Theorem 1 shows that incomplete-information stability and stability under forward-induction lead to the same set of stable matchings. Based on the result, it might be tempting to conclude that forward-induction reasoning is the key and only epistemic principle underlying incomplete-information stability. In this section we show that this is not the case.

We study more in detail the relation between the two solution concepts by comparing the two procedures (Λ^{ℓ}) and $(\mathfrak{R}^{n})$ not only the final predictions $\mathfrak{R}^{\infty}$ and Λ^{∞} but also at each step of the two iterations. To this end, for every n we consider the set

$$\mathfrak{S}^{n} = \left\{ \mathbf{m} \in \mathbf{M} : (\mathbf{m}, a) \in \mathfrak{R}_{i}^{n} \text{ for every } i \in I \right\}.$$

So, $\mathfrak{S}^{n}$ is the collection of matching outcomes with the property that mantaining the status-quo is, for every worker, a n-rationalizable strategy.

In the idealized limit, as n and ℓ go to infinity, $\mathfrak{S}^{n}$ and Λ^{ℓ} converge to the same set of predictions. Does a similar equivalence hold once we restrict the attention to the more realistic case of a finite order n of rationalizability as well as a finite level ℓ of incomplete information stability? Addressing this question will allow us to fully characterize the relation between the two solution concepts.

While there is no reason to presume a simple equivalence such as $\mathfrak{S}^{n} = \Lambda^{n}$ to be true for every n and every market, one may expect the two solution concepts to be logically

[4] See Liu, Mailath, Postelwaite and Samuelson (2014) for a more general statement.

nested in the sense that given n and ℓ, $\mathfrak{S}^n \subseteq \Lambda^\ell$ or $\Lambda^\ell \subseteq \mathfrak{S}^n$ is guaranteed to hold. As we show below, this is generally not true. This creates a difficulty for a clear comparison between the two solution concepts.

Given a level ℓ of incomplete information stability, we define the following bounds:

$$B(\ell) = \min\left\{n : \mathfrak{S}^n \subseteq \Lambda^\ell\right\} \quad \text{and} \quad b(\ell) = \max\left\{n : \Lambda^\ell \subseteq \mathfrak{S}^n\right\}$$

The maps b and B relate orders of forward-induction reasoning to levels of incomplete-information stability, and satisfy $\mathfrak{S}^{B(\ell)} \subseteq \Lambda^\ell \subseteq \mathfrak{S}^{b(\ell)}$ for every ℓ. They admit the following interpretation:

- Suppose $\mathfrak{S}^n \subseteq \Lambda^\ell$. Then, an outside observer who knows agents play n-rationalizable strategies will be able to infer that any matching that is not blocked is incomplete-information stable at level ℓ. The map B characterizes the *minimum* order n for which this is true.

- Suppose $\Lambda^\ell \subseteq \mathfrak{S}^n$. In this case, observing a matching outcome belonging to Λ^ℓ does not reject the hypothesis that players play n-rationalizable strategies. The map b describes, for every ℓ, the *highest* order of rationalizability that is not rejected by observing matchings that are incomplete-information stable at level ℓ.[5]

If there is a non-trivial gap between the two bounds, then any n that satisfies $B(\ell) > n > b(\ell)$ is such that Λ^ℓ and $\mathfrak{S}^n$ cannot be directly compared. That is, neither $\Lambda^\ell \subseteq \mathfrak{S}^n$ nor $\mathfrak{S}^n \subseteq \Lambda^\ell$ hold.

The maps b and B depend, in general, on the particular market under consideration. The next result establishes universal bounds that apply to any market.

Theorem 2 *For every $\ell \in \mathbb{N}$, b and B satisfy $3\ell \geq B(\ell)$ and $b(\ell) \geq 1 + 2\ell$.*

The result establishes the inclusions $\mathfrak{S}^{3\ell} \subseteq \Lambda^\ell \subseteq \mathfrak{S}^{1+2\ell}$ for every level ℓ. We now show that these bounds cannot be improved upon. For the next result, we summarize by a tuple (I, J, ν, ϕ, W, F) and a market $\mathbf{M}$ an instance of all the primitives of the model introduced in section 2.

Theorem 3 *For every N there exist a tuple (I, J, ν, ϕ, W, F) and a market $\mathbf{M}$ such that $B(\ell) = 3\ell$ and $b(\ell) = 1 + 2\ell$ for every $\ell \leq N$.*

Theorems 2 and 3 fully characterize the relation between incomplete-information stability and stability under forward induction. They establish tight bounds relating the

[5] That is, observing an outcome $\mathbf{m} \in \Lambda^\ell$ never rejects the hypothesis that players's decision to abstain is $b(\ell)$-rationalizable, but it may reject the hypothesis that players play n-rationalizable strategies for $n > b(\ell)$.

two solution concepts. In addition, the proof of Theorem 3 provides a simple example of a market where the gap $B(\ell) - b(\ell)$ can be made arbitrarily large.

Taken together, Theorems 1 and 3 show that while there is a strong connection in terms of predictions between incomplete-information stability and stability under forward-induction, the two solution concepts are quite distinct. In particular, to the extent that incomplete-information stability is meant to reflect an inference process performed by the agents in the market, Theorem 3 suggests that such an inference does not necessarily reflect forward-induction reasoning.

6.2 Illustrative Example

We now provide an example of a market where both stability under forward induction and incomplete-information stability lead to the conclusion that a certain matching outcome is not stable, but through different types of inference by the agents on the uninformed side of the market. The example will also allow us to illustrate some of the ideas underlying the main results in the paper.

	A	B		A	B	
worker payoffs:	4	3		4	3	
worker types, $\mathbf{w}$:	2	3		2	$\hat{3}$	
payment, $\mathbf{p}$:	-4	0		-4	0	
firm types, $\mathbf{f}$:	4	1	ε	4	1	ε
firm payoffs:	12	3	0	12	0	0
	L	C	E	L	C	E
matching outcome:		$\mathbf{m}$			$\hat{\mathbf{m}}$	

Figure 1

The market consists of the two matching outcomes $\mathbf{m}$ and $\hat{\mathbf{m}}$, as described in Figure 1. There are two workers, A and B, and three firms, L, C, and E. In both outcomes worker B is matched to firm C at wage 0 while worker A is matched to firm L at wage -4. Firm E is unmatched. The only uncertainty is about B's type, which can be either 3 in $\mathbf{m}$, or $\hat{3}$ in $\hat{\mathbf{m}}$. Firm E's type equals a known constant $\varepsilon \in [0, 1/2]$. We refer to $\hat{3}$ and ε as "bad" types and to the remaining types as "regular."

A match between two regular types w and f lead to standard premuneration values $\nu(w, f) = \phi(w, f) = wf$. A match between the two bad types produces payoffs $\nu(\hat{3}, \varepsilon) = \phi(\hat{3}, \varepsilon) = 3\varepsilon$. The remaining values are defined as follows: If $f \neq \varepsilon$ then $\nu(\hat{3}, f) = 3f$ and $\phi(\hat{3}, f) = 0$. If $w \neq \hat{3}$ then $\nu(w, \varepsilon) = 0$ and $\phi(w, \varepsilon) = w\varepsilon$. So, a match with type $\hat{3}$ is of no value to firms L and C. However, the premuneration value $\nu(\hat{3}, \cdot)$ worker B obtains from a match is the same as that of the regular type $w = 3$.

If worker B's actual type is 3, then the matching $\mathbf{m}$ is not incomplete-information stable. The first step of Definition 4 eliminates the outcome $\hat{\mathbf{m}}$. This is because worker B and firm E can form a blocking pair at transfer $q = \varepsilon/2$. Such a blocking pair increases E's payoff from 3 to $3 + \varepsilon/2$. The outcome $\mathbf{m}$ is then eliminated in the second step by considering a blocking pair between worker B and firm L at transfer $q = -1/2$.

Key to the argument is the inference made by firm L about B's type upon being involved in a blocking pair. Incomplete-information stability (or, more precisely, a strict interpretation of it) suggests the following line of reasoning for firm L: "Suppose worker B were of type $\hat{3}$. Then, she could have formed a blocking pair with firm E. This is because firm E, unlike L, would have accepted to match with B under any possible belief. However, the existence of such a profitable deviation would contradict the hypothesis that the matching is stable. Hence, her type must be 3. Therefore I agree to break the current matching and match with B."

If blocking pairs are formed by a noncooperative negotation, as in this paper, then we encounter a key difficulty in formalizing the inference described in the previous paragraph. It is perhaps not obvious whether worker B, if of type $\hat{3}$, would indeed choose to form a blocking pair with firm E. To illustrate, suppose ε is small. In this case, a worker of type $\hat{3}$ faces a nontrivial choice between making an offer to firm E and increasing her payoff by at most ε, or making an offer to firm L in the hope of being mistaken for a regular type and obtain a (potentially) much higher payoff. The choice between the two offers must ultimately depend on worker B's belief on how firm L will interpret an offer. However, if we admit the possibility for worker $\hat{3}$ to try to form a blocking pair with firm L, rather than E, than why should the same possibility be ruled out by firm L?

We now analyze this example using stability under forward-induction. We will reach three main conclusions. First, both $\mathbf{m}$ and $\hat{\mathbf{m}}$ are not stable under forward induction. Second, the inference made by firm L upon receiving an offer from worker B will be qualitatively different from the inference we described above and will explicitly hinge on forward induction reasoning. Finally, stability under forward-induction allows us to remain agnostic about whether worker B, were she of type $\hat{3}$, would be more likely to make an offer to firm L or firm E.

The conclusion that the matching outcome in Figure 1 is not stable under forward-induction is reached through a series of simple claims. In what follows, in order to ease the exposition, we consider a restricted game where worker B can either abstain or make one of two possible offers: an offer $s' = (\mathrm{L}, -1/2)$ to firm L at transfer $q = -1/2$ and an offer $s'' = (\mathrm{E}, \varepsilon/2)$ to firm E at transfer $q = \varepsilon/2$. In addition, we assume worker A can only abstain. Such assumptions are without loss of generality.

Claim 1 *When the matching outcome is* $\hat{\mathbf{m}}$*, abstaining is not* 2*-rationalizable for worker B and both offers* s' *and* s'' *are* 2*-rationalizable. I.e.* $(\hat{\mathbf{m}}, a) \notin \mathfrak{R}^2_B$ *and* $(\hat{\mathbf{m}}, s'), (\hat{\mathbf{m}}, s'') \in \mathfrak{R}^2_B$*. Both accepting and rejecting offer* s' *are* 2*-rationalizable strategies for firm E.*

Proof. In the first step we can eliminate the strategy where firm E rejects offer s''. This implies that in step $n = 2$ we can eliminate the pair $(\hat{\mathbf{m}}, a)$ where worker B is of type $\hat{3}$ and abstains from making an offer, since abstaining is now strictly worse than making offer s''. In addition, both offers s' and s'' are 2-rationalizable. Offer s'' is a best response to a belief concentrated on the event where firm L rejects offer s', if made. Offer s' is optimal with respect to any belief under which L accepts s' with probability sufficiently high. ■

Claim 2 *There is no strategy s for worker B such that $(\hat{\mathbf{m}}, s^*)$ is 3-rationalizable. If the outcome is $\mathbf{m}$, both abstaining and making offer s' are 3-rationalizable strategies for type 3. I.e. $(\{\hat{\mathbf{m}}\} \times S) \cap \mathfrak{R}^3 = \emptyset$ and $(\mathbf{m}, a), (\mathbf{m}, s) \in \mathfrak{R}^3_B$.*

Proof. Because $(\hat{\mathbf{m}}, a) \notin \mathfrak{R}^2_B$ and firm C knows the type of worker B, there can be no belief of firm C that, at the beginning of the game, assigns probability 1 to $\mathfrak{R}^2$ and to the event where worker B is of type $\hat{3}$ and abstains from making offers. Hence, $(\{\hat{\mathbf{m}}\} \times S) \cap \mathfrak{R}^3 = \emptyset$. The same argument applied in the proof of Claim 1 implies that $(\mathbf{m}, a), (\mathbf{m}, s') \in \mathfrak{R}^3_B$. ■

Claim 3 *Any strategy s for firm L such that $(\mathbf{m}, s)$ is 4-rationalizable must accept offer s'. Thus $(\mathbf{m}, a) \notin \mathfrak{R}^5_B$. Hence $\mathbf{m}$ is not stable under forward-induction.*

Proof. Claim 2 shows $(\{\hat{\mathbf{m}}\} \times S) \cap \mathfrak{R}^3 = \emptyset$ and $(\mathbf{m}, s') \in \mathfrak{R}^3_B$. Hence, conditional on observing offer s', property (P4-n) in Definition 4, evaluated at $n = 4$, implies that firm L must assign probability 1 to the event that worker B is of type 3. Hence, the firm must accept the offer. It follows that s' is the only 5-rationalizable strategy for type 3. So, the matching is not stable under forward-induction. ■

The conclusion that the match is not stable under forward induction is based on the following intuition. Consider the offer $s' = (\text{L}, -1/2)$ made by worker 3 to firm L. Firm L must interpret the offer by mantaining the highest possible degree of belief in the event that other players are rational and that the offer was unexpected (by everyone other than worker B). In particular, and this is the key aspect, firm L must take into account that the offer was unexpected to firm C even though the same firm *knew* B's actual type.

