Stable Matching under Forward-Induction Reasoning*

Luciano Pomatto
MEDS, Kellogg School of Management, Northwestern University

December 31, 2014

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 non-cooperative and 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 a class of 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 non-cooperative approach in matching markets.

*Please visit http://www.kellogg.northwestern.edu/faculty/pomatto/index.htm for the most current version. I am indebted to Alvaro Sandroni and Marciano Siniscalchi for many useful discussions and suggestions. I also thank Willemien Kets and Rakesh Vohra for helpful comments. All errors are my own.

1 Introduction

Over the past decades, models of matching markets have been applied to the design of college admissions, the analysis of housing markets, the study of labor and marriage markets, and the architecture of kidney exchange protocols. 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. In addition to added realism, incomplete information is critical to understand relevant phenomena such as costly signaling: workers would only put effort in signaling if their abilities were not entirely known ex-ante.

The crucial difficulty in the study of matching markets with incomplete information is 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.

This paper considers a non-cooperative approach to two-sided matching markets with incomplete information. I study a class of markets where one side (workers) has private information about characteristics of its members (for instance, their skills), that are payoff-relevant for both sides. 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 occurs through take-it-or-leave-it offers, which involve a small cost. If no offers are made, or all offers are rejected, then the default allocation is implemented.

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 I 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. Players' thought processes are modelled explicitly using the epistemic framework introduced by Battigalli and Siniscalchi (2002,2007). This approach formalizes players' thought processes by describing their hierarchies of beliefs about other players' strategies and payoff-types, and how these beliefs are updated throughout the game. Stability is defined by imposing three requirements on players' actions and hierarchies of beliefs, which, informally, are:

1. Agents are rational and abstain from making offers;
2. Players expect no offer to be made by other agents;
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 one, 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 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) (LMPS, henceforth). In particular, the result shows that stability under forward induction can be applied through a simple algorithm. The result provides a new connection between cooperative and noncooperative approaches in matching markets.

It follows by this characterization and the results in LMPS that the set of stable outcomes is always non-empty and is efficient under supermodularity. That is, when preferences are strictly supermodular, a stable matching outcome is ex-post efficient.

The noncooperative approach allows for a specific understanding of what thought processes can lead to stability 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 noncooperative approach reveals new difficulties in reaching stability and efficiency in matching markets. In this paper it is assumed that workers know each others' characteristics. This assumption appears to be reasonable in some cases (e.g., markets for experts) but not in others (e.g. large markets) where workers are not expected to know each other better than firms. This assumption, which does not appear in LMPS, compensates for the fact that when blocking pairs are formed through offers, agents must take into account the payoff other agents are currently obtaining from the status quo. This consideration is absent from a cooperative model, where players find themselves exogenously involved in a blocking pair.

The paper is organized as follows. Section 2 introduces the basic two-sided matching environment studied in this paper. Section 3 introduces incomplete information and presents a class of markets with one-sided payoff uncertainty. Section 4 describes how blocking pairs are formed. Section 5 presents the epistemic framework used to analyze the game. Section 6 formally defines stability under forward induction. Section 7 describes incomplete-informations tability, the main characterization result and provides examples. Section 8 discusses the result and provides extensions. Section 9 concludes.

1.1 Related Literature

This paper is related to the literature on the core. Starting with Wilson (1978), new notions of core for environments with incomplete information have been introduced in de Clippel (2007), Dutta and Vohra (2005), Myerson (2007), Peivandi (2013), Serrano and Vohra (2007) and Vohra (1999), among others. The current paper shares some similarities with Serrano and Vohra (2007), where blocking coalitions are formed non-cooperatively, as the equilibrium outcomes of a voting game.

A recent literature studies two-sided matching under incomplete information. Roth (1989) shows that there is no mechanism for which there exists a Bayesian Nash equilibrium whose outcome is always ex-post stable. In additions to LMPS, two-sided matching with incomplete information is studied in Chade (2006), Chade, Lewis and Smith (2011), Ehlres and Masso (2007) and Hoppe, Moldovanu and Sela (2009), among others. In particular, Chakraborty, Citanna and Ostrovsky (2010) provide a new notion of stability for centralized markets. In addition, Bikhchandani (2014) analyzes incomplete-information stability in markets without transferable utility and markets with two-sided incomplete information.

This paper builds upon the literature on forward-induction reasoning. The best rationalization principle was introduced in Battigalli (1996). Common strong belief in rationality was defined and characterized in Battigalli and Siniscalchi (2002) in the context of games with payoff-uncertainty, 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). As shown in Battigalli and Siniscalchi (2003), in two-players signaling games common strong belief in rationality and in a fixed distribution over messages and types provides a characterization of the set of self-confirming equilibria satisfying the iterated intuitive criterion. Unlike Battigalli and Siniscalchi (2003), in this paper players do not share a common belief over the payoff-types of the informed players. Another important difference between the current paper and the previous literature is in the information structure considered. In all of these papers, players' private information is described by a partition with a product structure. In particular, this implies that the characterization theorem in Battigalli and Siniscalchi (2007) cannot be applied directly to the current setup.

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 Cho and Kreps (1987), Govindan and Wilson (2009), Kohlberg and Mertens (1986), Man (2012), Pearce (1984), Reny (1992) and Van Damme (1989).

2 Two-Sided Matching Markets

I consider a two-sided matching environment with transferable utility, following Crawford and Knoer (1981) and LMPS. 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*. The sets I and J could equally represent consultants and companies, doctors and patients or scientists and universities. Each worker is endowed with an *attribute*, or *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. Denote by $\mathbf{w} \in W^I$ and $\mathbf{f} \in F^J$ the corresponding vectors (or *profiles*) of attributes.

A *matching function* is a map $\mu : I \to J \cup \{\emptyset\}$ that is injective on $\mu^{-1}(J)$. The function μ is interpreted as follows: 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. As formalized by the assumption that μ is injective on $\mu^{-1}(J)$, a worker is assigned to at most one firm and a firm can hire at most one worker.

A match between worker of type w and firm of type f gives rise, in the absence of monetary transfers, to a payoff of $\nu(w, f)$ for the worker and of $\phi(w, f)$ for the firm. Following LMPS, the functions ν and ϕ are referred to as *premuneration values*. The premuneration value of an unmatched worker or firm is equal to 0. 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 payment $\mathbf{p}_{i,\mu(i)} \in \mathbb{R}$ from firm $\mu(i)$ to worker i. Payments can be negative. Unmatched workers receive no payments. I use the notation $\mathbf{p}_{i,\emptyset} = \mathbf{p}_{\emptyset,j}$ for every i and j. Workers and firms have quasilinear preferences with respect to payments. So, 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. Notice that, using the notation described above, both expressions are equal to 0 when i and j are unmatched.

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 corresponding payment scheme. A matching outcome is *individually rational* if it gives a positive 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 is implemented. When the profiles of payoff-types $\mathbf{w}$ and $\mathbf{f}$ are common knowledge, 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

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

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 is well known, for fixed $\mathbf{w}$ and $\mathbf{f}$ a complete-information stable outcome always exists, and every stable outcome is efficient.

3 Incomplete Information

This standard framework is now altered by relaxing the assumption of complete information. I consider environments where agents are uncertain about each others' attributes, but where the uncertainty is not as severe to the point of precluding agents from making meaningful inference. In particular, I consider markets with one-sided, interim, incomplete information. The status quo allocation $(\mu, \mathbf{p})$ remains common knowledge, but workers have now an informational advantage over firms.

More specifically, in these class of markets the profile $\mathbf{f}$ of firms' payoff-types is common knowledge, while the vector $\mathbf{w}$ of workers' types is only known to belong to a subset $\mathbf{W} \subseteq W^I$. Single agents have the following information: Each firm j is informed of the the quality of the worker $\mu^{-1}(j)$ it is matched to under the default allocation; Conversely, each worker i knows both his own type $\mathbf{w}_i$ and the type of other workers. It should be emphasized that firms do not share a common belief over the set $\mathbf{W}$. In fact, firms will typically disagree about the quality of different workers.

The following examples illustrate these assumptions in specific environments:

- A new cohort of MBA students is graduating. In the job market, students from the same program are matched to firms. A student's type represents his newly acquired level of productivity. A firm's type represents both its productivity and the know-how that is gained by working for that firm. Information about firms is readily available and students, after having spent years together, have a good understanding of each other's productivity. Both facts are clear to the potential employers.

- A group of well established sellers (I) is matched to a group of new buyers (J). Each seller sells one indivisible unit of an experience good (i.e., a good whose quality is only known once the good is consumed). The true quality of each good is easily discernable by all sellers.

- A group of highly specialized workers (scientists, surgeons, consultants, etc.), is matched to a set of employers (deans, hospitals, public authorities, etc.). A worker's type summarizes his overall quality. While employers may have access to public information regarding each worker, the degree of specialization involved makes it difficult for non-specialist to have a good understanding of their skills.

Some of the assumptions require further comments. The premise that each firm j

knows the quality of worker $\mu^{-1}(j)$ can be interpreted in at least two ways. It can be interpreted, literally, as saying that worker i's type is disclosed to firm $\mu(i)$ at the time the allocation is presented but before it is actually implemented. Alternatively, it can be thought as the result of the same default allocation being implemented repeateadly over time. Under this different interpretation of the model, agents negotiate over new partnerships after the allocation has been implemented at least once, at a stage where each firm has learned the payoff-type of the worker it is matched to.

As discussed in the introduction, the assumption that workers know each other's quality emphasizes workers' informational advantage, and does not appear in the cooperative model of LMPS. As will become clear, this assumption will draw a clear demarcation between the two approaches.

4 The Blocking Game

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

4.1 Model

The set of players is given by $I \cup J$, where $|I| \geq 2$. An allocation $(\mu, \mathbf{p})$ is given, together with the profile $\mathbf{f}$ of firms' payoff-types. In this game of incomplete information the set of payoff-relevant states is given by the set $\mathbf{W}$ of workers' payoff-types. The objects $(\mu, \mathbf{p})$, $\mathbf{f}$ and $\mathbf{W}$ are parameters of the game which are common knowledge among all players. While conditions on players' beliefs will be introduced in the next sections, the informational assumptions introduced in section 3 are assumed to hold.