So firm L must ask: What is the "best" possible explanation that, ex-ante, could have justified firm C's belief that worker B was going to abstain? Such an explanation depends on B's type. Consider the case where B's type is $\hat{3}$. Then firm C must have thought that B believed that firm E was irrational. If not, then B would have expected E to accept offer s', making abstaining a non-optimal strategy. Now consider the case where B's type is 3. In this case, firm C could have expected B to abstain as a best response to the belief that firm I would have rejected offer s under the incorrect belief that B's type was $\hat{3}$. This explanation assigns a higher degree of belief of rationality to B's belief.

Under forward-induction reasoning, firm L favors the latter explanation, and so must rule out the possibility that B's type is 3. Anticipating this, B cannot rationally abstain from making offer s'. Thus, the matching is not stable under forward induction.

7 Extensions and Discussion

7.1 Grain of Truth Assumption

The grain of truth assumption plays two roles in the paper. First, without such assumption, and under general conditions on premuneration values and the set of payoff-types, a worker could abstain from making offers under the certain belief that his or her payoff-type is the worst payoff-type in the market (and so that no offer that is profitable for her would be accepted by any firm). So, as one may except, the grain of truth assumption shrinks the set of matchings that can be sustained as stable. The grain of truth assumption plays another, more subtle, role in the main results because it facilitates forward-induction reasoning. A firm, upon receiving an offer, must reason about the fact that this offer was unexpected even though other agents attached positive probability to the *actual* type of the worker who put forward the offer. As seen in the example in the previous section, this type of reasoning is germane to forward-induction reasoning in these markets.

7.2 Offers and Rejection

Stability under forward-induction requires abstaining from making offers to be a rationalizable strategy for every worker. Alternatively, it may be to natural to deem "stable" a matching where any profitable offer, if made would be rejected. The next theorem shows an equivalence between these two notions. For the next result, we say that a strategy s_j rejects the unilateral offer $s_i = (j, q)$ if s_j rejects offer s_i when the latter is the only offer made by any worker.

Theorem 4 *A matching outcome* $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ *is stable under forward induction if and only if there is a rationalizable pair* $(\mathbf{m}, s) \in \mathfrak{R}^\infty$ *such that for every firm* j*, the strategy* s_j *rejects any unilateral offer* $s_i = (j, q)$ *such that* $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$.

The result shows that a matching is stable if and only if we can find for every firm j a rationalizable strategy s_j that rejects any offer that, if accepted, would be profitable for the worker proposing it.

7.3 Strict Stability

The fact that a matching outcome is stable does not imply that abstaining is, for every worker, the *only* rationalizable strategy. We call an outcome $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ *strictly stable* if $\mathfrak{R}_i^\infty = \{(\mathbf{m}, a_i)\}$ for every i. We now show that strict stability is an unsuitably strong notion of stability. The next result provides a characterization.

Theorem 5 *Consider a market* $\mathbf{M}$*. A matching outcome* $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ *in* $\mathbf{M}$ *is strictly stable if and only if* $\mathbf{m} \in \Lambda^{\infty}$ *and there is no worker* i*, firm* j *and payment* q *such that*

$$\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$$

and

$$\phi(\mathbf{w}'_i, \mathbf{f}_j) - q \geq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$$

for some $\mathbf{w}' \in \mathbf{W}$ *such that* $\mathbf{m}' = (\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{\infty}$*,* $\mathcal{P}_j(\mathbf{m}') = \mathcal{P}_j(\mathbf{m}')$ *and* $\nu(\mathbf{w}'_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$.

A strict stable matching outcome is incomplete-information stable. In addition, under strict stability, a worker i and a firm j can block a matching outcome $\mathbf{m}$ as long as there is *some* payoff profile $\mathbf{w}'$ that makes the combination (i, j, q) profitable for firm j and such that the resulting outcome $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p})$ is incomplete-information stable. An immediate implication of the result is that a strict stable matching outcome must be complete-information stable. In addition, it is possible to constructs markets $\mathbf{M}$ that contain multiple complete-information stable outcomes but no strict stable outcomes.

7.4 Conclusions

This paper proposes a new notion of stability for markets with one-sided uncertainty. Stability is formulated in an epistemic framework. Its definition is based on two main ideas. First, as in many real life situations, an existing allocation can only be altered if agents actively engage in negotiation. Second, forward-induction reasoning provides a non-equilibrium theory of belief revision that is particularly suitable for describing how beliefs are updated throughout the negotiation phase. To test the usefulness of this approach, the main theorem of this paper establishes an equivalence result between stability under forward-induction and incomplete information stability. The latter is a solution concept recently introduced in the literature and which satisfies surprising properties in terms of existence and efficiency. The result shows that stability under forward-induction can be applied through a simple algorithm and provides epistemic foundations for incomplete-information stability.

A Appendix

A.1 Definition of Q

We now make formal the assumption, introduced in Section 3.1, that the set Q is a sufficiently fine grid. To this end, since W and F are finite, we can find a large enough open interval $(\alpha, \beta) \subseteq \mathbb{R}$ such that it is without loss of generality to restrict the attention, in the definition of incomplete-information stability, to payments q that belong to (α, β). Given $\varepsilon > 0$, we call a finite set $A \subseteq [\alpha, \beta]$ an ε-*grid* if every open subinterval of (α, β) of diameter ε intersects A.

We assume that Q is an ε-grid where $\varepsilon \leq \varepsilon^*$ and the bound $\varepsilon^* > 0$ is described below.

For every $\mathbf{m} \in \mathbf{M} \backslash \Lambda^{\infty}$ let $n_{\mathbf{m}} \geq 0$ be such that $\mathbf{m} \in \Lambda^{n_{\mathbf{m}}} \backslash \Lambda^{n_{\mathbf{m}}+1}$. For every $\mathbf{m} \in \mathbf{M} \backslash \Lambda^{\infty}$ consider the set $P_{\mathbf{m}}$ of pairs (i, j) such that for some payment $q \in (\alpha, \beta)$ the combination (i, j, q) has the property that it $n_{\mathbf{m}}$-blocks the outcome $\mathbf{m}$. For every $(i, j) \in P_{\mathbf{m}}$ select one payment $q(i, j, \mathbf{m})$ such that $(i, j, q(i, j, \mathbf{m}))$ blocks $\mathbf{m}$.

Because the definition of incomplete-information stability involves only strict inequalities for each $q(i, j, \mathbf{m})$ there exists a small enough $\varepsilon(i, j, \mathbf{m}) > 0$ such that any $q' \in \mathbb{R}$ that is at distance at most $\varepsilon(i, j, \mathbf{m})$ from q has the properties that q' belongs to (α, β) and (i, j, q') also $n_{\mathbf{m}}$-blocks the outcome $\mathbf{m}$. We define ε^* be the minimal $\varepsilon(i, j, \mathbf{m})$ across all payments $q(i, j, \mathbf{m})$.

By construction, the bound ε^* has the following property. For every $\varepsilon \leq \varepsilon^*$ and every ε-grid A, if there exists a combination (i, j, q) that $n_{\mathbf{m}}$-blocks an outcome $\mathbf{m} \in \mathbf{M}$ then there exists $q' \in A$ such that the combination (i, j, q') also $n_{\mathbf{m}}$-blocks the same outcome.

A.2 Preliminaries

Given any subset $\Psi \subseteq \mathfrak{R}$ and player k denote by Ψ_k and Ψ_{-k} the projections of Ψ on $\mathbf{M} \times S_k$ and $\mathbf{M} \times S_{-k}$, respectively. For every k and conditional probability system (henceforth, CPS) b_k we will denote by $b_{k,h}$ the probability measure $b_k(\cdot|h)$.

As shown in the next lemma, for a given worker i it is enough to consider an initial probability $\rho_i \in \Delta(\mathbf{M} \times S_{-i})$ rather than a fully specified conditional probability system.

Lemma 1 *Fix $n \geq 0$, $\mathbf{m} \in \mathbf{M}, i \in I$ and a strategy $s_i \in S_i$. Let $\rho_i \in \Delta(\mathbf{M} \times S_{-i})$ be a probability measure such that $\rho_i(\{\mathbf{m}\} \times S_{-i}) > 0, \rho_i(\mathcal{P}_i(\mathbf{m}) \times S_{-i}) = 1$ and s_i and ρ_i satisfy properties* (P1-n)-(P3-n). *Then there exists a CPS b_i such that $b_{i,\varnothing} = \rho_i$ and s_i and b_i satisfy properties* (P1-n)-(P4-n).

Proof. The CPS b_i is easily defined as follows. Let $b_{i,\varnothing} = \rho_i$. Denote by H^A_{-i} be the set of histories following no offers from workers other than i. For every $h \in H^A_{-i}$ let $b_{i,h} = b_{i,\varnothing}$. Now consider all histories $h \notin H^A_{-i}$ such that $h \neq \varnothing$ and (1) holds for $m = n - 1$. For every such history define $b_{i,h}$ to satisfy $b_{i,h}\left(\left(\{\mathbf{m}\} \times S_{-j}(h)\right) \cap \mathfrak{R}^m_{-i}\right) = 1$. Proceeding inductively, we can decrease m and repeat the argument at every step. Because $\mathfrak{R}^0_{-i} = \mathfrak{R}_{-i}$, then for every history there exists $m \leq n-1$ such that (1) holds. So, we obtain a collection of conditional probabilities $b_i = (b_{i,h})_{h \in H}$. We need to verify that b_i is a well defined CPS. Because $b_{i,\varnothing}$ assigns probability 1 to no offer being made by other workers, only histories in H^A_{-i} have initial strictly positive probability. For every such history h we have $b_{i,h} = b_{i,\varnothing}$, so Bayesian updating is respected. Hence, b_i is a well defined CPS. By construction, the pair (s_i, b_i) satisfies (P1-n)-(P4-n). ■

As recorded below, for a fixed matching outcome $\mathbf{m}$ the set $\{s \in S : (\mathbf{m}, s) \in \mathfrak{R}^n\}$ has a product structure. The result follows immediately from Definition 2 and its proof is omitted.

Lemma 2 *Fix $s \in S$ and $\mathbf{m} \in \mathbf{M}$. If $(\mathbf{m}, s_k) \in \mathfrak{R}^n_k$ for each k then $(\mathbf{m}, s) \in \mathfrak{R}^n$.*

We conclude this subsection with a lemma on the composition of multiple strategies. Recall that H_j denotes the set of histories at which firm j has received at least one offer.

Lemma 3 *Fix $n \geq 0$, $\mathbf{m} \in \mathbf{M}$, and $j \in J$. Consider a finite sequence*

$$(\mathbf{m}^1, s^1_j), ..., (\mathbf{m}^m, s^m_j) \text{ in } \mathfrak{R}^n_j$$

such that $\mathcal{P}_j(\mathbf{m}^1) = \ldots = \mathcal{P}_j(\mathbf{m}^m)$. If a strategy s_j is such that

$$s_j(h) \in \left\{s^1_j(h), ..., s^m_j(h)\right\} \quad \text{for all } h \in H_j$$

then $(\mathbf{m}, s_j)$ belongs to $\mathfrak{R}^n_j$.

Proof. For every $r = 1, ..., m$, let b^r_j be a consistent CPS such that s^r_j and b^r_j satisfy properties (P1-n)-(P4-n). For every $h \in H_j$, let $r(h) \in \{1, ..., m\}$ be such that $s_j(h) = s^{r(h)}_j(h)$. Define the CPS b_j as $b_{j,h} = b^{r(h)}_{j,h}$ for every $h \in H_j$ and $b_{j,h} = b^1_{j,h}$ for every $h \in H \backslash H_j$. The CPS b_j is well defined. To see this, notice that the only history different from $\varnothing$ that is reached with positive probability under $b_{j,\varnothing} = b^1_{j,\varnothing}$ is the history h^* in which all workers abstain from making offers. Because $h^* \notin H_j$ then $b_{j,h^*} = b^1_{j,h^*}$. Hence, the requirement of Bayesian updating is respected. Since each b^r_j is consistent it follows that b_j is consistent as well. In addition, because $b_{j,\varnothing} = b^1_{j,\varnothing}$ then b_j satisfies (P2-n) and (P3-n). We now verify that (P4-n) holds. For every $m \in \{0, ..., n-1\}$ and every history h, if $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}^m_{-j} \neq \emptyset$ then $b^{r(h)}_{j,h}$ assigns probability 1 to $\mathfrak{R}^m_{-j}$, hence $b_{j,h}$ assigns probability 1 to $\mathfrak{R}^m_{-j}$ as well. Thus property (P4-n) is satisfied. Finally, the action $s_j(h)$ is optimal with respect to $b^{r(h)}_{j,h} = b_{j,h}$ at every history $h \in H_j$. Hence s_j is optimal with respect to b_j. Therefore, we can conclude that $(\mathbf{m}, s_j) \in \mathfrak{R}^n_j$. ■

A.3 Proof of Theorems 1 and 2

Let $S_I = \prod_{i \in I} S_i$ and $S_J = \prod_{j \in J} S_j$. For every n, denote by $\mathfrak{R}_I^n$ the projection of $\mathfrak{R}^n$ on $\mathbf{M} \times S_I$ and by $\mathfrak{R}_J^n$ the projection on $\mathfrak{R}^n$ on $\mathbf{M} \times S_J$. Also let $a_I = (a_i)_{i \in I}$ and for each i denote by a_{-i} the vector $(a_k)_{k \in I \setminus \{i\}}$.

Lemma 4 *For every $\mathbf{m} \in \mathbf{M}$, $i \in I$, $n \geq 1$ and $s_i \in S_i$,*

1. *If $(\mathbf{m}, s_i) \in \mathfrak{R}_i^n$ then $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$;*
2. *If $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$ then $(\{\mathbf{m}\} \times S) \cap \mathfrak{R}^n \neq \emptyset$.*

Proof. (1) Suppose $(\mathbf{m}, s_i) \in \mathfrak{R}_i^n$. Consider a worker $k \neq i$ (recall that $|I| \geq 2$ by assumption). Then, since $(\mathbf{m}, s_k) \in \mathfrak{R}_k^n$, there must exist a corresponding CPS b_k such that $b_{k,\varnothing}\left(\mathfrak{R}_{-k}^{n-1}\right) = 1$, $b_{k,\varnothing}(\{\mathbf{m}\} \times S_{-k}) > 0$ and $b_{k,\varnothing}(A_i) = 1$, where $A_i = \{(\mathbf{m}, s_{-k}) : s_i = a_i\}$. Therefore

$$b_{k,\varnothing}\left(\mathfrak{R}_{-k}^{n-1} \cap (\{\mathbf{m}\} \times S_{-k}) \cap A_i\right) > 0$$

So, in particular, $(\mathbf{m}, a_i) \in \mathfrak{R}_i^{n-1}$. Because s_i and i are arbitrary, it follows that $(\mathbf{m}, a_i) \in \mathfrak{R}_i^{n-1}$ for every $i \in I$. Hence, Lemma 2 implies $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$.

(2) Let $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$. Then $(\{\mathbf{m}\} \times \{a_I\} \times S_J) \cap \mathfrak{R}^{n-1} \neq \emptyset$. Thus, for each player k we can find a probability $\rho_k \in \Delta(\mathfrak{R}_{-k})$ assigning probability 1 to $(\{\mathbf{m}\} \times \{a_{-k}\} \times S_J) \cap \mathfrak{R}_{-k}^{n-1}$.[6]

The probability ρ_k can then be extended to a consistent CPS b_k such that $b_{k,\varnothing} = \rho_k$. To this end, define a vector $(b_{k,h})_{h \in H}$ as follows. Let $b_{k,\varnothing} = \rho_k$. As in the proof of Lemma 1, let H_{-k}^A be the set of histories following no offers from workers $I \setminus \{k\}$. For every $h \in H_{-k}^A$ let $b_{k,h} = b_{k,\varnothing}$. Now consider all histories $h \notin H_{-k}^A$ such that $h \neq \varnothing$ and (1) holds for $m = n - 1$. For every such history define $b_{k,h}$ to assign probability 1 to $(\mathcal{P}_k(\mathbf{m}) \times S_{-k}(h)) \cap \mathfrak{R}_{-i}^m$. Proceeding inductively, decrease m and repeat the argument to obtain a vector $b_k = (b_{k,h})_{h \in H}$. We need to verify that b_k is a well defined conditional probability system. Because $b_{k,\varnothing}$ assigns probability 1 all workers abstaining from making offers (except possibly for k), only histories in H_{-k}^A are reached with strictly positive probability under $b_{k,\varnothing}$. For every such history h we have $b_{k,h} = b_{k,\varnothing}$. Hence, b_k is a well defined conditional probability system. By construction it is consistent. In addition, by definition b_k believes in no competing offers, and it is immediate to verify it satisfies properties (P3-n) and (P4-n). Any strategy s_k that is optimal with respect to b_k is such that the pair (s_k, b_k) satisfies properties (P1-n)-(P4-n). A profile s of such strategies satisfies $(\mathbf{m}, s) \in \mathfrak{R}^n$. Hence $(\{\mathbf{m}\} \times S) \cap \mathfrak{R}^n \neq \emptyset$. ■

[6]The assumption that ρ_k assigns probability 1 to the actual outcome ω simplifies the notation but is inessential to the argument.

The next two lemmas provide conditions that are sufficient and necessary for a matching outcome $\mathbf{m}$ to satisfy $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$. To ease notation we denote by $[i, j, q]$ the second-stage history reached when all workers except for i abstain from making offers and i makes offer (j, q).

Lemma 5 *For every $n \geq 1$, $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$ if and only $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$ and there exists a strategy profile $(s_j^*)_{j \in J}$ such that:*

1. *$(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-1}$ for every j; and*
2. *$s_j^*(h) = r$ for every j and every history $h = [i, j, q]$ that satisfies*

$$\nu(\mathbf{w}_i, f_j) + q > \nu\left(\mathbf{w}_i, f_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}.$$

Proof. Let $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$. Consider an offer (j, q) by worker i such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ and fix $h = [i, j, q]$. We claim there must exist a strategy of firm j, which we denote by $s_j^{i,q}$, with the properties that $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{n-1}$ and $s_j^{i,q}(h) = r$.

Suppose, as a way of contradiction, that such a strategy does not exists. Then offer (j, q) is accepted (i.e. $s_j(h) = i$) by any strategy $s_j \in S_j$ such that $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{n-1}$. Let b_i be a CPS that satisfies the grain of truth assumption and such that $b_{i,\varnothing}\left(\mathfrak{R}_{-i}^{n-1}\right) = 1$. Then, $b_{i,\varnothing}$ must attach strictly positive probability to the event where $s_j(h) = i$. Therefore, if b_i is a consistent CPS that satisfies properties (P2-n) and (P3-n) then a_i cannot be optimal with respect to b_i. This contradicts the assumption that $(\mathbf{m}, a_i) \in \mathfrak{R}_i^n$ and concludes the proof of the claim.

Given a firm j, define the set

$$D_j = \left\{s_j^{i,q} : i \in I, q \in Q \text{ and } \nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}\right\}$$

We now compose the strategies in D_j into a new strategy s_j^* as follows: For every history $h = [i, j, q]$ such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$, let $s_j^*(h) = s_j^{i,q}(h)$. For any other history $h \in H_j$, let $s_j^*(h) = s_j(h)$ for some strategy s_j such that $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{n-1}$. Because the set D_j is finite Lemma 3 implies $(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-1}$. This concludes the first part of the proof.

We now prove the converse implication. Because $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$ we know from Lemma 4 that $(\{\mathbf{m}\} \times S) \cap \mathfrak{R}^n \neq \emptyset$. We now show that $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$. Let $s_J^* = (s_j^*)_{j \in J}$ be a profile of strategies that satisfies conditions (1) and (2) in the statement. For every worker i, let $\rho_i \in \Delta(\mathfrak{R}_{-i})$ assign probability 1 to $(\mathbf{m}, a_{-i}, s_J^*)$. Because $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$ and $(\mathbf{m}, s_J^*) \in \mathfrak{R}_J^{n-1}$ then $(\mathbf{m}, a_I, s_J^*) \in \mathfrak{R}^{n-1}$ by Lemma 3. So, ρ_i assigns probability 1 to $\mathfrak{R}_{-i}^{n-1}$. Strategy a_i is optimal with respect to ρ_i. Using Lemma 1, we can define a consistent CPS b_i such that $b_{i,\varnothing} = \rho_i$ and a_i and b_i satisfy properties (P1-n)-(P4-n). Hence $(\mathbf{m}, a_i) \in \mathfrak{R}_i^n$. By repeating the construction for every $i \in I$ we obtain $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$. ∎

Lemma 6 *Let* $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, p)$. *For every* $n \geq 2$, $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$ *if and only if* $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$ *and there is no worker* i *and strategy* $s_i = (j, q)$ *such that* $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ *and*

$$\phi(\mathbf{w}_i', \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$$

for all and at least one profile $\mathbf{w}' \in \mathbf{W}$ *such that*

$$\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } ((\mathbf{w}', \mathbf{f}, \mu, p), s_i, a_{-i}) \in \mathfrak{R}_I^{n-2}. \tag{7}$$

Proof of Lemma 6. We first prove the "only if" part. Suppose $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$ and $s_i = (j, q)$ and $\mathbf{w}'$ are such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ and

$$\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } ((\mathbf{w}', \mathbf{f}, \mu, p), s_i, a_{-i}) \in \mathfrak{R}_I^{n-2}.$$

We now show that $\phi(\mathbf{w}_i'', \mathbf{f}_j) - q \leq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$ for some profile $\mathbf{w}''$ that satisfies (7).

Since $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$, we can apply Lemma 5. Let $(s_j^*)_{j \in J} \in S_J$ be the corresponding profile of strategies. In particular, $(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-1}$ for every j. For each j, let b_j^* be a consistent CPS such that s_j^* and b_j^* satisfy properties (P1-$(n-1)$)-(P4-$(n-1)$).