The game is played in two stages. In each stage, actions are played simultaneously.

- In the first stage each worker i decides between two actions: to *abstain* ("a") or to make an *offer* $(j, q) \in (J - \{\mu(i)\}) \times Q$, where $Q \subseteq \mathbb{R}$ is finite.

Offers are observed by all players. Informally, an offer (j, q) means that worker i is willing to break the status quo and form a new partnership with firm j for a wage q. The finiteness of Q implies that offers are made using a discrete currency.

- In the second stage each firm that 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. They receive payoffs

$$\nu(\mathbf{w}_i, \mathbf{f}_j) + q \text{ and } \phi(\mathbf{w}_i, \mathbf{f}_j) - q,$$

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.

The remaining agents are matched according to the original allocation $(\mu, \mathbf{p})$ and obtain the corresponding payoffs

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

In addition, offers are costly. A worker who does not abstain pays a monetary cost $c > 0$. This cost is subtracted from the payoffs described above.

The structure of the game is common knowledge.

4.2 Discussion

The game has two features that play an important role in the analysis. The first is that 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 the interpretation of the core under complete information (see, for instance, the discussion in Myerson (1991) about the assumptions implicit in the interpretation of the core). It is assumed, in the description of the game, that players originally matched to agents involved in a blocking offer become unmatched (unless part of a blocking offer themselves). This assumption is not crucial. The reason is that, for the purpose of this paper, the fact that a blocking offer has formed is enough to conclude that the matching is not stable.

The presence of a cost $c > 0$ rules out the possibility for an offer to be a best response to the belief that it will be rejected. In turn, this simplifies the inference that firms can make about workers' types when receiving an offer.

4.2.1 Assumptions

Three assumptions are made about the parameters of the game: the allocation $(\mu, \mathbf{p})$ is individually rational, the set Q is a fine grid and the cost c is small. All three assumptions could be easily relaxed, but at the cost of complicating the exposition.

Formally, I assume that $(\mathbf{f}, \mu, \mathbf{p})$ is such that for each profile $\mathbf{w} \in \mathbf{W}$ the resulting matching outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is individually rational. Denote by Λ^{ir} the set of all matching outcomes $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ such that $(\mu, \mathbf{f}, \mathbf{p})$ satisfies this property. This condition can be replaced by the assumption that each player can unilaterally leave the match and obtain a payoff of 0, regardless of other players' actions. The main result can be easily adapted to this more general case.

Given $(\mu, \mathbf{p}, \mathbf{f})$, the set Q is taken to be a fine grid parametrized by $\varepsilon > 0$ and denoted by Q_ε. To this end, fix a number $q^* \in \mathbb{R}_+$ large enough such that for every $j \in J$ and every $w \in W$ we have $\phi(w, \mathbf{f}_j) - q^* < \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. Similarly, let $q_* \in \mathbb{R}_-$ be low enough such for every $i \in I$ and every $f \in F$ we have $\nu(\mathbf{w}_i, f) + q^* < \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$. For each $\varepsilon > 0$, define $Q_\varepsilon \subseteq \mathbb{R}$ to be a finite set such that any subinterval of $[q_*, q^*]$ of length ε intersects Q_ε.

Given $(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$, the cost $c > 0$ is assumed to be small enough that for every i, for every $w, w' \in W$, $f \in F$ and $q \in Q_\varepsilon$ it satisfies $\nu(w, f) + q > \nu\left(w, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ if and only if $\nu(w', f) + q - c > \nu\left(w', \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ (because of the finiteness of W and Q_ε, c is well defined). So, the cost does not affect any relevant strict inequality.

4.2.2 Notation

Given the parameters $\mu, \mathbf{p}, \mathbf{f}$ and ε, denote the game by $\Gamma(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$. I will refer to $\Gamma(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$ as the *blocking game.* The notation used to describe history and strategies is entirely standard. However, 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 players plan their action conditional on every realization of $\mathbf{w}$. So, strategies are not defined as a function of $\mathbf{w}$.

Let H denote the set of all non-terminal histories. The set H is identified with the empty (or *initial*) history $\varnothing$ together with the collection $D \subseteq 2^{I \times J \times Q_\varepsilon}$ of all feasible combinations of offers. In particular, $h = \{(i, j, q)\}$ denotes the history that follows an offer (j, q) from worker i when all other workers have abstained from making offers.

For each worker i, let $H_i = \{\varnothing\}$. For each firm, denote by $H_j \subseteq D$ the subset of histories where j has received at least one offer. A strategy of worker i is a function $s_i : H_i \to \{a\} \cup ((J - \{\mu(i)\}) \times Q_\varepsilon)$. To ease the notation we will write $s_i = (j, q)$ to refer to the strategy s_i defined as $s_i(\varnothing) = (j, q)$. The same applies to the notation $s_i = a$. 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. 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 are consistent with h. If $h = \varnothing$ then $S_{-k}(\varnothing) = S_{-k}$. If $h \in D$, then $S_{-k}(h) = D_{-k}(h) \times S_J$, where $D_{-k}(h)$ is the set of strategies of workers in $I - \{k\}$ that lead to the offers observed at h.

5 Beliefs and Possible Worlds

One possible approach would be to analyze the blocking game by looking at its Nash equilibria and their refinements. While natural, this approach would diverge from the main purpose of this paper, which is to arrive at a definition of stability by explicitly modeling players' beliefs and thought processes. For this purpose, it more convenient to analyze the game epistemically. This section introduces the necessary formalism. It follows Battigalli and Siniscalchi (1999,2002).

5.1 Type Structures

The first step is to formalize conditional beliefs. Let Θ be an abstract metric space of uncertainty (e.g., $\Theta = \mathbf{W}$). A *conditional probability system* (or CPS) for player k is a collection of conditional probabilities[1]

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

with the property that for every history h that is reached by $b_k(\cdot|\varnothing)$ with strictly positive probability, $b_k(\cdot|h)$ is derived from $b_k(\cdot|\varnothing)$ by applying Bayes' rule (recall that $\varnothing$ denotes the initial (empty) history). [2] So, a conditional probability systems specify

[1]Given a metric space X, we denote by $\Delta(X)$ the corresponding set of Borel probability measures.

[2]More formally, given b_k, for every history h such that $S_{-k}(h)$ has positive probability under $b_k(\cdot|\varnothing)$, the probability $b_k(\cdot|h)$ is derived using Bayes' rule by updating $b_k(\cdot|\varnothing)$ on the event $\Theta \times$

players' beliefs at the beginning of the game and conditional on any history, including histories that the player deems impossible at the beginning of the game. Given Θ, denote by $\Delta^H(\Theta \times S_{-k})$ the set of player k's conditional probability systems. [3]

In deciding whether or not to abstain, a worker will take into account how firms will revise their beliefs in response to an offer. Modeling this idea requires not only to consider conditional beliefs but also to consider hierarchies of conditional beliefs. Players' beliefs are modelled using the language of type structures, as defined in Battigalli and Siniscalchi (1999).

Definition 2 A *type-structure* $\mathcal{T} = \left((T_k, \beta_k)_{k \in I \cup J}\right)$ defines for each player k a compact metric space T_k of *epistemic types* and a continous *belief map*

$$\beta_k : T_k \to \Delta^H(\mathbf{W} \times S_{-k} \times T_{-k})$$

The structure $\mathcal{T}$ is *belief-complete* if, in addition, the maps $(\beta_k)_{k \in I \cup J}$ are onto.

Each epistemic type in the set T_k describes a possible mental state of player k. The function β_k maps to each type a CPS $\beta_k(t_k) = (\beta_k(t_k)(\cdot|h))_{h \in H}$, describing player k's beliefs about workers' attributes, opponents' stragies and epistemic types. As in Harsanyi (1967), an epistemic type t_k defines probabilities over the epistemic type of other players. So, each type t_k describes player k's beliefs about $\mathbf{W}$ and S_{-k}; about $\mathbf{W}$ and S_{-k} *and* other players' beliefs about W and S; and so forth. So, each epistemic type encodes a full hierarchy of beliefs. In what follows, a belief-complete type structure $\mathcal{T}$ is fixed. Belief-completeness is a richness assumption that is standard in the literature. [4]

5.2 Essential Ingredients

The definition of type structure will not be applied directly. Instead, the focus will be on simpler objects that are derived from the type structure.

$S_{-k}(h)$.

[3] As shown by Battigalli and Siniscalchi (1999), when Θ is compact, the space $\Delta^H(\Theta \times S_{-k})$ is a closed subset of the compact metric space $\prod_{h \in H} \Delta(\Theta \times S_{-k}(h))$ in the relative topology.

[4] See, for instance, Battigalli and Friedenberg (2013) on the connection between belief-completeness and forward-induction reasoning. As shown in Battigalli and Siniscalchi (1999), a belief-complete type structure can always be constructed by taking the set of types to be the set of all hierarchies of conditional CPSs.

A *state of the world* is a combination $(\mathbf{w}, s, t)$ specifying workers' payoff-types together with players' actions and beliefs. The set of all states of the world is denoted by $\Omega = \mathbf{W} \times S \times T$. So, Ω is the set of all conceivable ways in which players can play and think about the game.

An *event* is a (measurable) subset of Ω. Events of interests will be: the event where all workers abstain; the event where players' strategies are optimal given their beliefs; etc. An important class of events is defined as follows. Given an agent k and an event $E \subseteq \Omega$, denote by E_{-k} the projection of E on $\mathbf{W} \times S_{-k} \times T_{-k}$, and let

$$B_{k,h}(E_{-k}) = \{(\mathbf{w}, s, t) : \beta_k(t_k)(E_{-k}|h) = 1\}$$

be the event where player k *believes* E_{-k} *at history* h. Thus, $B_{k,h}(E_{-k})$ is the collection of all the states of the world where player k attaches probability 1 to the event E_{-k} conditional on having reached history h.