Let $h = [i, j, q]$. Since $(\mathbf{w}', \mathbf{f}, \mu, p) \in \mathcal{P}_j(\mathbf{m})$ we have $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2} \neq \emptyset$. So, $b_{j,h}^*$ must assign probability 1 to $\mathfrak{R}_{-j}^{n-2}$. Because $s_j^*(h) = r$ then, in order for r to be optimal with respect to $b_{j,h}^*$, the latter must attach strictly positive probability to some profile $\mathbf{w}'' \in \mathbf{W}$ such that

$$\mathbf{w}''_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } \phi(\mathbf{w}_i'', \mathbf{f}_j) - q \leq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}.$$

This concludes the first part of the proof.

We now prove the "if" part. Let $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$ and assume that the other conditions in the "if" part of the statement are satisfied. We now show that $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$. For every firm j, let H_j^* be the set of histories of the form $h = [i, j, q]$ for some offer $s_i = (j, q)$ such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ and $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2} \neq \emptyset$.

For every $h \in H_j^*$ we can define, by assumption, a probability $\rho_{j,h} \in \Delta\left(\mathfrak{R}_{-j}^{n-2}\right)$ whose marginal on $\mathbf{M}$ assigns probability 1 to an outcome $\left(\mathbf{w}^h, \mathbf{f}, \mu, \mathbf{p}\right) \in \mathcal{P}_j(\mathbf{m})$ where

$$\phi\left(\mathbf{w}_i^h, \mathbf{f}_j\right) - q \leq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}.$$

We now extend the vector $(\rho_h)_{h \in H_j^*}$ to a CPS. First, the vector is extented to the collection of all histories h such that $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2} \neq \emptyset$. To this end, define the probability $\rho_{j,\varnothing}$ to satisfy $\text{marg}_{\mathbf{M} \times S_I} \rho_{j,\varnothing}(\mathbf{m}, a_I) = 1$ and $\rho_{j,\varnothing}\left(\mathfrak{R}_{-j}^{n-2}\right) = 1$. This is possible since $(\mathbf{m}, a_I)$ belongs to $\mathfrak{R}_I^{n-1} \subseteq \mathfrak{R}_I^{n-2}$ by assumption. If h is the history following no offers to any firm, let $\rho_{j,h} = \rho_{j,\varnothing}$. For any other history h such that $h \notin H_j^*$ but $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2} \neq \emptyset$, let $\rho_{j,h}$ assign probability 1 to $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2}$.

The resulting vector of conditional probabilities can now be extended to a CPS. Recall that $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$. So, we can apply Lemma 5 and obtain a profile $s_J^* = (s_j^*)_{j \in J}$ of strategies that satisfy $(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-2}$ for every j as well as condition (2) of that Lemma. For each j, let b_j^* be a consistent CPS such that s_j^* and b_j^* satisfy properties (P1-$(n-2)$)-(P4-$(n-2)$). Define a CPS b_j such that

$$\begin{aligned} b_{j,h} &= \rho_{j,h} \text{ if } h \text{ is such that } (\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2} \neq \emptyset \text{ and} \\ b_{j,h} &= b_{j,h}^* \text{ otherwise} \end{aligned}$$

(Battigalli (1997) applies a similar argument).[7] It is immediate to verify b_j is consistent.

Now let s_j be a strategy such that:

(i). $s_j(h) = r$ for every $h \in H_j^*$;

(ii). $s_j(h)$ is a best response to$b_{j,h}$ for every $h \in H_j \backslash H_j^*$ such that $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2} \neq \emptyset$; and

(iii). $s_j(h) = s_j^*(h)$ for every other history $h \in H_j$.

We now verify that $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{n-1}$. By definition, $s_j(h)$ is a best response to to $b_{j,h}$ at every $h \in H_j$. So s_j is optimal with respect to b_j. By the definition of $\rho_{j,\varnothing}$, b_j also satisfies (P2-$(n-1)$) and (P3-$(n-1)$).

To verify (P4-$(n-1)$), let $m \in \{0, ..., n-2\}$ and h be such that $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^m \neq \emptyset$. If $m = n-2$ then $b_{j,h}\left(\mathfrak{R}_{-j}^{n-2}\right) = \rho_{j,h}\left(\mathfrak{R}_j^{n-2}\right) = 1$. If $m < n-2$, then $b_{j,h}\left(\mathfrak{R}_{-j}^m\right) = b_{j,h}^*\left(\mathfrak{R}_{-j}^m\right) = 1$. So, (P4-$(n-1)$) is satisfied. We can therefore conclude that $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{n-1}$.

We can now repeat this construction for every j. Consider the resulting profile $(s_j)_{j \in J}$ $\in S_J$. Let $s_i = (j, q)$ be an offer such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$, and let $h = [i, j, q]$. If $(\mathbf{m}, s_i, a_{-i}) \in \mathfrak{R}_I^{n-2}$ then $h \in H_j^*$ so $s_j(h) = r$ as required by (i) above. If $(\mathbf{m}, s_i, a_{-i}) \notin \mathfrak{R}_I^{n-2}$ then the intersection $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{n-2}$ is empty, hence $s_j(h) = s_j^*(h) = r$, as implied by (iii).

To conclude, the strategy profile $(s_j)_{j \in J}$ satisfies properties (1) and (2) in the statement of Lemma 5. Because $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{n-1}$, then the same lemma implies $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$. ■

[7] As before, to verify that the CPS b_j is well-defined, we need to verify that Bayes' rule is applied after all histories that has positive probability under $b_{j,\varnothing}$. The only such history is the history h following no offers to any firm. But in that case $b_{j,h} = b_{j,\varnothing}$, hence Bayes' rule is trivially respected.

Lemma 7 *Let* $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$*. If the offer* $s_i = (j, q)$ *is such that*

$$\nu(\mathbf{w}_i, \mathbf{f}_j) + q \;>\; \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)} \text{ and} \tag{8}$$

$$\phi(\mathbf{w}_i, \mathbf{f}_j) - q \;\geq\; \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j} \tag{9}$$

then $(\mathbf{m}, s_i) \in \mathfrak{R}_i^n$*.*

Proof. Because $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$, we can apply Lemma 5. Let $(s_j^*)_{j\in J}$ be a profile that satisfies conditions (1) and (2) in the statement of that Lemma. Fix a worker i and an offer $s_i = (j, q)$ such that (8) and (9) hold. Let $h = [i, j, q]$. Define s_j as $s_j(h) = i$ and $s_j(h') = s_j^*(h')$ for every $h' \in H_j$ different from h. So, the strategy s_j accepts the offer (j, q) and rejects any other offer that if accepted would improve i's payoff strictly above the status quo.

We now claim that $(\mathbf{m}, s_i) \in \mathfrak{R}_i^m$ and $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{m-1}$ for every $m \in \{1, ..., n\}$. The proof proceeds by induction on m. Given (8), the claim is easily seen to hold for $m = 1$. Suppose it is true for $m \in \{1, ..., n-1\}$. We now show that $(\mathbf{m}, s_i) \in \mathfrak{R}_i^{m+1}$ and $(\mathbf{m}, s_j) \in \mathfrak{R}_j^m$. Let b_j^* be a consistent CPS such that s_j^* and b_j^* satisfy properties (P1-$(n-1)$)-(P4-$(n-1)$). Define a new CPS b_j as follows: if $h = [i, j, q]$ then $b_{j,h}$ assigns probability 1 to

$$\left(\mathbf{m}, s_i, a_{-i}, \left(s_{\hat{j}}^*\right)_{\hat{j}\in J\setminus\{j\}}\right)$$

and if $h' \neq h$ then $b_{j,h'} = b_{j,h'}^*$. Notice that b_j is a well defined and consistent CPS.

Inequality (9) implies that $s_j(h)$ is optimal with respect to $b_{j,h}$. It follows that s_j and b_j satisfy (P1-m). Because $(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-1}$ then b_j^* satisfies (P2-$(n-1)$) and (P3-$(n-1)$). Hence, b_j^* satisfies (P2-m) and (P3-m). It follows then that b_j also satisfies (P2-m) and (P3-m). To verify (P4-m), consider first the history $h = [i, j, q]$. Because $(\mathbf{m}, s_i) \in \mathfrak{R}_i^m \subseteq \mathfrak{R}_i^{m-1}$ by the inductive hypothesis and $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n \subseteq \mathfrak{R}_I^{m-1}$ by assumption, then, by Lemma 2, we have $(\mathbf{m}, s_i, a_{-i}) \in \mathfrak{R}_I^{m-1}$. Similarly, because $(\mathbf{m}, s_j) \in \mathfrak{R}_j^{m-1}$ and $\left(\mathbf{m}, s_{\hat{j}}^*\right) \in \mathfrak{R}_{\hat{j}}^{n-1}$ for every $\hat{j} \neq j$, we have

$$\left(\mathbf{m}, s_j, \left(s_{\hat{j}}^*\right)_{\hat{j}\in J-\{j\}}\right) \in \mathfrak{R}_J^{m-1}$$

Hence, using the fact that $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{m-1}$, we obtain

$$\left(\mathbf{m}, s_i, a_{-i}, \left(s_{\hat{j}}^*\right)_{\hat{j}\in J\setminus\{j\}}\right) \in \mathfrak{R}_{-j}^{m-1}$$

so $(\mathbf{M} \times S_{-j}(h)) \cap \mathcal{P}_j(\mathbf{m}) \cap \mathfrak{R}_{-j}^{m-1} \neq \emptyset$, hence $b_{j,h}$ assigns probability 1 to $\mathfrak{R}_{-j}^{m-1}$. It follows from the definition of b_j^* and the fact that $(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-1}$ that property (P4-(m)) is verified with respect to any other history $h' \neq h$. We can conclude that $(\mathbf{m}, s_j) \in \mathfrak{R}_j^m$.

Let $\rho_i \in \Delta(\mathfrak{R}_{-i})$ assign probability 1 to

$$\left(\mathbf{m}, a_{-i}, s_j, \left(s_{\hat{j}}^*\right)_{\hat{j}\in J\setminus\{j\}}\right) \tag{10}$$

By the inductive hypothesis $(\mathbf{m}, s_i) \in \mathfrak{R}_I^m$. As shown above $(\mathbf{m}, s_j) \in \mathfrak{R}_j^m$ hence

$$\left(\mathbf{m}, s_j, \left(s_{\hat{j}}^*\right)_{\hat{j} \in J \setminus \{j\}}\right) \in \mathfrak{R}_J^m$$

By assumptioin $(\mathbf{m}, a_I) \in \mathfrak{R}_I^m$. It follows that (10) belongs to $\mathfrak{R}_{-i}^m$. Hence $\rho_i\left(\mathfrak{R}_{-i}^m\right) = 1$. Moreover, $s_i = (j, q)$ is optimal with respect to the probability ρ_i. By applying Lemma 1, we can define a consistent CPS b_i such that $b_{i,\varnothing} = \rho_i$ and such that s_i and b_i satisfy (P1-$(m+1)$)-(P4-$(m+1)$). Therefore $(\mathbf{m}, s_i) \in \mathfrak{R}_i^{m+1}$. This concludes the proof of the inductive step. We conclude that $(\mathbf{m}, s_i) \in \mathfrak{R}_I^n$. ■

If (i, j, q) is a combination that satisfies (2)-(6) in the definition of Λ^n, then we say the outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is *n-blocked* by (i, j, q). The next two lemmas are the main steps in the proof of Theorem 2.