In the present framework players do not formally have beliefs about their own beliefs or strategies. However, to simplify the language, given an event $E \subseteq \Omega$, I will occasionaly refer to $B_{k,h}(E_{-k})$ as the event where player k *believes* E *at history* h.

6 Stability Under Forward Induction

This section returns to the main question of this paper by introducing a new notion of stability. A matching outcome will be said to be stable if, given the corresponding blocking game, it is possible for workers to rationally abstain from making offers under suitable restrictions on players' beliefs.

6.1 Basic Conditions

The first step is to define rationality. Given a player k, a strategy s_k and a pair $(\mathbf{w}, s_{-k})$ in $\mathbf{W} \times S_{-k}$, let $U_k(s_k, s_{-k}, \mathbf{w})$ denote the corresponding payoff for player k. Player k is *rational in the state of the world* $\omega = (\mathbf{w}^*, s^*, t^*)$ if for every history $h \in H_k$ the strategy s_k^* solves[5]

$$\max_{s_k \in S_k} \sum_{(\mathbf{w}, s_{-k})} \left(\mathrm{marg}_{\mathbf{W} \times S_{-k}} \beta_k(t_k^*)(\mathbf{w}, s_{-k}|h) \right) U(s_k, s_{-k}, \mathbf{w}).$$

[5]Given a CPS b_k, $\mathrm{marg}_{\mathbf{W} \times S_{-k}} b_k(\cdot|h)$ denotes the marginal of $b_k(\cdot|h)$ on $\mathbf{W} \times S_{-k}$.

So, a player is *rational* if he plays optimally at every history he is asked to act, given his belief. Denote by R the set of states of the world where all players are rational.

Another basic assumption is for players not to expect others to make offers. The event where all workers abstain is denoted by $A = \{(\mathbf{w}, s, t) : \; s_i = a \text{ for all } i \in I\}$. Let

$$B_{\varnothing}(A) = \bigcap\nolimits_k B_{k,\varnothing}(A_{-k})$$

be the event where each player k expects, at the beginning of the game, that all workers (other possibly than k) will abstain from making offers.

Players' beliefs are required to be consistent with their information. To ease the notation, given a realized $\mathbf{w} \in \mathbf{W}$ define $[\mathbf{w}]_i = \{\mathbf{w}\} \times S_{-i} \times T_{-i}$ for every worker i and $[\mathbf{w}]_j = \left\{\tilde{\mathbf{w}} : \tilde{\mathbf{w}}_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}\right\} \times S_{-j} \times T_{-j}$ for each firm j. Given a player k, the event $[\mathbf{w}]_k$ represents the information k holds at the beginning of the game. As described in section 3, workers' information corresponds to the whole vector $\mathbf{w}$. For each firm, its information is given by the payoff-type of the worker it is matched to under the default allocation. To connect players' beliefs and information, define the event

$$C = \left\{(\mathbf{w}, s, t) : (\mathbf{w}, s, t) \in \bigcap\nolimits_{k \in I \cup J} \bigcap\nolimits_h B_{k,h}\left([\mathbf{w}]_k\right)\right\}.$$

The event C is the collection of state the world where, at every history, workers have correct beliefs about each other's payoff-type and firms have correct beliefs about the quality of the worker they are matched to under the status-quo.

A state of the world $(\mathbf{w}, s, t)$ satisfying

$$(\mathbf{w}, s, t) \in A \cap R \cap C \cap B_{\varnothing}(A) \tag{1}$$

describes a situation where workers rationally abstain from making offers and all players correctly expect the status-quo allocation to be implemented. This idea is in the spirit of self-confirming equilibrium (i.e., as in Fudenberg and Levine (1993)). These basic conditions are now extended by imposing further restrictions on players' rationality and on how beliefs are revised conditional on (unexpected) offers.

6.2 Strong Belief

Forward-induction reasoning is described through the notions of strong belief and common strong belief, introduced in Battigalli and Siniscalchi (2002). The next definition adapts their definition to the current setting.

Definition 3 Player k *strongly believes* an event $E \subseteq \Omega_{-k}$ in the state $(\mathbf{w}, s, t)$ if $[\mathbf{w}]_k \cap E \neq \emptyset$ and $(\mathbf{w}, s, t) \in B_{k,h}(E)$ for every history h such that

$$(\mathbf{W} \times S_{-k}(h) \times T_{-k}) \cap [\mathbf{w}]_k \cap E \neq \emptyset. \tag{2}$$

Given an event $E \subseteq \Omega_{-k}$, denote by $SB_k(E)$ the set of states of the world where k strongly believes E. To illustrate the definition, consider the case where k is a firm (an analogous interpretation applies in when k is a worker). When the history h is reached, the evidence available to k consists of the attribute of worker $\mu^{-1}(k)$, described by $[\mathbf{w}]_k$, and the set of strategies that could have led to h, described by $S_{-k}(h)$. Expression (2) holds whenever this evidence does not logically contradict E. Under the assumption of strong belief, at any such history player k is *required* to believe E. So, player k strongly believes the event E if he believes it at the beginning of the game and continues to believe it as long as the event is not contradicted by the evidence. Given an event $E \subseteq \Omega$, denote by $SB(E) = \bigcap_{k \in I \cup J} SB_k(E_{-k})$ the event where each player strongly believes E.

Notice, in definition 3, the requirement $[\mathbf{w}]_k \cap E \neq \emptyset$. This restriction implies that for an event to be strongly believed, the event must be believed at the beginning of the game. That is, it implies $SB_k(E) \subseteq B_{k.\varnothing}(E)$ for every player k and every event E.

6.3 Stability

The notion of strong belief allows to formalize the idea that players will interpret offers according to the highest possible degree of sophistication that can be attached to their proponent. Let $Z^1 = SB(R \cap B_{\varnothing}(A) \cap C)$. Proceeding inductively, define for every $n > 1$ the event

$$Z^n = Z^{n-1} \cap SB\left(Z^{n-1}\right).$$

The sequence $\{Z^1, Z^2, Z^3, ...\}$ is decreasing and imposes increasingly stringent requirements on players' beliefs.

The event Z^1 describe a situation where the basic conditions $R \cap B_{\varnothing}(A) \cap C$ are strongly believed by all players. Strong belief in $R \cap B_{\varnothing}(A) \cap C$ implies that, whenever possible, players interpret an offer as being rational and unexpected by all players (other than the proponent).[6] The event Z^2 describes a further degree of sophistication, in which, in addition, players strongly believe the event $R \cap B_{\varnothing}(A) \cap C \cap SB(R \cap B_{\varnothing}(A) \cap C)$. So, Z^2 describes a situation where, in response to an offer, whenever possible players hold the belief that others are rational, did not expect the offer to be made, and, in turn, interpret the offer as rational and unexpected by their opponents. The limit of this iteration leads to the event

$$SB^{\infty}(R \cap B_{\varnothing}(A) \cap C) = \bigcap\nolimits_{n=1}^{\infty} Z^n \tag{3}$$

which, following the terminology in Battigalli and Siniscalchi (2002), defines *common strong belief in* $R \cap B_{\varnothing}(A) \cap C$.

Under common strong belief in the basic condition $R \cap B_{\varnothing}(A) \cap C$, the hypothesis $\bigcap_{n=1}^{\infty} Z^n$ is believed by all players at the beginning of the game. Conditional on history h being reached, if h is inconsistent with the event $\bigcap_{n=1}^{\infty} Z^n$ then each player k revises his beliefs by assigning probability one to $Z_{-k}^{\bar{n}}$, where $\bar{n}$ is the largest n such that Z_{-k}^{n} is compatible with h.[7] So, players interpret unexpected moves according to the strongest assumption among $\{Z^1, Z^2, Z^3, ...\}$ that is not contradicted by the evidence. This idea leads to the following definition.

Definition 4 A matching outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is *stable under forward induction* if, for every $\varepsilon > 0$, given the blocking game $\Gamma(\mathbf{f}, \mu, \mathbf{p}, \varepsilon)$, there is a state of the world $\omega = (\mathbf{w}, s, t)$ such that

$$\omega \in A \cap R \cap B_{\varnothing}(A) \cap C \cap SB^{\infty}(R \cap B_{\varnothing}(A) \cap C). \tag{4}$$

For a matching outcome to be stable it must be possible for players to rationally abstain from making offers when their beliefs are described by common strong belief in rationality and in the belief that no offers will be made. The next sections illustrate the logic behind this stability under forward induction and its implications.

[6] From now on I will omit to mention the event C in the informal discussion.

[7] Formally, the highest n such that $(\mathbf{W} \times S_{-k}(h) \times T_{-k}) \cap [\mathbf{w}]_k \cap Z_{-k}^{n} \neq \emptyset$.

7 Characterization

7.1 Incomplete-Information Stability

A notion of stability under incomplete information was recently introduced by LMPS. Its definition takes the form of an iterative elimination procedure.

Definition 5 (LMPS (2014)) Let Λ^0 be the set of individually rational matching outcomes. Inductively, for each $n > 0$ define Λ^n as the set of outcomes $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ such that $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{n-1}$ and there is no $i \in I$, $j \in J$ and $q \in \mathbb{R}$ such that

$$\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)} \tag{5}$$

and

$$\phi(\mathbf{w}'_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j} \tag{6}$$

for all $\mathbf{w}' \in \mathbf{W}$ that satisfy

$$\begin{aligned} (\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) &\in \Lambda^{n-1}, & (7)\\ \mathbf{w}'_{\mu^{-1}(j)} &= \mathbf{w}_{\mu^{-1}(j)}, \; \textit{and} & (8)\\ \nu(\mathbf{w}'_i, \mathbf{f}_j) + q &> \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}. & (9) \end{aligned}$$

The set of *incomplete-information stable* matching outcomes is $\Lambda^\infty = \bigcap_{n=1}^{\infty} \Lambda^n$.

To illustrate the definition, consider first the case where $n = 1$. An outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is eliminated in the first iteration if it is possible to find a worker i and a firm j who can form a 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 the firm can consider "plausible". As required by (7)-(9), j restricts the attention to type profiles $\mathbf{w}'$ that satisfy individual rationality, 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 firms consider plausible. Requirement (7) says that in the second iteration, when evaluating a potential block, firms only consider payoff-types such that the current matching outcome belongs to Λ^1. Proceeding inductively, in the n-th step of the procedure firms maintain the assumption that the matching outcome has survived $n-1$ of the elimination process.