Lemma 8 *For every $n \geq 0$ and $\mathbf{m} \in \mathbf{M}$, if $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{3n}$ then $\mathbf{m} \in \Lambda^n$.*

Proof. The proof proceeds by induction. The result is vacuously true when $n = 0$. Now assume the result is true for $n \geq 0$. Let $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \mathbf{M} \setminus \Lambda^{n+1}$. We now show that $(\mathbf{m}, a_I) \notin \mathfrak{R}^{3n+3}$. Assume that $\mathbf{m} \in \Lambda^n \setminus \Lambda^{n+1}$. This assumption is without loss of generality since, if $\mathbf{m} \notin \Lambda^n$ then $(\mathbf{m}, a_I) \notin \mathfrak{R}_I^{3n}$ by the inductive hypothesis. By the definition of Q (see section A.1) it follows that we can find a tuple (i, j, q) where $q \in Q$ is a payment that leads to no ties and such that (i, j, q) $(n+1)$-blocks $\mathbf{m}$.

So, (i, j, q) satisfies $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ and $\phi(\mathbf{w}'_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$ for all $\mathbf{w}' \in \mathbf{W}$ such that:

$$\begin{aligned}
(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) &\in \Lambda^n, \\
\mathbf{w}'_{\mu^{-1}(j)} &= \mathbf{w}_{\mu^{-1}(j)}, \text{ and} \\
\nu(\mathbf{w}'_i, \mathbf{f}_j) + q &> \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}.
\end{aligned}$$

Because $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n$, it follows that $\phi(\mathbf{w}_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$.

We now show that Lemma 6 implies $(\mathbf{m}, a_I) \notin \mathfrak{R}_I^{3n+3}$. Assume, by way of contradiction, that $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{3n+3}$ and consider the offer $s_i = (j, q)$. Because $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ and $\phi(\mathbf{w}_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$, Lemma 7 implies $(\mathbf{m}, s_i, a_{-i}) \in \mathfrak{R}_I^{3n+3}$. Hence, $(\mathbf{m}, s_i, a_{-i}) \in \mathfrak{R}_I^{3n+1}$. Consider now any profile $\mathbf{w}'$ that, as $\mathbf{w}$, satisfies

$$\begin{aligned}
\mathbf{w}'_{\mu^{-1}(j)} &= \mathbf{w}_{\mu^{-1}(j)}, \text{ and} \\
\left((\mathbf{w}', \mathbf{f}, \mu, p), s_i, a_{-i}\right) &\in \mathfrak{R}_I^{3n+1}.
\end{aligned}$$

We now show that $\mathbf{w}'$ must satisfy $\phi(\mathbf{w}'_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. By Lemma 6 this will imply $(\mathbf{m}, a_I) \notin \mathfrak{R}_I^{3n+3}$.

Let $\mathbf{m}' = (\mathbf{w}', \mathbf{f}, \mu, p)$. Because $(\mathbf{m}', s_i, a_{-i}) \in \mathfrak{R}_I^{3n+1}$, Lemma 4 implies $(\mathbf{m}', a_I) \in \mathfrak{R}_I^{3n}$. By the inductive hypothesis we conclude that $\mathbf{m}' \in \Lambda^n$. By assumption, $\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$. In addition, because s_i is optimal with then $\nu(\mathbf{w}'_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$.[8] Therefore, since (i,j,q) $(n+1)$-blocks $\mathbf{m}$, we conclude that $\phi(\mathbf{w}'_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. This concludes the proof of the result. ■

Lemma 9 *For every $n \geq 1$ and every $\mathbf{m} \in \mathbf{M}$, if $\mathbf{m} \in \Lambda^n$ then $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{1+2n}$.*

Proof. Any outcome $\mathbf{m} \in \mathbf{M}$ satisfies $(\mathbf{m}, a_I) \in \mathfrak{R}_I^1$. This follows from the fact that a_i is optimal under the belief that all offers are rejected. Therefore $\Lambda^0 = \mathbf{M} = \mathfrak{R}^1$.

Proceeding inductively, assume the result is true for $n \geq 0$. Let $\mathbf{m} = (\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ be such that $(\mathbf{m}, a_I) \notin \mathfrak{R}_I^{1+2n+2}$. We show that $\mathbf{m} \notin \Lambda^{n+1}$. It is without loss of generality to assume $\mathbf{m} \in \Lambda^n$ and $(\mathbf{m}, a_I) \in \mathfrak{R}_I^{1+2n}$ (if $(\mathbf{m}, a_I) \notin \mathfrak{R}_I^{1+2n}$ then $\mathbf{m} \notin \Lambda^n$ by the inductive hypothesis). So, $(\mathbf{m}, a_I) \in \mathfrak{R}_I^m \backslash \mathfrak{R}_I^{m+1}$, where $m \in \{1+2n, 1+2n+1\}$.

By Lemma 6 there exists an offer $s_i = (j, q)$ such that

$$\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$$

and

$$\phi(\mathbf{w}'_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$$

for every and at least one profile $\mathbf{w}'$ such that

$$\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } (\mathbf{m}', s_i, a_{-i}) \in \mathfrak{R}_I^{m-1} \tag{11}$$

where $\mathbf{m}' = (\mathbf{w}', \mathbf{f}, \mu, p)$.

We now show that (i,j,q) $(n+1)$-blocks $\mathbf{m}$. In order to reach this conclusion we need to show that every profile $\mathbf{w}''$ that, as $\mathbf{w}$, satisfies

$$\mathbf{m}'' = (\mathbf{w}'', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n,\ \mathbf{w}''_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } \nu(\mathbf{w}''_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}''_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)} \tag{12}$$

has the property that $\phi(\mathbf{w}''_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$.

The next step in the proof is to show that any $\mathbf{m}''$ that satisfies (12) must also satisfy $(\mathbf{m}'', s_i, a_{-i}) \in \mathfrak{R}_I^{m-1}$. By (11) this will imply $\phi(\mathbf{w}''_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$, establishing that (i,j,q) $(n+1)$-blocks $\mathbf{m}$.

To this end fix an outcome $\mathbf{m}''$ that satisfies (12). By the inductive hypothesis, we know that $(\mathbf{m}'', a_I) \in \mathfrak{R}_I^{1+2n}$. Because $m \leq 1+2n+1$, then $\mathfrak{R}_I^{1+2n+1} \subseteq \mathfrak{R}_I^m$ so $\mathfrak{R}_I^{1+2n} \subseteq \mathfrak{R}_I^{m-1}$. Thus, $(\mathbf{m}'', a_I) \in \mathfrak{R}_I^{m-1}$. We now show that $(\mathbf{m}'', s_i, a_{-i}) \in \mathfrak{R}_I^{m-1}$. This conclusion is reached in three steps.

[8] Furthermore, since q leads to no ties then $\nu(\mathbf{w}'_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$.

First, fix a matching outcome $\mathbf{m}'$ that satisfies (11). Because $(\mathbf{m}', s_i) \in \mathfrak{R}_i^{m-1}$ there must exist a pair $(\mathbf{m}, \tilde{s}_J) \in \mathfrak{R}_J^{m-2}$ such that $\mathbf{m} \in \mathcal{P}_i(\mathbf{m}')$ and $\tilde{s}_j$ accepts the offer $s_i = (j, q)$, i.e. $\tilde{s}_j([i, j, q]) = i$. If not, then s_i could not be optimal with respect to a consistent CPS that assigns probability 1 to $\mathfrak{R}_{-i}^{m-2}$.

Second, because $(\mathbf{m}'', a_I) \in \mathfrak{R}_I^{m-1}$ we can apply Lemma 5 and obtain a profile $(\mathbf{m}'', s_J^*)$ in $\mathfrak{R}_J^{m-2}$ with the property that every offer $(\hat{\jmath}, \hat{q})$ by player i such that $\nu(\mathbf{w}_i'', \mathbf{f}_{\hat{\jmath}}) + \hat{q} > \nu\left(\mathbf{w}_i'', \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ is rejected by $s_{\hat{\jmath}}^*$. That is, $s_{\hat{\jmath}}^*([i, \hat{\jmath}, \hat{q}]) = r$.

Third, consider a probability $\rho_i \in \Delta(\mathfrak{R}_{-i})$ that has support $(\mathbf{m}'', a_{-i}, s_J^*)$ and $(\mathbf{m}, a_{-i}, \tilde{s}_J)$. By construction, it satisfies $\rho_i\left(\mathfrak{R}_{-i}^{m-2}\right) = 1$ and the strategy $s_i = (j, q)$ is the unique best reply to ρ_i. By Lemma 1, ρ_i can be extended to a consistent CPS b_i such that $b_{i,\varnothing} = \rho_i$ and the pair (s_i, b_i) satisfies (P1-$(m-1)$)-(P4-$(m-1)$) with respect to the outcome $\mathbf{m}''$. Hence $(\mathbf{m}'', s_i, a_{-i}) \in \mathfrak{R}_I^{m-1}$.

Therefore $\mathbf{m}''$ satisfies (11). Thus $\phi(\mathbf{w}_i'', \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. Because this is true for every $\mathbf{m}''$ that satisfies (12), we conclude that (i, j, q) $(n+1)$-blocks the outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$. So $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^{n+1}$. ■

Proof of Theorem 1. Consider a market $\mathbf{M}$. Since $\mathbf{M} \times S$ is finite there exists N large enough such that either $\mathfrak{R}^N = \emptyset$ or $\mathfrak{R}^\infty = \mathfrak{R}^N$. Similarly, there exists n such that $\Lambda^\infty = \Lambda^n$. Therefore $\Lambda^n = \Lambda^\ell$ for every $\ell \geq n$. Let $\mathbf{m} \in \Lambda^\infty$. By Theorem 2, taking $\ell \geq N$ we obtain $\Lambda^\ell \subseteq \Lambda^N \subseteq \mathfrak{S}^{1+2N} \subseteq \mathfrak{S}^N = \mathfrak{S}^\infty$. Hence $\mathbf{m} \in \mathfrak{S}^\infty$. Conversely, let $\mathbf{m} \in \mathfrak{S}^\infty$. If $\ell \geq N$ then $\mathbf{m} \in \mathfrak{S}^\ell = \mathfrak{S}^{3\ell}$ and Theorem 2 implies $\mathbf{m} \in \mathfrak{S}^{3\ell} \subseteq \Lambda^\ell$. Since ℓ is arbitrary then $\mathbf{m} \in \Lambda^\infty$. ■

A.4 Proof of Theorem 3

The market is composed of two groups of agents, labelled α and β. In each group there is one worker and N firms. The set of types is $W = F = \{1, 2, \ldots, N\}$. In each group there is one firm of each type. A match between a worker of type w and a firm of type f in group α leads to premuneration denoted by $\nu_\alpha(w, f)$ and $\phi_\alpha(w, f)$, and defined as follows. For every $f < N$,

$$\nu_\alpha(w, f) = 2^w \text{ if } w = f \text{ , } \nu_\alpha(w, f) = 2^{w-1} \text{ if } w < f, \ \nu_\alpha(w, f) = 0 \text{ if } w < f$$

$$\phi_\alpha(w, f) = 0 \text{ if } w = f \quad \text{and} \quad \phi_\alpha(w, f) = -\kappa \text{ if } w \neq f$$

where $\kappa > 2^{N+10}$. For $f = N$,

$$\nu_\alpha(w, N) = 0 \text{ if } w < N \text{ and } \nu_\alpha(w, N) = 2^N \text{ if } w = N$$

$$\phi_\alpha(w, N) = 0 \text{ for all } w.$$

A match among a worker of type w and firm of type f in group β leads to premuneration payoffs:

$$\nu_\beta(w, f) = 2^w \text{ if } w = f, \ \nu_\beta(w, f) = 2^{w+2} \text{ if } w < f, \ \nu_\beta(w, f) = 0 \text{ if } w < f$$

$$\phi_\beta(w,f)=0 \text{ if } w=f \quad \text{and} \quad \phi_\beta(w,f)=-\kappa \text{ if } w\neq f$$

We assume a match between two agents that are in different groups lead to a premuneration of -1 both for the worker and the firm, irrespectively of their types. Hence, from now on, we can omit without loss of generality matches among agents in different groups.