The procedure is motivated in LMPS by the idea that a stable matching persists over time. To be stable, a matching must present no blocking opportunities. In addition, it must remain immune to blocking when, over time, agents draw inference about workers' types from the observation that no block has occurred.

This definition of stability satisfies two surprising property: existence and efficiency under standard supermodularity assumptions.

Lemma 1 (LMPS) *Every complete-information stable matching outcome is incomplete-information stable. In particular, for every* $\mathbf{w}$ *and* $\mathbf{f}$*, there exists an allocation* $(\mu, \mathbf{p})$ *such that* $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ *is incomplete-information stable.*

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

Theorem 1 (LMPS) *Let* $W \subset \mathbb{R}$ *and* $F \subset \mathbb{R}$*. Assume that* ν *and* ϕ *are strictly increasing and strictly supermodular. Then, every incomplete information stable matching outcome is efficient.*[8]

The assumption of supermodularity implies that every efficient outcome must be positively assortative.

7.2 Characterization Theorem

The next theorem, which is the main result of the paper, characterize the set of matching outcomes that are stable under forward induction.

Theorem 2 *A matching outcome* $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{ir}$ *is stable under forward induction if and only if it is incomplete information stable.*

The result shows that stability under forward induction, despite its elaborate definition, can be described using a simple iterative procedure. In addition, it establishes, together with Theorem 1, that forward-induction reasoning leads to efficiency under supermodular premuneration values. From a different perspective, Theorem 2 provides epistemic foundations for incomplete-information stability, which can be interpreted as the outcome of non-cooperative negotiation under the assumption that players revise their beliefs according to the logic of forward-induction reasoning. The next examples illustrate the intuition behind the result.

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

7.3 Examples

Consider the market described in Figure 1. Adam and Ann are matched at a wage -4. Ann's payoff-type is 4, while Adam's type is 2. Bob is unmatched and his payoff-type can be either 3 or 0. Premuneration values are given by $\phi(w, f) = \nu(w, f) = wf$ for all pairs of payoff-types.[9]

	Adam	Bob
worker payoffs:	4	0
worker types, $\mathbf{w}$:	2	3(0)
payments, $\mathbf{p}$:	-4	
firm types, $\mathbf{f}$:	4	
firm payoff:	12	
	Ann	

Figure 1

Assume Bob's actual type is 3. Then, the matching outcome is not stable under incomplete information. To see this, consider the combination given by Bob, Ann and the payment $q = -1$. This combination is such that the matching outcome is eliminated in the first iteration of Definition 5.

While it follows from Theorem 2 that the outcome is not stable under forward induction, it might be helpful to see the underlying intuition at a more informal level. Suppose, by way of contradiction, that ω is a state of the world that satisfies (4) and in which Bob's type is 3. As we now show, in the state ω Bob cannot be rational.

Consider the offer from Bob to Ann at a payment $q = -1$. In the state ω, Ann strongly believes the event $R \cap B_{\varnothing}(A) \cap C$. So, conditional on the offer being made, Ann believes that Bob is rational unless the event $R \cap B_{\varnothing}(A) \cap C$ is contradicted by the offer. To see that the event is not contradicted, we need to verify that there exists at least one state of the world ω' (belonging to C) where making such an offer is a rational strategy and Ann believes, at the beginning of the game, that Bob will abstain. For instance, a state $\omega' \in C$ such that Bob's type is 3, he expects the offer under consideration to be accepted, believes that any more profitable offer would be rejected, and such that, in addition, Alice expects no offer to be made. The existence of such a state follows from

[9] In the examples in this subsection, premuneration values are not strictly supermodular. This is only for simplicity. Similar arguments apply to examples where strict supermodularity is satisfied.

the assumption that the underlying type structure $\mathcal{T}$ is belief-complete.

So, conditional on the offer being made, Ann *will* believe that Bob is rational. Because offers are costly, the offer can only be a rational strategy if his type is not 0. Thus, Ann must conclude that Bob's type is 3. Hence, expecting a payoff of 13 and being rational, she will accept the offer. Because Bob believes at the beginning of the game that Ann strongly believes $R \cap B_{\varnothing}(A) \cap C$, he anticipates that such an offer will be accepted. Hence, it cannot be a rational choice for Bob to abstain from making an offer to Ann. So, the outcome is not stable under forward induction.

The next example, presented in Figure 2, illustrates one more step in the players' thought processes.

	Adam	Bob	
worker payoffs:	4	1(0)	
worker types, **w**:	2	$3(3^*)$	
payments, **p**:	-4	0	
firm types, **f**:	4	f	f^*
firm payoffs:	12	0	0
	Ann	Alice	Charlie

Figure 2

In this example, Bob is matched to Alice and his type can be either 3 or 3^*. It will be convenient to think of 3 as the "good" type and of 3^* as the "bad" type. An unmatched agent, named Charlie, is of type f^*.

Premuneration values for matches that do not appear in Figure 2 are defined as follows. Adam has no incentives to leave the default match. Formally, $\nu(2, f) = \phi(2, f) = 0$ and $\nu(2, f^*) = \phi(2, f^*) = 0$. In the case where Bob is of the good type, a match between Bob and Ann produces a premuneration value of 12 for both parties, i.e., $\nu(3, 4) = \phi(3, 4) = 12$. Conversely, a match between Bob and Charlie produces 0 surplus, i.e., $\nu(3, f^*) = \phi(3, f^*) = 0$. In case where Bob is of the bad type, he has an incentive to deviate from the existing allocation by matching with either Charlie or Ann. A match between Bob and Charlie produces strictly positive utility for both agents. In particular, $\nu(3^*, f^*) = \phi(3^*, f^*) = 1$. A match between Bob and Ann gives utility $\nu(3^*, 4) = 12$ to Bob but $\phi(3^*, 4) = 0$ to Ann.

Assume Bob's actual type is 3. We now show that the resulting outcome is not incomplete-information stable. The first iteration of Definition 5 eliminates the alternative outcome where Bob's type is 3^*. To see this, consider the combination given by Bob, Charlie and the payment $q = -\frac{1}{2}$. Because $\nu(3^*, f^*) = 1$, this match is profitable for Bob. Because $q < 0$, the match is also profitable for Charlie regardless of Bob's type. This combination show that this alternative outcome does not belong to Λ^1. Now let's return to the current outcome in which Bob is of the good type. It can be verified that this outcome survives the first iteration. To see that it is eliminated in the second iteration, consider the combination given by Bob, Ann and the payment $q = -1$. Because the only outcome belonging to Λ^1 is the actual outcome, the combination increases the payoff of both Bob and Ann and satisfies (7)-(9).

I now provide an intuition for why the outcome is not stable under forward induction. First notice that, being of type 3, Bob's only incentive to deviate from the status quo is by matching with Ann. For Bob to propose a profitable offer to Ann, the offer must involve a payment q such that $12 + q > 1$, i.e., $q > -11$. Any such offer is such that both the good type and the bad type would benefit from it. However, in case Bob's type was 3^*, Ann would obtain a payoff of $-q < 12$. So, for Ann to accept offer she must rule out the possibility that Bob is of the bad type.

Consider an offer from Bob to Ann at a wage $q = -1$. Ann interprets the offer according to the "best" possible explanation that is consistent with players' rationality and the belief that this offer was unexpected by everyone other than Bob. In particular, and this is the key aspect, Ann takes into account that the offer is unexpected to Alice and that Alice knows Bob's type.

So Ann must ask herself: What is the "best" possible explanation that, ex-ante, could have justified Alice's belief that Bob was going to abstain? The explanation depends on Bob's type. If Bob's type is 3^*, then Alice must have thought that Bob believed that Charlie is irrational. If not, then Bob would have made the offer $q = -\frac{1}{2}$ to Charlie who, being rational, would have accepted. More generally, any explanation that does not rule out the bad type involves Alice believing that Bob expected Charlie to act irrationally. If instead Bob's type is the good type, then instead Alice could have thought, for instance, that Bob was going to abstain, expecting Ann to rationally reject any profitable offer under the incorrect belief that Bob's type was the bad type. Because Ann opts for the "best" possible explanation, she must rule out the bad type. Hence, she accepts the offer. Anticipating this, Bob cannot rationally abstain from making an offer to Ann. Thus, the matching is not stable under forward induction.

8 Extensions and Discussion

8.1 Offers and Rejection

Stability under forward induction is defined by considering states of the world where workers abstain from making offers. It is possible, however, that in some of these states of the world firms' strategies are such that profitable offers, if made, would be actually accepted. This leads to the question of whether stability allows workers to refrain from starting a negotiation because of excessively pessimistic beliefs. The next result shows that requiring offers to be rejected does not restrict the set of stable matching outcomes.

Theorem 3 *A matching outcome* $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ *is stable under forward induction if and only if there is a state of the world* $\omega = (\mathbf{w}, s, t)$ *that satisfies*

$$\omega \in A \cap R \cap B_{\varnothing}(A) \cap C \cap SB^{\infty}(R \cap B_{\varnothing}(A) \cap C)$$

and such that for every worker i *and every offer* (j, q)*, if* $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu(\mathbf{w}_i, \mathbf{f}_{\mu(i)}) + \mathbf{p}_{i,\mu(i)}$ *then the offer is rejected by firm* j *(i.e.,* $s_j(\{i, j, q\}) = r$*).*

If the matching outcome is stable it is possible to find a state of the world that satisfies the same epistemic conditions of Definition 4 and where, in addition, any offer that would improve a worker's payoff above the status-quo allocation is rejected.

8.2 Strict Stability

Stability asks whether it is rational for workers to abstain, under suitable epistemic conditions. A more stringent definition would require abstention to be, in addition, the *only* action that is rational for workers. To illustrate this approach, let $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ be an outcome that is stable under forward induction. The outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is said to be *strictly stable* if for every $\varepsilon > 0$ and every state of the world $\omega = (\mathbf{w}, s, t)$,

$$\text{if } \omega \in R \cap B_{\varnothing}(A) \cap C \cap SB^{\infty}(R \cap B_{\varnothing}(A) \cap C) \text{ then } \omega \in A.$$

So, an outcome is strictly stable if workers abstain from making offers in every state of the world where players are rational, given the epistemic conditions defining stability under forward induction. The next result provides a characterization of strict stability in terms of incomplete-information stability.

Theorem 4 *Given an incomplete-information stable matching outcome* $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$*, the following are equivalent:*

1. $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ *is strictly stable.*

2. *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 profile $\mathbf{w}' \in \mathbf{W}$ *such that*
$$\begin{aligned} (\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) &\in \Lambda^\infty, \\ \mathbf{w}'_{\mu^{-1}(j)} &= \mathbf{w}_{\mu^{-1}(j)}, \\ \nu(\mathbf{w}'_i, \mathbf{f}_j) + q &> \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}. \end{aligned}$$

An outcome that is strictly stable must be complete-information stable. This follows from the fact that any complete-information stable outcome belongs to Λ^∞. However, unlike complete-information stability, outcomes that are strictly stable under forward induction may not exist for a fixed profile of types. To see this, consider the following example:

	Adam	Bob
worker payoffs:	$1 + p^*$	$1 + p^*$ $(3 + p^*)$
worker types, $\mathbf{w}$:	1	1(3)
payments, $\mathbf{p}$:	p^*	p^*
firm types, $\mathbf{f}$:	1	1
firm payoffs:	$1 - p^*$	$1 - p^*$
	Ann	Alice

Figure 3

Premuneration values are given by the product of the corresponding pairs of types. Bob can be of type 1 or 3. His actual type is 1.

Denote by $\mathbf{w}^1$ and $\mathbf{w}^3$ the type profiles such that Bob's type is 1 and 3, respectively. It is immediate to check that, given the type profile $\mathbf{w}^1$, there is a unique complete-information stable matching outcome, up to a relabeling of workers. In this outcome, denoted by $(\mathbf{w}^1, \mathbf{f}, \mu, \mathbf{p}^*)$, both workers are matched and at the same wage p^*. Given $\mathbf{w}^1$, this is the only candidate for a strictly stable outcome. It can be shown that the outcome remains incomplete-information stable when changing Bob's type from 1 to 3. That is, $(\mathbf{w}^3, \mathbf{f}, \mu, \mathbf{p}^*) \in \Lambda^\infty$. Suppose, for concreteness, that Bob is matched with Alice, as in the Figure 3. Consider the blocking pair formed by Bob and Ann at a payment q, where $q \in (p^*, 2 + p^*)$. We now show that this combination violates condition (2) of Theorem 4. Because $q > p^*$, it leads to an increase in Bob's payoff from $3 + p^*$ to $3 + q$. In addition, because $q < 2 + p^*$ then it also lead to an increase in Ann's payoff from $1 - p^*$ to $3 - q$ in the case where Bob's type is 3. Hence, $(\mathbf{w}^1, \mathbf{f}, \mu, \mathbf{p}^*)$ is not strictly stable.

8.3 Conclusion

This paper proposes a new notion of stability for markets with one-sided uncertainty. Stability is formulated in a non-cooperative and 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.

9 Appendix

9.1 Preliminaries

Throughout the Appendix, definitions and statements will refer to the blocking game $\Gamma(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$, given the parameters $(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$. It is necessary to introduce some additional notation. Let $\Sigma = \mathbf{W} \times S$. For every subset $\Psi \subseteq \Sigma$ and every player $k \in I \cup J$, let Ψ_k and Ψ_{-k} be the projections of Ψ on $\mathbf{W} \times S_k$ and $\mathbf{W} \times S_{-k}$, respectively. Given $\mathbf{w}$, define $\langle \mathbf{w} \rangle_i = \{\mathbf{w}\} \times S_{-i}$ for every worker i, and $\langle \mathbf{w} \rangle_j = \left\{ \mathbf{w}' : \mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \right\} \times S_{-j}$ for every firm j.

Given a CPS $b_k \in \Delta(\Theta \times S_{-k})$ for some space of uncertainty Θ, it is convenient to use the notation $b_{k,h}$ to denote the probability $b_k(\cdot|h)$. For every $k \in I$ and CPS $b_k \in \Delta(\mathbf{W} \times S_{-k})$, denote by $r_k(b_k) \subseteq S_k$ the set of strategies that are sequential best replies to b_k. That is, each strategy $s_k^* \in r_k(b_k)$ solves

$$\max_{s_k \in S_k} \sum_{(\mathbf{w}, s_{-k})} (b_{k,h}(\mathbf{w}, s_{-k})) U(s_k, s_{-k}, \mathbf{w})$$

for every $h \in H_k$.

In what follows I characterize the pairs of payoff-types and strategy profiles that are consistent with the epistemic conditions

$$\begin{aligned} R^1 &= R \cap B_{\varnothing}(A) \cap C, \\ R^n &= R^{n-1} \cap SB\left(R^{n-1}\right) \quad \text{for every } n \geq 2, \text{ and} \\ R^{\infty} &= \bigcap\nolimits_{n=1}^{\infty} R^n. \end{aligned}$$

In particular, it follows that $R^{\infty} = R \cap B_{\varnothing}(A) \cap C \cap SB^{\infty}(R \cap B_{\varnothing}(A) \cap C)$.

As a first step, we show that each event R^n is closed (hence measurable). The next result is standard.

Lemma 2 *For each n, the set R^n is closed.*

Proof. Consider first the case where $n = 1$. Given a player k, a type $t_k \in T_k$ let $f_k(t_k) \in \Delta^H(\mathbf{W} \times S_{-k})$ be the CPS defined as $f_{k,h}(t_k) = \text{marg}_{\mathbf{W} \times S_{-k}} \beta_{k,h}(t_k)$ for every h. It is well known that $f_k(t_k)$ is a well defined CPS and that the function f_k is continuous (see, e.g., Battigalli and Siniscalchi (2007)) It is routine to verify that this implies that R^1 is closed. It follows from standard arguments that for each player k, each closed set $F \subseteq \Omega_{-k}$ and each history h, the event $B_{k,h}(F)$ is closed. This implies that $B_{\varnothing}(A)$ is closed. Finally, notice that we can write C as

$$C = \bigcup_{\mathbf{w} \in \mathbf{W}} \left((\{\mathbf{w}\} \times S \times T) \cap \bigcap_{k \in I \cup J} \bigcap_{h \in H} B_{k,h}([\mathbf{w}]_k) \right)$$

since each $[\mathbf{w}]_k$ is closed and $\mathbf{W}$ is finite, it follows that C is closed.

The proof is concluded by verifying that for each closed set $F \subseteq \Omega$, the event $SB(F)$ is closed. Let $\mathbf{W}_{k,F}$ be the set of profiles $\mathbf{w}$ such that $[\mathbf{w}]_k \cap F \neq \emptyset$. We have

$$SB_k(F_{-k}) = \bigcup_{\mathbf{w} \in \mathbf{W}_{k,F}} \left((\{\mathbf{w}\} \times S \times T) \cap \bigcap_{h:(\mathbf{W} \times S_{-k}(h) \times T_{-k}) \cap [\mathbf{w}]_k \cap F_{-k} \neq \emptyset} B_{k,h}(F_{-k}) \right) \cap C$$

Because each projection F_{-k} is closed and $\mathbf{W}_{k,F}$ it follows that $SB_k(F_{-k})$ is closed. Hence $SB(F)$ is closed as well. ■

9.2 Rationalizability

Definition 6 *Consider the following procedure. Let $\Sigma^0 = \Sigma$. Inductively, for every $n > 0$ define Σ^n to be set of pairs $(\mathbf{w}, s) \in \Sigma$ such that for each player k there exists a CPS $b_k \in \Delta^H(\Sigma_{-k})$ such that*

(P1-n) $s_k \in r_k(b_k)$,

(P2-n) $b_{k,\varnothing}\left(\Sigma_{-k}^{n-1}\right) = 1$,

(P3-n) $b_{k,\varnothing}\left(\left\{\left(\mathbf{w}', s'_{-k}\right) : s'_i = a \text{ for all } i \in I - \{k\}\right\}\right) = 1$,

(P4-n) $b_{k,h}\left(\langle \mathbf{w} \rangle_k\right) = 1$ *for all* $h \in H$

(P5-n) *for all* $h \in H$ *and all* $m \in \{0, ..., n-1\}$, *if*

$$(\mathbf{W} \times S_{-k}(h)) \cap \langle \mathbf{w} \rangle_k \cap \Sigma_{-k}^m \neq \emptyset \tag{10}$$

then $b_{k,h}\left(\Sigma_{-k}^m\right) = 1$.

Properties (P1-n)-(P5-n) reflect the epistemic conditions introduced in the body of the paper. By (P1-n), strategies are best replies to beliefs. By (P2-n), each player initially believes that other players play strategies compatible with Σ^n_{-k}. According to (P3-n), each player k expects at the beginning of the game that all workers (except k, if $k \in I$) will abstain from making offers. Property (P4-n) says that players' beliefs about $\mathbf{W}$ are consistent with their information. Property (P5-n) plays a crucial role and reflects the strong belief assumptions. Notice that, given (P5-n), property (P2-n) is equivalent to the requirement $\langle \mathbf{w} \rangle_k \cap \Sigma^n_{-k} \neq \emptyset$.

The procedure is, essentially, an instance of Δ-rationalizability, as defined in Battigalli and Siniscalchi (2003), but adapted to the present framework.

We first show that beliefs of workers at histories different other than the initial history play no role.

Because workers only act in the first stage of the game, in the case where $k \in I$ property (P1-n) only involves the strategy s_k and the initial probability $b_{k,\varnothing}$. Hence, given a worker i, the question of whether properties (P1-n)-(P3-n) are satisfied can be applied (with slight abuse of notation) to any strategy $s_i \in S_i$ and probability $\rho_i \in \Delta(\mathbf{W} \times S_{-i})$.