Premuneration values are defined so that firms have an incentive to match with a worker of the same type. In group α, workers have an incentive to match to a firm of equal type. In group β, workers have an incentive to match to a firm of a lower type.We take Q to be an ε-grid where $0<\varepsilon<1$.

A matching outcome is complete-information stable if and only if in each group a worker is matched to a firm of the same type at a transfer $q\in[2^{-w},0]$ where w is the worker's type. We denote by C the set of complete-information stable outcomes. We also denote by $\mathbf{m}_\alpha(w)$ the matching outcome where the worker in group α is of type w and unmatched, while the worker in group β is of type N and matched to firm N. Similarly, $\mathbf{m}_\beta(w)$ is the matching outcome where the worker in group β is of type w and unmatched, while the worker in group α is of type N and matched to firm N. It remains to define the market $\mathbf{M}$. We let

$$\mathbf{M}=C\cup\{\mathbf{m}_\alpha(w):w\leq N\}\cup\{(\mathbf{m}_\beta(w)):w\leq N\}$$

The next lemma characterizes the set of level ℓ incomplete-information stable outcomes.

Lemma 10 *For all $\ell=0,\ldots,N$,*

$$\Lambda^\ell=C\cup\{\mathbf{m}_\alpha(w):w\leq N-\ell\}\cup\{(\mathbf{m}_\beta(w)):w\leq N-\ell\}.$$

Proof. For $\ell=0$ this holds by definition. Assume the claim is true for $\ell-1$. We wish to show it holds for ℓ. In group α, we have $\mathbf{m}_\alpha(N-\ell)\notin\Lambda^\ell$. As a way of contradiction, assume $\mathbf{m}_\alpha(N-\ell)\in\Lambda^\ell$. Consider a blocking pair between worker $w=N-\ell$ and firm $N-\ell$ at transfer $q\in(-2^{N-\ell},0)$. For any $w'<N-\ell$ we have $\nu_\alpha(w',N-\ell)+q<0$. In addition, $\nu_\alpha(w,N-\ell)+q>0$ and $\phi_\alpha(w,N-\ell)+q>0$. This shows that $\mathbf{m}_\alpha(N-\ell)\notin\Lambda^\ell$.

We now show that $\mathbf{m}_\alpha(w)\in\Lambda^\ell$ for all $w<N-\ell$. Fix $w<N-\ell$ and consider a candidate blocking pair between worker $w<N-\ell$ and a firm f at transfers q. We must have $q<0$ and $\nu_\alpha(w,f)-q>0$. Hence $f\leq w$. Because $\nu_\alpha(N-\ell,f)\geq\nu_\alpha(w,f)$, then $\nu_\alpha(N-\ell,f)-q>0$. Hence both types w and $N-\ell$ would profit from the blocking pair. It follows that property (4) in the definition of Λ^ℓ is violated. Hence $\mathbf{m}_\alpha(w)\in\Lambda^\ell$.

We now consider group β. We have $\mathbf{m}_\beta(N-\ell)\notin\Lambda^\ell$. Similarly to what we have shown in group α, by the inductive hypothesis, worker $w=N-\ell$ and firm $N-\ell$ can form a blocking pair at transfer $-q\in\left(2^{N-\ell},\frac{3}{2}2^{N-\ell-1}\right)$. Hence $\mathbf{m}_\beta(N-\ell)\notin\Lambda^\ell$. It remains to show that $\mathbf{m}_\beta(w)\in\Lambda^\ell$ for all $w<N-\ell$. To this end, consider a worker of type $w<N-\ell$. Consider a candidate blocking pair between w and firm $f\leq w$. Because

$\nu_\beta(w,f) \le \nu_\beta(N-\ell, f)$, then the candidate blocking pair would be profitable both for type w and $N-\ell$. As in the case of group α, it follows that property (4) in the definition of Λ^ℓ is violated. This concludes the proof of the claim. ■

We now study rationalizability. We first consider group α. To simplify the notation, we let $\mathfrak{R}^n_{\alpha,k}$ be the set of pairs $(\mathbf{m}_\alpha(w), s_k) \in \mathfrak{R}^n_k$ where $w \in W$ and $s_k \in S_k$. We also define

$$S^n_{\alpha,k}(w) = \left\{ s_k \in S_k : (\mathbf{m}_\alpha(w), s_k) \in \mathfrak{R}^n_{\alpha,k} \right\}$$

Without risk of confusion, we denote a firm j by its type $f = \mathbf{f}_j$ and denote by i the worker in group α.

Lemma 11 *Fix $w, w' \in W$. If $a \in S^m_{\alpha,i}(w) \cap S^m_{\alpha,i}(w')$ then $S^{m+1}_{\alpha,f}(w) = S^{m+1}_{\alpha,f}(w')$ for every f.*

Proof. Since $a \in S^m_{\alpha,i}(w) \cap S^m_{\alpha,i}(w')$, Lemma 4 implies $S^{m+1}_{\alpha,f}(w) \neq \emptyset$ and $S^{m+1}_{\alpha,f}(w') \neq \emptyset$. In addition, we have $\mathcal{P}_f(\mathbf{m}_\alpha(w)) = \mathcal{P}_f(\mathbf{m}_\alpha(w'))$ since in both matching outcomes firm f is unmatched. It is immediate to verify that this implies $S^{m+1}_{\alpha,f}(w) = S^{m+1}_{\alpha,f}(w')$. ■

The next lemmas characterizes the sequence $\left(S^{2n}_{\alpha,k}(w)\right)_n$.

Lemma 12 *Let $a \in S^m_{\alpha,i}(w) \cap S^m_{\alpha,i}(w')$, where $w < w'$. Then for every f there exists a strategy $s_f \in S^{m+1}_{\alpha,f}(w)$ such that for every $s_i = (f,q)$, if $\nu_\alpha(w,f) + q > 0$ then s_i is rejected by s_f.*

Proof. Fix f and consider an offer $s_i = (f,q)$ such that $\nu_\alpha(w,f) + q > 0$. We claim that for every $r \le m$, $\{f\} \neq \left\{\tilde{w} : s_i \in S^r_{\alpha,i}(\tilde{w})\right\}$. To this end, notice that if $\nu_\alpha(w,f) + q > 0$ then $f \le w$. Moreover, since $w' > w$, then $\nu_\alpha(w',f) + q > 0$. Assume, without loss of generality, that $\left\{\tilde{w} : s_i \in S^r_{\alpha,i}(\tilde{w})\right\}$ is nonempty. This means s_i is accepted by some strategy $s_f \in S^{r-1}_{\alpha,f}(w'')$, where $w'' \in \left\{\tilde{w} : s_i \in S^r_{\alpha,i}(\tilde{w})\right\}$. Lemma 11 implies $S^{r-1}_{\alpha,f}(w'') = S^{r-1}_{\alpha,f}(w) = S^{r-1}_{\alpha,f}(w')$. By assumption we have $a \in S^r_{\alpha,i}(w)$. Let b_i be a corresponding conditional probability system with respect to which abstaining is optimal for i. As in the proof of Lemma 4, we can obtain a new conditional probability system b'_i that agrees with b_i except for its marginal on $S_{\alpha,f}$, where b'_i assigns probability 1 to a strategy in $s_f \in S^{r-1}_{\alpha,f}(w)$ that accepts offer s_i. Then, s_i is optimal with respect to b'_i, and we obtain $s_i \in S^r_{\alpha,i}(w)$. By repeating the same argument for w', we obtain $s_i \in S^r_{\alpha,i}(w')$. So, $\{w, w'\} \subseteq \left\{\tilde{w} : s_i \in S^r_{\alpha,i}(\tilde{w})\right\}$.

Let H^* be the set of histories h obtained when an offer $s_i = (f,q)$ is made to f and satisfies $\nu_\alpha(w,f) + q > 0$. Given $h \in H^*$, let $r(h)$ be the largest $r \le m$ such that $\{\tilde{w} : s_i \in S^r_i(\tilde{w})\}$. The claim implies we can construct a probability $\rho_h \in \Delta(\mathbf{M} \times S_{-f})$ that assigns probability 1 to $(\{\mathbf{m}_\alpha(\tilde{w})\} \times S_{-f}) \cap \mathfrak{R}^{r(h)}_{-f}$, where $\tilde{w}$ is a type other than f. Let $(\rho_h)_{h \in H^*}$ be the resulting array of probabilities. As in the proof of Lemma 4, this array can be extended to a consistent conditional probability system b_f that satisfies properties (P2-$(m+1)$)-(P4-$(m+1)$). Let s_f be a strategy that is optimal with respect to b_f. Then $s_f \in S_f^{m+1}(w)$ and s_f rejects any offer $s_i = (f, q)$ such that $\nu_\alpha(w, f) + q > 0$. ■

Lemma 13 *For every $n \in \{0, \ldots, N-1\}$,*

1. $S_{\alpha,i}^{2(n+1)}(w) = \emptyset$ *for all $w > N - n$;*
2. $S_{\alpha,i}^{2(n+1)}(N-n) \neq \emptyset$ *and $s_i \in S_{\alpha,i}^{2(n+1)}(N-n)$ only if s_i is an offer to firm $N-n$;*
3. $a \in S_{\alpha,i}^{2(n+1)}(w)$ *for all $w < N - n$;*
4. *For every $w < N - n$ and every f there exists $s_f \in S_{\alpha,f}^{2(n+1)}(w)$ with the property that if $s_i = (f, q)$ is such that $\nu_\alpha(w, f) + q > 0$ then s_i is rejected by s_f.*

Proof. The proof is by induction on n. We first consider $n = 0$. For every w, $s_{\alpha,i} = (f, q) \in S_{\alpha,i}^1(w)$ if and only if $\nu_\alpha(w, f) + q > 0$. It is also immediate to verify that for every $f < N$ and w, $s_f \in S_{\alpha,f}^1(w)$ if and only if s_f accepts any offer (f, q) where $q < -\kappa$ and rejects any offer where $q > 0$. If $f = N$ then $s_f \in S_{\alpha,f}^1(w)$ if and only if s_f accepts any offer where $q < 0$ and rejects any offer where $q > 0$.

When $n = 0$ (1) holds trivially. To see that (2) holds, notice that for a worker of type N, an offer to firm $f = N$ at wage $q \in (-1, 0)$, if accepted, provides a payoff greater than the payoff the worker could obtain from any other offer to a firm $f \neq N$ that is accepted by some strategy in $S_{\alpha,f}^1(N)$. In addition, an offer to $f = N$ at transfer $q < 0$ is accepted by every strategy $s_f \in S_{\alpha,f}^1(N)$. Hence, $s_i \in S_{\alpha,i}^2(N)$ only if s_i is an offer to firm N. Claim (3) follows from the fact that for any $w < N$ and any offer $s_i = (f, q)$ such that $\nu_\alpha(w, f) + q > 0$, s_i is rejected by some strategy in $S_{\alpha,f}^1(w)$, together with an application of Lemma 5. Claim (4) follows from Lemma 12.

Assume (1)-(4) hold for n. Conditions (1) and (2) imply, by Lemma 4,

$$S_{\alpha,i}^{2n+3}(w) = \emptyset \text{ for every } w > N - n - 1 \tag{13}$$

Condition (4) and Lemma 3 imply

$$a \in S_{\alpha,i}^{2n+3}(w) \text{ for all } w < N - n. \tag{14}$$

Let $s_i = (f, q)$ be an offer where $f = N - n - 1$ and $q \in (-1, 0)$. Lemma 7 and the fact that $a \in S_{\alpha,i}^{2n+2}(N - n - 1)$ imply $s_i \in S_{\alpha,i}^{2n+2}(N - n - 1)$. In addition, we have $s_i \notin S_{\alpha,i}^{2n+2}(w')$ for any $w' \neq N - n - 1$. For $w' > N - n - 1$ this follows directly from (1) and (2). For $w' < N - n - 1$ this follows from the fact that $\nu_\alpha(w', f) + q \leq 0$. Therefore, given $f = N - n - 1$, any strategy $s_f \in S_{\alpha,f}^{2n+3}(N - n - 1)$ must accept offer s_i.