Lemma 3 *Fix $n \geq 0$, $\mathbf{w} \in \mathbf{W}, i \in I$ and a strategy $s_i \in S_i$. Let $\rho_i \in \Delta(\mathbf{W} \times S_{-i})$ be a probability such that $\rho_i(\{\mathbf{w}\} \times S_{-i}) = 1$ 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 (P1-n)-(P5-n).*

Proof. The CPS b_i is 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 (10) holds for $m = n - 1$. For every such history define $b_{i,h}$ to satisfy $b_{i,h}\left((\{\mathbf{w}\} \times S_{-j}(h)) \cap \Sigma^m_{-i}\right) = 1$. Proceeding inductively, decrease m and repeat the argument at every step. Because $\Sigma^0_{-i} = \Sigma_{-i}$, then for every history there exists an m such that (10) 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)-(P5-n). ■

It is immediate to check that the sequence (Σ^n) is decreasing. Let $\Sigma^\infty = \bigcap_{n=1}^\infty \Sigma^n$. We first observe that for a fixed $\mathbf{w}$ each set Σ^n and each event R^n have a product structure. Recall that for every n and player k the set R^n_k denotes the projection of R^n on $\mathbf{W} \times S_k \times T_k$.

Lemma 4 *Let $\omega = (\mathbf{w}, s, t)$ be such that $(\mathbf{w}, s_k, t_k) \in R_k^n$ for every player k. Then $(\mathbf{w}, s, t) \in R^n$.*

Proof. It follows immediately from the definitions that the result is true for $n = 1$. Assume it holds for every $m \in \{0, ..., n-1\}$, where $n > 0$. Let $(\mathbf{w}, s_k, t_k) \in R_k^n \subseteq R_k^{n-1}$ for each k. Then $\omega = (\mathbf{w}, s, t) \in R^{n-1}$ by assumption. We need to show that $\omega \in SB(R^n)$. Every player k satisfies $\beta_{k,\varnothing}(t_k)\left(R_{-k}^{n-1}\right)$. So, $\omega \in \bigcap_k B_{k,\varnothing}(R^{n-1})$. In addition, at any history h, if $(\mathbf{W} \cap S_{-k}(h) \cap T_{-k}) \cap \langle \mathbf{w} \rangle_k \cap R_{-k}^{n-1} \neq \emptyset$ then $\beta_{k,h}(t)\left(R_{-k}^{n-1}\right)$, given that $(\mathbf{w}, s_k, t_k) \in R_k^n$. Since this is true for all players, it follows that $\omega \in R^n$. ■

Lemma 5 *If $(\mathbf{w}, s_k) \in \Sigma_k^n$ for each k then $(\mathbf{w}, s) \in \Sigma^n$. In addition, for every j and every $\mathbf{w}' \in \mathbf{W}$ such that $(\{\mathbf{w}'\} \times S) \cap \Sigma^n \neq \emptyset$ and $\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$, the strategy s_j satisfies $(\mathbf{w}', s_j) \in \Sigma_j^n$.*

Proof. The proof of the first claim parallels the proof of Lemma 4. The proof of the second claim follows from the fact that $\langle \mathbf{w}' \rangle_j = \langle \mathbf{w} \rangle_j$. ■

9.3 Epistemic Characterization

We now prove that the procedure presented in definition 6 characterizes the behavioral implications of the epistemic conditions $R \cap B_\varnothing(A) \cap C \cap SB^\infty(R \cap B_\varnothing(A) \cap C)$. The proof of the next result is adapted from Battigalli and Siniscalchi (2007).

Theorem 5 *The events $(R^n)_{n=1}^\infty$ satisfy*

$$\mathrm{proj}_\Sigma R^n = \Sigma^n \text{ for every } n$$

and $\mathrm{proj}_\Sigma R^\infty = \Sigma^\infty$.

Proof. Given a player k, a type $t_k \in T_k$ let $f_k(t_k) \in \Delta^H(\mathbf{W} \times S_{-k})$ be the CPS defined as

$$f_{k,h}(t_k) = \mathrm{marg}_{\mathbf{W} \times S_{-k}} \beta_{k,h}(t_k)$$

for every h. It is routine to check that $f_k(t_k)$ is a well defined CPS. It follows from the belief-completeness of the type structure that each map f_k is onto.

Let $n = 1$ and fix a state $\omega = (\mathbf{w}, s, t) \in R \cap C \cap B_\varnothing(A)$. Fix a player k and consider the CPS $f_k(t_k)$. Because $\omega \in R$, it follows that s_k is a sequential best reply $f_k(t_k)$. So the pair $(s_k, f_k(t_k))$ satisfies (P1-n). Since $\omega \in C \cap B_\varnothing(A)$ and $\Sigma^0 = \Sigma$, $f_k(t_k)$ satisfies $f_{k,h}(t_k)(\langle \mathbf{w} \rangle_k) = \beta_{k,h}(t_k)([\mathbf{w}]_k)$ for every h and

$$f_{k,h}(t_k)\left(\left\{\left(\mathbf{w}', s'_{-k}\right) : s'_i = a \text{ for all } i \in I - \{k\}\right\}\right) = 1.$$

Since $\Sigma^0 = \Sigma$, property (P5-n) holds vacuously. By repeating this argument for each player we conclude that $(\mathbf{w}, s) \in \Sigma^1$.

Now fix a profile $\mathbf{w}$, a player k and a strategy s_k such that $(\mathbf{w}, s_k) \in \Sigma^1_k$. Let b_k be a CPS such that the pair (s_k, b_k) satisfies (P1-n)-(P5-n). By the belief completeness of the type structure $\mathcal{T}$ we can find for each player k a type $\tau^1_k(\mathbf{w}, s_k) \in T_k$ such that $f_k(\tau^1_k(\mathbf{w}, s_k)) = b_k$. Now consider the resulting state of the world

$$\omega = \left(\mathbf{w}, (s_k)_{k \in I \cup J}, \left(\tau^1_k(\mathbf{w}, s_k)\right)_{k \in I \cup J}\right)$$

It is immediate to verify that the equality $f_k(\tau^1_k(\mathbf{w}, s_k)) = b_k$ implies $\omega \in R$ (because of (P1-n)), $\omega \in C$ (because of (P4-n)) and $\omega \in B_\varnothing(A)$ (because of (P3-n)). So, $\omega \in R^1$ and $(\mathbf{w}, s_k, \tau^1_k(\mathbf{w}, s_k)) \in R^1_k$ for each $k \in I \cup J$.

The proof now proceeds inductively. Assume there exists $n > 0$ such that for every $m \in \{1, ..., n-1\}$ we have (i) $R^m_k = \Sigma^m_k$ for every $k \in I \cup J$ and (ii) there exist maps $\tau^m_k : \mathbf{W} \times S_k \to T_k$, $k \in I \cup J$ such that

$$(\mathbf{w}, s_k, \tau^m_k(\mathbf{w}, s_k)) \in R^m_k$$

for each k and each $(\mathbf{w}, s_k) \in \Sigma^m_k$.

We first notice that for every k and $m \in \{1, ..., n-1\}$, we have $\Sigma^m_{-k} = \mathrm{proj}_{\Sigma_{-k}} R^m_{-k}$. By the inductive hypothesis, the result is true if either $\Sigma^m = \emptyset$ or $R^m = \emptyset$. So assume both sets are non-empty. Let $(\mathbf{w}, s_{-k}) \in \Sigma^m_{-k}$. For every $k \neq l$, we have $(\mathbf{w}, s_l) \in \Sigma^m_l$. Choose a strategy s_k such that $(\mathbf{w}, s_k) \in \Sigma^m_k$, and denote by $(\mathbf{w}, s) = (\mathbf{w}, s_k, s_{-k})$ the resulting profile. By Lemma 5, it follows that $(\mathbf{w}, s) \in \Sigma^m$. Then $(\mathbf{w}, s, t) \in R^m$ for some $t \in T$ by the inductive hypothesis. Hence $(\mathbf{w}, s_{-k}, t_{-k}) \in R^m_{-k}$, so $\Sigma^m_{-k} \subseteq \mathrm{proj}_{\Sigma_{-k}} R^m_{-k}$. The proof that $\mathrm{proj}_{\Sigma_{-k}} R^m_{-k} \subseteq \Sigma^m_{-k}$ is analogous and based on Lemma 4.

We now prove the inductive step. Fix a state $\omega = (\mathbf{w}, s, t) \in R^n$. We show that $(\mathbf{w}, s) \in \Sigma^n$. Given a player k, consider the CPS $f_k(t_k)$. By replicating the steps in the proof of the case $n = 1$, we conclude that the pair $(s_k, f_k(t_k))$ satisfies (P1-n), (P3-n), and (P4-n). By assumption, $\omega \in R^{n-1} \cap B_{\varnothing}\left(R^{n-1}\right)$. Hence $\beta_{k,\varnothing}\left(t_k\right)\left(R_{-k}^{n-1}\right) = 1$. So,

$$f_{k,\varnothing}\left(t_k\right)\left(\operatorname{proj}_{\Sigma_{-k}} R_{-k}^{n-1}\right) = 1.$$

By the inductive hypothesis, we have $\operatorname{proj}_{\Sigma_{-k}} R_{-k}^{n-1} = \Sigma_{-k}^{n-1}$, hence $f_{k,\varnothing}\left(t_k\right)\left(\Sigma_{-k}^{n-1}\right) = 1$. So, $\left(s_k, f_k\left(t_k\right)\right)$ satisfies (P2-n). It remains to prove (P5-n). Let $m \in \{0, ..., n-1\}$ and fix a history h such that

$$\left(\mathbf{W} \times S_{-k}\left(h\right)\right) \cap \langle \mathbf{w} \rangle_k \cap \Sigma_{-k}^m \neq \emptyset.$$

By the inductive hypothesis, we have $\left(\mathbf{W} \times S_{-k}\left(h\right) \times T_{-k}\right) \cap [\mathbf{w}]_k \cap R_{-k}^m \neq \emptyset$. So, because $\omega \in SB\left(R^m\right)$, it must be that $\beta_{k,h}\left(t_k\right)\left(R_{-k}^m\right) = 1$. Hence, by applying the inductive hypothesis again, we have

$$f_{k,h}\left(t_k\right)\left(\Sigma_{-k}^m\right) = f_{k,h}\left(t_k\right)\left(\operatorname{proj}_{\Sigma_{-k}} R_{-k}^m\right) = \beta_{k,h}\left(t_k\right)\left(R_{-k}^m\right) = 1$$

We can conclude that $f_k\left(t_k\right)$ satisfies (P5-n). By repeating the argument for every player k we conclude $\left(\mathbf{w}, s\right) \in \Sigma^n$.

In the other direction, fix a profile $\mathbf{w}$, a player k and a strategy s_k such that $\left(\mathbf{w}, s_k\right) \in \Sigma_k^n$. We now show we can choose a type $\tau_k^n\left(\mathbf{w}\,, s_k\right) \in T_k$ such that $\left(\mathbf{w}, s_k, \tau_k^n\left(\mathbf{w}, s\right)\right) \in R_k^n$. Let b_k be a CPS such that $\left(s_k, b_k\right)$ satisfies (P1-n)-(P5-n). We use the following notation: for every player l and every $\left(\mathbf{w}', s_l'\right) \in \mathbf{W} \times S_l$ let

$$m_l\left(\mathbf{w}', s_l'\right) = \max\left\{m = 0, ..., n-1 : \left(\mathbf{w}', s_l'\right) \in \Sigma_l^m\right\}.$$

By the belief-completeness of the type space and by replicating the proof Lemma 6 in Battigalli and Siniscalchi (2007), we can define a type $\tau_k^n\left(\mathbf{w}, s_k\right) \in T_k$, such that

$$\beta_{k,h}\left(\tau_k^n\left(\mathbf{w}, s_k\right)\right)\left(\mathbf{w}', s_{-k}', \left(\tau_l^{m_l\left(\mathbf{w}', s_l'\right)}\left(\mathbf{w}', s_l'\right)\right)_{l \neq k}\right) = b_{k,h}\left(\mathbf{w}', s_{-k}'\right) \tag{11}$$

for all $h \in H$, $s_{-k}' \in S_{-k}\left(h\right)$ and $\mathbf{w}' \in \mathbf{W}$.

Notice that $f_k\left(\tau_k^n\left(\mathbf{w}, s_k\right)\right) = b_k$. Hence, it follows easily from (P1-n) and (P3-n) that

$$\left(\mathbf{w}, s_k, \tau_k^n\left(\mathbf{w}, s_k\right)\right) \in \operatorname{proj}_{\Omega_k}\left(R \cap B_{k,\varnothing}\left(A\right)\right) \tag{12}$$

By (P2-n) we have $b_{k,\varnothing}\left(\Sigma_{-k}^{n-1}\right)=1$. Let $\left(\mathbf{w}',(s_l')_{l\neq k}\right)$ be in the support of $b_{k,\varnothing}$. Then $m_l\left(\mathbf{w}',s_l'\right)=n-1$ for every l. Hence $\left(\mathbf{w}',s_l',\tau_l^{n-1}\left(\mathbf{w}',s_l'\right)\right)_{l\neq k}\in R_{-k}^{n-1}$. Hence

$$\beta_{k,\varnothing}\left(\tau_k^n\left(\mathbf{w},s_i\right)\right)\left(R_{-k}^n\right)=1 \tag{13}$$

By (P4-n) and the definition of b_k, we have

$$\beta_{k,h}\left(\tau_k^n\left(\mathbf{w},s_i\right)\right)\left([\mathbf{w}]_k\right)=b_{k,h}\left(\langle\mathbf{w}\rangle_k\right)=1 \text{ for all } h\in H. \tag{14}$$

We now need to verify the strong belief property. Let $m\in\{0,...,n-1\}$ and $h\in H$ be such that

$$\left(\mathbf{W}\times S_{-k}(h)\times T_{-k}\right)\cap[\mathbf{w}]_k\cap R_{-k}^m\neq\emptyset$$

hence, $\left(\mathbf{W}\times S_{-k}(h)\right)\cap\langle\mathbf{w}\rangle_k\cap\left(\text{proj}_{\Sigma_{-k}}R_{-k}^m\right)\neq\emptyset$. By the inductive hypothesis, $\left(\mathbf{W}\times S_{-k}(h)\right)\cap\langle\mathbf{w}\rangle_k\cap\Sigma_{-k}^m\neq\emptyset$. Thus, from (P5-$n$), it follows $b_{k,h}\left(\Sigma_{-k}^m\right)=1$. Now consider a profile

$$\omega_{-k}'=\left(\mathbf{w}',\left(s_l',\tau_l^{m_l\left(\mathbf{w}',s_l'\right)}\left(\mathbf{w}',s_l'\right)\right)_{l\neq k}\right)$$

in the support of $\beta_{k,h}\left(\tau_k^n\left(\mathbf{w},s_k\right)\right)$. Because $b_{k,h}\left(\Sigma_{-k}^m\right)=1$ it follows that $m_l\left(\mathbf{w}',s_l'\right)\geq m$ for every $l\neq k$. Hence

$$\left(\mathbf{w}',s_l',\tau_l^{m_l\left(\mathbf{w}',s_l'\right)}\left(\mathbf{w}',s_l'\right)\right)\in R_l^{m_l\left(\mathbf{w}',s_l'\right)}\subseteq R_l^m, \text{ for all } l\neq k.$$

Hence $\omega_{-k}'\in R_{-k}^m$. Thus,

$$\beta_{k,h}\left(\tau_k^n\left(\mathbf{w},s_k\right)\right)\left(R_{-k}^m\right)=1. \tag{15}$$

Now repeat the construction for every player and consider the resulting state of the world $\omega=\left(\mathbf{w},s,\left(\tau_k^n\left(\mathbf{w},s_k\right)\right)_{k\in I\cup J}\right)$. It follows from (12) that $\omega\in R\cap B_\varnothing(A)$, while (14) implies $\omega\in C$. Moreover, (13) implies $\omega\in B_\varnothing\left(R^{n-1}\right)$. Finally, from (15) we conclude

$$\omega\in\bigcap_{m=0}^{n-1}SB\left(R^m\right), \text{ for all } k\in I\cup J.$$

Therefore $\omega\in R\cap C\cap B_\varnothing(A)\cap\bigcap_{m=0}^{n-1}SB\left(R^m\right)=R^n$. This concludes the proof of the inductive step, and the proof of the first claim in the Theorem.

The proof that $proj_\Sigma R^\infty=\Sigma^\infty$ follows from the fact that each event R^n is closed and Ω is compact and can be replicated almost verbatim from the proof of Proposition

2 in Battigalli and Siniscalchi (2007). ■

We conclude this subsection with a lemma about composing different strategies.

Lemma 6 *Fix $n \geq 0$, $\mathbf{w} \in \mathbf{W}$ and $j \in J$. Consider a finite sequence $\left(\mathbf{w}^1, s_j^1\right), ..., \left(\mathbf{w}^m, s_j^m\right)$ in Σ_j^n such that $\mathbf{w}_{\mu^{-1}(j)} = \mathbf{w}^1_{\mu^{-1}(j)} = ... = \mathbf{w}^m_{\mu^{-1}(j)}$. If a strategy s_j is such that*

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

then $(\mathbf{w}, s_j)$ belongs to Σ_j^n.

Proof. For every $r = 1, ..., m$, let b_j^r be a CPS such that s_j and b_j^r satisfy (P1-n)-(P5-n). For every $h \in H_j$, let $r(h) \in \{1, ..., m\}$ be such that $s_j(h) = s_j^{r(h)}(h)$. Define the CPS b_j as $b_{j,h} = b_{j,h}^{r(h)}$ for every $h \in H_j$ and $b_{j,h} = b_{j,h}^1$ for every $h \in H - H_j$. The CPS b_j is well defined (the only non-terminal history different from $\varnothing$ that is reached with positive probability under $b_{j,\varnothing}$ is the history h following no offers by any worker. Because $h \notin H_j$ then $b_{j,h} = b_{j,h}^1$. Since $b_{j,\varnothing} = b_{j,\varnothing}^1$ then the requirement of Bayesian updating is respected). By construction, s_j and b_j satisfy (P2-n), (P3-n) and (P4-n). For every $m \in \{0, ..., n-1\}$ and every history h, if $(\mathbf{W} \times S_{-j}(h)) \times \langle \mathbf{w} \rangle_j \times \Sigma_{-j}^m \neq \emptyset$ then $b_{j,h}^{r(h)}\left(\Sigma_{-j}^m\right) = 1$ hence $b_{j,h}\left(\Sigma_{-j}^m\right) = 1$. Thus (P5-$n$) holds. Finally, the action $s_j(h)$ optimal with respect to $b_{j,h}^{r(h)} = b_{j,h}$ at every history $h \in H_j$. Hence $(\mathbf{w}, s_j) \in \Sigma_j^n$. ■

9.4 Proof of Theorem 2

For every n, denote by Σ_I^n the projection of Σ^n on $\mathbf{W} \times \prod_{i \in I} S_i$ and by Σ_J^n the projection on Σ^n on $\mathbf{W} \times \prod_{j \in J} S_j$. To simplify the notation, we denote by a_i the strategy of player i where i abstains from making offers. Also let $a_I = (a_i)_{i \in I}$ and denote by a_{-i} the vector $(a_{\hat{\imath}})_{\hat{\imath} \in I - \{i\}}$.

Theorem 2 is proved by showing that given a matching outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{ir}$ (and the corresponding blocking game $\Gamma(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$), we have $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{\infty}$ if and only if $(\mathbf{w}, a) \in \Sigma^{\infty}$.