Let $w < N-n-1$. By the inductive hypothesis we have $a \in S_{\alpha,i}^{2n+2}(w) \cap S_{\alpha,i}^{2n+2}(N-n-1)$. Hence, by applying Lemma 12 with $m = 2n+2$ we obtain that for every f there exists $s_f \in S_{\alpha,f}^{2n+3}(w)$ with the property that if $s_i = (f,q)$ is such that $\nu_\alpha(w,f) + q > 0$ then s_i is rejected by s_f.

We now verify that conditions (1)-(4) hold with respect to $n+1$. By (13) we have $S_i^{2n+4}(w) = \emptyset$ for every $w > N-n-1$. As established above, any offer $s_i = (f,q)$ where $f = N-n-1$ and $q \in (-1,0)$ is accepted by any strategy s_f in $S_f^{2n+3}(w)$. Hence $a \notin S_i^{2n+4}(N-n-1)$ and $s_i \in S_i^{2n+4}$ only if s_i is an offer to $f = N-n-1$. Hence (2) holds. In addition, for every $w < N-n-1$ and every $s_i = (f,q)$ such that $\nu_\alpha(w,f) - q > 0$, s_i is rejected by some strategy $s_f \in S_f^{2n+3}(w)$. Hence $a \in S^{2n+4}(w)$. This implies (3). To establish property (4), for every $w < N-n-1$ we have $a \in S_i^{2n+3}(w) \cap S_i^{2n+3}(N-n-1)$. Hence, by applying, as above, Lemma 12 with $m = 2n+3$ we obtain that for every f there exists $s_f \in S_f^{2n+4}(w)$ with the property that if $s_i = (f,q)$ is such that $\nu_\alpha(w,f) + q > 0$ then s_i is rejected by s_f. ■

We now consider group β. Let $\mathfrak{R}_{\beta,k}^n$ be the set of pairs $(\mathbf{m}_\beta(w), s_k) \in \mathfrak{R}_k^n$ where $w \in W$ and $s_k \in S_k$. Let

$$S_{\beta,k}^n(w) = \left\{ s_k \in S_k : (\mathbf{m}_\beta(w), s_k) \in \mathfrak{R}_{\beta,k}^n \right\}$$

The next lemmas characterizes the sequence $\left(S_{\beta,k}^{3n}(w)\right)_n$.

Lemma 14 *Let $a \in S_{\beta,i}^m(w) \cap S_{\beta,i}^m(w')$, where $w < w'$. Then for every f there exists a strategy $s_f \in S_{\beta,f}^{m+1}(w)$ such that for every $s_i = (f,q)$ such that $\nu_\alpha(w,f) + q > 0$, s_i is rejected by s_f.*

Proof. The result follows by replicating the proof of Lemma 12. ■

Lemma 15 *For every $n \geq 0$,*

1. $S_{\beta,i}^{3(n+1)}(w) = \emptyset$ *for all* $w > N-n$*;*
2. $S_{\beta,i}^{3(n+1)}(N-n) \neq \emptyset$ *and* $a \notin S_{\beta,i}^{3(n+1)}(N-n) \neq \emptyset$*;*
3. $a \in S_{\beta,i}^{3(n+1)}(w)$ *for all* $w < N-n$.
4. *For every $w < N-n$ and every f there exists $s_f \in S_{\beta,f}^{3(n+1)}(w)$ with the property that if $s_i = (f,q)$ is such that $\nu_\alpha(w,f) + q > 0$ then s_i is rejected by s_f.*

Proof. The proof is by induction on n. We first describe $S_{\beta,i}^1(w)$ and $S_{\beta,f}^1(w)$. It is immediate to verify that for every w we have $s_i = (f,q) \in S_{\beta,i}^1(w)$ if and only if $f \leq w$ and $\nu_\alpha(w,f) + q > 0$. It is also immediate to verify that for every f and w, $s_f \in S_{\beta,f}^1(w)$ if and only if s_f accepts any offer (f,q) where $q < -\kappa$ and rejects any offer where $q > 0$.

We now consider $S^2_{\beta,i}(w)$ and $S^2_{\beta,f}(w)$. For every w, $s_i = (f,q) \in S^1_{\beta,i}(w)$ if and only if $\nu_\alpha(w,f) + q > 0$ and $q < 0$. For every $f < N$, $s_f \in S^2_{\beta,f}(w)$ if and only if s_f accepts any offer (f,q) where $q < -\kappa$ and rejects any offer where $q > 0$. For $f = N$, every $s_f \in S^2_{\beta,f}(w)$ must accept any offer $s_i = (N,q)$ where $q < 0$.

When $n = 0$ (1) holds vacuously. To see that (2) holds, notice that for a worker of type N, an offer to firm $f = N$ at transfer $q \in (-1,0)$ provides a strictly positive payoff and is accepted by any strategy $s_f \in S^2_{\beta,f}(N)$, so $a \notin S^3_{\beta,i}$. In addition $a \in S^2_{\beta,i}(N)$ hence $S^3_{\beta,i}(N) \neq \emptyset$. This concludes the proof of (2). Claim (3) follows from the fact that for any $w < N$ and any offer $s_i = (f,q)$ such that $\nu_\alpha(w,f) + q > 0$, s_i is rejected by some strategy in $S^2_{\beta,f}(w)$, together with an application of Lemma 5. To verify (4), let $w < N$. As shown above, $a \in S^2_{\beta,i}(w) \cap S^2_{\beta,i}(N)$. Hence, Lemma 14 implies (4).

Before proceeding with the inductive steps we prove the following fact: *If $m \geq 3$, $a \in S^m_{\beta,i}(w)$ and $S^m_{\beta,i}(w+1) \neq \emptyset$, then for every f there exists a strategy $s_f \in S^{m+1}_{\beta,f}(w)$ with the property that any offer $s_i = (f,q)$ such that $\nu_\alpha(w,f) + q > 0$ is rejected by s_f*

Recall that any offer $s_i = (f,q)$ such that $f \leq w$ and $q \in (2^{-f}, 0]$ satisfies $s_i \in S^m_{\beta,i}(\tilde{w})$ for $\tilde{w} = f$. This follows from Lemma 4. Now let $s_i = (f,q)$ be an offer such that $\nu_\alpha(w,f) + q > 0$. Thus $f \leq w$. Let r be the largest $\tilde{r} \leq m$ such that $\left\{\tilde{w} : s_i \in S^{\tilde{r}}_{\beta,i}(\tilde{w})\right\} \neq \emptyset$. We now argue that $\{f\} \neq \left\{\tilde{w} : s_i \in S^r_{\beta,i}(\tilde{w})\right\}$. Consider first the case where $q > 0$ or $q \leq -2^f$. In this case $s_i \notin S^2_{\beta,i}(\tilde{w})$ if $\tilde{w} = f$ but $\left\{\tilde{w} : s_i \in S^2_{\beta,i}(\tilde{w})\right\} \neq \emptyset$. Now suppose $q \in (-2^f, 0]$. We now show that $s_i \in S^m_{\beta,i}(w+1)$. As shown above, for every offer $s'_i = (f', q')$ such that $f' \leq w$ and $q \leq 2^{-f'}$ or $q' > 0$, there exists a strategy $s_{f'} \in S^{m-1}_{\beta,f'}(w+1)$ that rejects s'_i. In addition, because $a \in S^m_i(w)$, then for any offer $s'_i = (f', q')$ such that $f' \leq w$ and $q' \in (2^{-f'}, 0]$ there is a strategy $s_{f'} \in S^{m-1}_{f'}(w) = S^{m-1}_{f'}(w+1)$ rejecting s_i. Finally, since $s_i \in S^m_i(\tilde{w})$ for $\tilde{w} = f$, there exists a strategy in $S^{m-1}_f(\tilde{w}) = S^{m-1}_f(w)$ that accepts s_i. By applying Lemma 3 we can therefore find a profile of firms' strategies in $\mathfrak{R}^n_\beta$ such that among all offers $s'_i = (f', q')$ such that $f' \leq w$ and s'_i is profitable for $w+1$, only offer s_i is accepted. Let b_i a consistent conditional probability system for i that assigns probability 1 to such a profile of strategies. To show that s_i is optimal with respect to b_i it remains to prove that no offer to firm $f = w+1$ is optimal under b_i. Offer s_i, if accepted, provides a payoff greater or equal to $2^{w+3} - q \geq 2^{w+3} - 2^w$. This payoff is higher then the payoff worker $w+1$ can obtain from any offer to firm $f' = w+1$ at a negative wage. Hence s_i is optimal with respect to b_i. Hence $s_i \in S^m_i(w+1)$. It follows that $\{f\} \neq \left\{\tilde{w} : s_i \in S^r_{\beta,i}(\tilde{w})\right\}$. Given this fact, we can now replicate the argument used in the proof of Lemma 12 to conclude the proof of the claim.

Assume claims (1)-(4) to be true for n. Conditions (1) and (2) imply, by Lemma 4,

$$S^{3n+4}_{\beta,i}(w) = \emptyset \text{ for every } w \geq N - n \tag{15}$$

Condition (4) and Lemma 5 imply

$$a \in S^{3n+4}_{\beta,i}(w) \text{ for all } w < N-n. \tag{16}$$

We now show that for every $w < N-n$ there exists $s_f \in S^{3n+4}_{\beta,f}(w)$ such that for every offer $s_i = (f,q)$, if $\nu_\alpha(w,f)+q > 0$ then s_f rejects s_i. In the case $w = N-n-1$, the claim follows from $a \in S^{3n+3}_{\beta,i}(N-n-1)$, $S^{3n+3}_{\beta,i}(N-n) \neq \emptyset$ and the fact proved above. If $w < N-n-1$, the claim follows from Lemma 14 together with $a \in S^{3n+3}_{\beta,i}(w) \cap S^{3n+3}_{\beta,i}(N-n-1)$.

We now study the $(3n+5)$-th step of the rationalizability procedure. It follows from (15) and (16) that $S^{3n+5}_{\beta,i}(w) = \emptyset$ for every $w \geq N-n$. In addition for every $w < N-n$, every offer that is profitable for w is rejected by some strategy $s_f \in S^{3n+4}_{\beta,f}(w)$. Hence Lemma 5 implies $a \in S^{3n+5}_{\beta,i}(w)$.

Now consider an offer $s_i = (N-n-1,q)$ where $q \in (-1,0)$. Because $a \in S^{3n+4}_{\beta,i}(N-n-1)$ then Lemma 7 implies $s_i \in S^{3n+4}_{\beta,i}(N-n-1)$. In addition, $s_i \notin S^{3n+4}_{\beta,i}(w)$ for any $w \neq N-n-1$. For $w < N-n-1$ this follows from $\nu_\alpha(w,f)+q < 0$. For $w > N-n-1$ we have $S^{3n+4}_{\beta,i}(w) = \emptyset$. It follows that any strategy $s_f \in S^{3n+5}_{\beta,f}(N-n-1)$ must accept offer s_i. Finally, by applying, as above, Lemma 14, we obtain that for every $w < N-n-1$ there exists $s_f \in S^{3n+5}_{\beta,f}(w)$ such that for every offer $s_i = (f,q)$, if $\nu_\alpha(w,f)+q > 0$ then s_f rejects s_i.

We now verify that conditions (1)-(4) hold with respect to $n+1$. By (15) we have $S^{3n+6}_{\beta,i}(w) = \emptyset$ for every $w > N-n-1$. As established above, any offer $s_i = (f,q)$ where $f = N-n-1$ and $q \in (-1,0)$ is accepted by any strategy s_f in $S^{3n+5}_{\beta,f}(w)$. Hence $a \notin S^{3n+6}_{\beta,i}(N-n-1)$. Hence (2) holds. In addition, for every $w < N-n-1$ and every $s_i = (f,q)$ such that $\nu_\alpha(w,f)-q > 0$, s_i is rejected by some strategy $s_f \in S^{3n+5}_{\beta,f}(w)$. Hence Lemma 5 implies $a \in S^{3n+6}_{\beta,i}(w)$. This implies (3). Property (4) follows the fact we established in the first part of the proof, $a \in S^{3n+6}_{\beta,i}(N-n-2)$ and $S^{3n+6}_{\beta,i}(N-n-1) \neq \emptyset$.
■

We can now conclude the proof of the theorem. Let $\mathfrak{S}^n_\alpha = \left\{w : (\mathbf{m}_\alpha(w), a) \in \mathfrak{R}^n_{i,\alpha}\right\}$ and $\mathfrak{S}^n_\beta = \left\{w : (\mathbf{m}_\alpha(w), a) \in \mathfrak{R}^n_{i,\beta}\right\}$. So, $\mathfrak{S}^n = C \cup \mathfrak{S}^n_\alpha \cup \mathfrak{S}^n_\beta$. Lemma 13 shows that for every n we have $a \in S^{2n+1}_{\alpha,i}(N-n)$ (as implied by $S^{2n+2}_{\alpha,i}(N-n) \neq \emptyset$ and Lemma 4) and $a \notin S^{2n+2}_{\alpha,i}(N-n)$. Hence

$$\begin{aligned}
\mathfrak{S}^{2n+1}_\alpha &= \{\mathbf{m}_\alpha(1), \ldots, \mathbf{m}_\alpha(N-n)\} \\
\mathfrak{S}^{2n+2}_\alpha &= \{\mathbf{m}_\alpha(1), \ldots, \mathbf{m}_\alpha(N-n-1)\} \\
\mathfrak{S}^{2n+3}_\alpha &= \{\mathbf{m}_\alpha(1), \ldots, \mathbf{m}_\alpha(N-n-1)\}
\end{aligned}$$

Similarly, 15 implies, for every n,

$$\begin{aligned}
\mathfrak{S}_\beta^{3n} &= \{\mathbf{m}_\beta(1), \ldots, \mathbf{m}_\beta(N-n)\} \\
\mathfrak{S}_\beta^{3n+1} &= \{\mathbf{m}_\beta(1), \ldots, \mathbf{m}_\beta(N-n)\} \\
\mathfrak{S}_\beta^{3n+2} &= \{\mathbf{m}_\beta(1), \ldots, \mathbf{m}_\beta(N-n)\} \\
\mathfrak{S}_\beta^{3n+3} &= \{\mathbf{m}_\beta(1), \ldots, \mathbf{m}_\beta(N-n-1)\}
\end{aligned}$$

Given $\ell = 0, \ldots, N$ we have $\mathfrak{S}^n \subseteq \Lambda^\ell$ if and only if $\mathfrak{S}_\alpha^n \subseteq \mathfrak{S}_\beta^n \subseteq \{\mathbf{m}_1(1), \ldots, \mathbf{m}_1(N-\ell)\}$, i.e. if and only if $n \geq 3\ell$. Similarly, $\Lambda^\ell \subseteq \mathfrak{S}^n$ if and only if $n \leq 1 + 2\ell$. Hence $B(\ell) = 3\ell$ and $b(\ell) = 1 + 2\ell$.

A.5 Proofs of Other Results

Proof of Theorem 4. Let $\mathbf{m}$ be stable under forward induction. So $(\mathbf{m}, a_I) \in \mathfrak{R}_I^\infty$. Let n be such that $\mathfrak{R}^{n-1} = \mathfrak{R}^\infty$. Since $(\mathbf{m}, a_I) \in \mathfrak{R}_I^n$, by Lemma 6 there exists a strategy profile (s_j^*) such that $(\mathbf{m}, s_j^*) \in \mathfrak{R}_j^{n-1} = \mathfrak{R}_j^\infty$ for every j and $s_j^*(h) = r$ for every j and every history $h = [i, j, q]$ that satisfies $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$. Lemma 2 implies $\left(\mathbf{m}, a_I, (s_j^*)_{j\in J}\right) \in \mathfrak{R}^{n-1} = \mathfrak{R}^\infty$. ■

Proof of Theorem 5. Suppose $\mathbf{m}$ is not strictly stable. Let $(\mathbf{w}, s_i) \in \mathfrak{R}_i^\infty$, where $s_i = (j, q)$. We can choose $n \geq 0$ large enough so that $\mathfrak{R}^\infty = \mathfrak{R}^n = \mathfrak{R}^{n-2}$. There must exist a strategy s_j such that $(\mathbf{w}, s_j) \in \mathfrak{R}_j^\infty$ and s_j accepts the offer (j, q), i.e., $s_j(\{(i, j, q)\}) = i$. Let b_j a CSP such that s_j and b_j satisfy (P1-n)-(P4-n). By (P4-n) it must be that $b_{j,h}\left(\mathfrak{R}_{-j}^\infty\right) = 1$. Hence, there is a profile $\mathbf{w}' \in \mathbf{W}$ in the support of $b_{j,h}$ such that $\phi(\mathbf{w}_i', \mathbf{f}_j) - q \geq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$, $\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$ and $(\mathbf{w}', s_i) \in \mathfrak{R}_i^\infty$. By Lemma 2, $(\mathbf{w}', a_I) \in \mathfrak{R}_I^\infty$. Hence $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^\infty$. This concludes the proof.

We now show the "if" part of the proof. Suppose we can find a tuple (i, j, q) and a profile $\mathbf{w}' \in \mathbf{W}$ such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + q$, $\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$, and

$$\begin{aligned}
\phi(\mathbf{w}_i', \mathbf{f}_j) - q &\geq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}, && (17)\\
\nu(\mathbf{w}_i', \mathbf{f}_j) + q &> \nu\left(\mathbf{w}_i', \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}, \text{ and} && (18)\\
(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) &\in \Lambda^\infty.
\end{aligned}$$

Because $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^\infty$ then $(\mathbf{w}', a_I) \in \mathfrak{R}_I^\infty$. Let $s_i = (j, q)$. Then, (17), (18) and Lemma 7 imply $(\mathbf{w}', s_i) \in \mathfrak{R}_i^\infty$. We now show that $(\mathbf{w}, s_i) \in \mathfrak{R}_i^\infty$, concluding that (1) must be violated. The proof is similar to the proof of Lemma 7. Because $(\mathbf{w}', s_i) \in \mathfrak{R}_i^\infty$, there must exist a strategy s_j such that s_j accepts the offer (j, q) and $(\mathbf{w}', s_j) \in \mathfrak{R}_j^\infty$. Because $(\mathbf{w}, a_I) \in \mathfrak{R}_I^\infty$, by Lemma 5 we can find a strategy profile $\left(s_j^*\right)_{j\in J}$ such that $\left(\mathbf{w}, s_j^*\right) \in \mathfrak{R}_j^\infty$

for every j and such that any offer $(\hat{\jmath}, \hat{q})$ by worker i that, if accepted, would improve worker i's payoff above the default allocation is rejected by strategy $s_{\hat{\jmath}}^{*}$. Now define a new strategy s_j' as follows. At the history h corresponding to the offer (j, q) from worker i, let $s_j'(h) = s_j(h) = i$. At every other history h, $s_j'(h) = s_j^{*}(h)$. By Lemma 3, $\left(\mathbf{w}, s_j'\right) \in \mathfrak{R}_j^{\infty}$. Let b_i' be a conditional probability system such that $b_{i,\varnothing}'$ is concentrated on

$$\left(\mathbf{w}, a_{-i}, s_j', \left(s_{\hat{\jmath}}^{*}\right)_{\hat{\jmath} \in J - \{j\}}\right).$$

Under b_i' the offer $s_i = (j, q)$ is a strict best response. It is immediate to check that s_i and b_i' satisfy (P1-n)-(P4-n), where $\mathfrak{R}^n = \mathfrak{R}^{\infty}$. Hence $(\mathbf{w}, s_i) \in \mathfrak{R}_i^{\infty}$. Thus, m is not strictly stable. ■

References

Battigalli, P. (1997): “Strategic Rationality Orderings and the Best Rationalization Principle,” *Games and Economic Behavior*, 13(2).

Battigalli, P. and A. Prestipino. (2013): “Transparent Restrictions on Beliefs and Forward-induction Reasoning in games with Asymmetric Information,” *Advances in Theoretical Economics*, 13(1).

Battigalli, P. and A. Friedenberg. (2013): “Forward Induction Reasoning Revisited,” *Theoretical Economics*, 7(1).

Battigalli, P. and M. Siniscalchi. (2002): “Strong belief and Forward Induction Reasoning,” *Journal of Economic Theory*, 106(2).

Battigalli, P. and M. Siniscalchi. (2003): “Rationalization and Incomplete Information,” *Advances in Theoretical Economics*, 3(1).

Bikhchandani, S. (2014): “Two-sided Matching with Incomplete Information,” *mimeo.*

Chakraborty, A., A. Citanna, and M. Ostrovsky (2010): “Two-Sided Matching With Interdependent Values,” *Journal of Economic Theory*, 145(1).

Chade, H. (2006): “Matching with Noise and the Acceptance Curse,” *Journal of Economic Theory*, 129(1).

Chade, H., G. Lewis, and L. Smith (2014): “Student Portfolios and the College Admission Problem,” *The Review of Economic Studies*, 81.

Chen, Y-C, and G. Hu (2017): “Learning by Matching,” *mimeo.*

Cho, I. and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics*, 52(2).

de Clippel, G. (2007): "The Type-agent Core for Exchange Economies with Asymmetric Information," *Journal of Economic Theory*, 135(1).

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

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

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

Fudenberg, D. and D. Levine (1993): "Self-confirming Equilibrium," *Econometrica*, 61(3).

Govindan, S. and R. Wilson (2009): "On Forward Induction," *Econometrica*, 77(1).

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

Kohlberg, E. (1981): "Some Problems with the Concept of Perfect Equilibrium." *Rapp. Rep. NBER Conf. Theory Gen. Econ. Equilibr. K. Dunz. N. Singh, Univ. Calif. Berkeley.*

Kohlberg, E., and J. F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54(5).

Liu, Q., (2017): "Stable Belief and Stable Matching," *mimeo.*

Liu, Q., G. J. Mailath, A. Postlewaite and L. Samuelson. (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82(2).

Man, P. "Forward Induction Equilibrium," *Games and Economic Behavior*, 75(1).

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

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

Peivandi, A. (2013): "Participation and Unbiased Pricing in CDS Settlement Mechanisms," *mimeo.*

Reny, P. J. (1992): "Backward Induction, Normal Form Perfection and Explicable Equilibria," *Econometrica*, 60(3).

Roth, A. (1989): “Two-sided Matching with Incomplete Information about others’ Preferences,” *Games and Economic Behavior*, 1(2).

Roth, A. and M. Sotomayor (1992): *Two-sided Matching: A Study in Game-theoretic Modeling and Analysis.* Cambridge University Press.

Serrano, R. and R. Vohra (2007): “Information Transmission in Coalitional Voting Games,” *Journal of Economic Theory*, 134(1).

Sobel, J., L. Stole and I. Zapater (1990): “Fixed-Equilibrium Rationalizability in Signaling Games,” *Journal of Economic Theory*, 52.

Shapley, L.S. and M. Shubik. (1971): “The Assignment Game I: The Core,” *International Journal of Game Theory*, 1(1).

Van Damme, E. (1989): “Stable Equilibria and Forward Induction,” *Journal of Economic Theory*, 48.

Vohra, R. (1999): “Incomplete Information, Incentive Compatibility, and the Core,” *Journal of Economic Theory*, 86(1).

Wilson, R. (1978): “Information, Efficiency, and the Core of an Economy,” *Econometrica*, 46(4).