Lemma 7 *For every $\mathbf{w} \in \mathbf{W}$, $i \in I$, $n \geq 1$ and $s_i \in S_i$,*

1. *If $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^n$ then $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$;*

2. *If $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$ then $(\{\mathbf{w}\} \times S) \cap \Sigma^n \neq \emptyset$.*

Proof. (1) Let $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^n$ and consider a worker $\hat{\imath} \neq i$ (recall that $|I| \geq 2$ by assumption). There must exists a CPS $b_{\hat{\imath}}$ such that $b_{\hat{\imath},\varnothing}\left(\Sigma_{-\hat{\imath}}^{n-1}\right) = 1$ and $\text{marg}_{\mathbf{W}\times S_i} b_{\hat{\imath},\varnothing}(\mathbf{w}, a_i) = 1$ (by (P3-n)). So $(\mathbf{w}, a_i) \in \Sigma_i^{n-1}$. Because $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^{n-1}$, then Lemma 5 implies $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$.

(2) Because $(\{\mathbf{w}\} \times a_I \times S_J) \cap \Sigma^{n-1} \neq \emptyset$ then for each player k we can find a probability $\rho_k \in \Delta(\Sigma_{-k})$ that satisfies

$$\rho_k\left(\left\{\left(\mathbf{w}, s'_{-k}\right) \in \Sigma_{-k}^{n-1} : s'_{\hat{\imath}} = a \text{ for all } \hat{\imath} \in I - \{k\}\right\}\right) = 1.$$

We now show the existence of a CPS b_k such that $b_{k,\varnothing} = \rho_k$. Define a vector $(b_{k,h})_{h\in H}$ as follows. Let $b_{k,\varnothing} = \rho_k$. As in the proof of Lemma 3, let H_{-k}^A be the set of histories following no offers from workers $I - \{k\}$. 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 (10) holds when $m = n - 1$. For every such history define $b_{k,h}$ to satisfy $b_{k,h}\left(\langle\mathbf{w}\rangle_k \cap \Sigma_{-i}^m\right) = 1$. Proceeding inductively, decrease m and repeat the argument to obtain the vector $b_k = (b_{k,h})_{h\in H}$. We need to verify that b_k is a well defined CPS. Because $b_{k,\varnothing}$ assigns probability 1 to no offer being made (except possibly by 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 CPS. Any strategies s_k that satisfies $s_k \in r_k(b_k)$ is such that the pair (s_k, b_k) satisfies (P1-n)-(P5-n). A profile s of such strategies satisfies $(\mathbf{w}, s) \in \Sigma^n$. ■

The next two lemmas provide conditions, analogous to the idea of blocking pair, that are sufficient and necessary for a profile $\mathbf{w}$ to satisfy $(\mathbf{w}, a_I) \in \Sigma_I^n$.

Lemma 8 *For every $n \geq 0$, $(\mathbf{w}, a_I) \in \Sigma_I^n$ if and only $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$ and there exists a strategy profile $\left(s_j^*\right)_{j\in J}$ such that (1) $\left(\mathbf{w}, s_j^*\right) \in \Sigma_j^{n-1}$ 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 $(\mathbf{w}, a_I) \in \Sigma_I^n$. Consider an offer (j, q) by worker i such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$. Letting $h = \{(i, j, q)\}$, there must exists a strategy of firm j, which we denote by $s_j^{i,q}$, such that $(\mathbf{w}, s_j) \in \Sigma_j^{n-1}$ and $s_j^{i,q}(h) = r$. If not, then the offer (j, q) would be accepted with probability 1 under any conditional probability $b_{i,\varnothing}$ of any CPS b_i that satisfies (P2-n), contradicting the assumption that $(\mathbf{w}, a_i) \in \Sigma_i^n$. Given a firm

j, define the set

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

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 other strategy s_j such that $(\mathbf{w}, s_j) \in \Sigma_j^{n-1}$. Because the set D_j is finite Lemma 6 implies $\left(\mathbf{w}, s_j^*\right) \in \Sigma_j^{n-1}$. This concludes the proof of the "only if" part.

We now prove the "if" part. Let $\left(s_j^*\right)_{j\in J}$ be a profile of strategies that satisfies conditions (1) and (2) in the statement. For every worker i, let $\rho_i \in \Delta(\Sigma_{-i})$ assign probability 1 to $\left(\mathbf{w}, a_{-i}, \left(s_j^*\right)_{j\in J}\right)$. Because $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$ and $(\mathbf{w}, s_J^*) \in \Sigma_J^{n-1}$ then $\rho_i\left(\Sigma_{-i}^{n-1}\right) = 1$. The strategy a_i is a (strict) best response with respect to ρ_i. Using Lemma 3, we can define a CPS b_i such that $b_{i,\varnothing} = \rho_i$ and a_i and b_i satisfy (P1-n)-(P5-n). Now repeat the construction for every $i \in I$. Because $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$ then we know from Lemma 7 that $(\{\mathbf{w}\} \times S) \cap \Sigma^n \neq \emptyset$. Thus, $(\mathbf{w}, a_I) \in \Sigma_I^n$. ■

Lemma 9 *For every $n \geq 2$, $(\mathbf{w}, a_I)$ belongs to Σ_I^n if and only if $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$ 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*

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

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

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

Proof. We first prove the "only if" part. Suppose $(\mathbf{w}, a_I) \in \Sigma_I^n$. If $s_i = (j,q)$ is such that $\nu(\mathbf{w}_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$ and $\mathbf{w}'$ satisfies

$$\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } (\mathbf{w}', s_i, a_{-i}) \in \Sigma_I^{n-2}$$

then, letting $h = (\{i,j,q\})$, we have $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2} \neq \emptyset$.

Using the fact that $(\mathbf{w}, a_I) \in \Sigma_I^n$, apply Lemma 8 and let $\left(s_j^*\right)_{j\in J}$ be a profile that satisfies $\left(\mathbf{w}, s_j^*\right) \in \Sigma_j^{n-1}$ for every j and also satisfies condition (2) of that Lemma. For each j, let b_j^* be a CPS such that s_j^* and b_j^* satisfy properties (P1-$(n-1)$)-(P5-$(n-1)$). So, b_j^* must satisfy $b_{j,h}^*\left(\Sigma_{-j}^{n-2}\right) = 1$. Because $s_j^*(h) = r$ then $b_{j,h}^*$ must attach strictly positive probability to a profile $\mathbf{w}' \in \mathbf{W}$ such that $\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$ 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 $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$ and assume that the other conditions in the "if" statement are satisfied. We now show that $(\mathbf{w}, a_I) \in \Sigma_I^n$. For every firm j, let H_j^* be the set of histories $h = \{(i, j, q)\}$ following a single 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 $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2} \neq \emptyset$. For every $h \in H_j^*$ we can define, by assumption, a probability $\rho_{j,h} \in \Delta\left(\Sigma_{-j}^{n-2}\right)$ such that

$$\mathrm{marg}_{\mathbf{W}\times S_I} \rho_{j,h}\left(\left(\mathbf{w}^h, s_i, a_{-i}\right)\right) = 1$$

where $\mathbf{w}^h$ satisfies $\phi\left(\mathbf{w}_i^h, \mathbf{f}_j\right) - q \leq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$ and $\mathbf{w}^h_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$. We now extend the vector $(\rho_h)_{h\in H_j^*}$ to the collection of all histories h such that $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2} \neq \emptyset$. To this end, define the probability $\rho_{j,\varnothing}$ to satisfy $\mathrm{marg}_{\mathbf{W}\times S_I} \rho_{j,\varnothing}(\mathbf{w}, a_I) = 1$ and $\rho_{j,\varnothing}\left(\Sigma_{-j}^{n-2}\right) = 1$. This is well defined since $(\mathbf{w}, a_I)$ belongs to Σ_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 $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2} \neq \emptyset$, let $\rho_{j,h}$ satisfy $\rho_{j,h}\left((\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2}\right) = 1$.

The resulting vector of conditional probabilities can now be extended to a CPS. Recall that $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$. So, we can apply Lemma 8 and obtain a profile $s_J^* = \left(s_j^*\right)_{j\in J}$ of strategies that satisfy $\left(\mathbf{w}, s_j^*\right) \in \Sigma_j^{n-2}$ for every j together with condition (2) of that Lemma. For each j, let b_j^* be a CPS such that s_j^* and b_j^* satisfy (P1-$(n-2)$)-(P5-$(n-2)$). Define a CPS b_j such that

$$\begin{aligned} b_{j,h} &= \rho_{j,h} \text{ if } h \text{ is such that } (\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2} \neq \emptyset \text{ and} \\ b_{j,h} &= b_{j,h}^* \text{ otherwise} \end{aligned}$$

(see Battigalli (1997) for a similar argument).[10]

Given j, define a strategy s_j such that: (i) $s_j(h) = r$ for every $h \in H_j^*$, (ii) $s_j(h)$ is a best reply to $b_{j,h}$ for every $h \in H_j - H_j^*$ such that $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{n-2} \neq \emptyset$, and (iii) $s_j(h) = s_j^*(h)$ for every other history $h \in H_j$. We now show that the resulting strategy satisfies $(\mathbf{w}, s_j) \in \Sigma_j^{n-1}$. By definition, $s_j(h)$ is optimal with respect to $b_{j,h}$ at every $h \in H_j$. So (s_j, b_j) satisfies (P1$(n-1)$). By the definition of $\rho_{j,\varnothing}$, it satisfies

[10] As before, to verify that the CPS b_i 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.

(P2-$(n-1)$) and (P3-$(n-1)$). Property (P4-$(n-1)$) is easily seen to hold. To verify (P5-$(n-1)$), let $m \in \{0, ..., n-2\}$ and h be such that $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^m \neq \emptyset$ holds. If $m = n-2$ then $b_{j,h}\left(\Sigma_{-j}^{n-2}\right) = \rho_{j,h}\left(\Sigma_j^{n-2}\right) = 1$. If $m < n-2$, then $b_{j,h}\left(\Sigma_{-j}^m\right) = b_{j,h}^*\left(\Sigma_{-j}^m\right) = 1$. So, (P5-$(n-1)$) is satisfied. Hence $(\mathbf{w}, s_j) \in \Sigma_j^{n-1}$.

Repeat the construction for every j and consider the resulting profile $(s_j)_{j \in J}$ of workers' strategies. 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 $h = \{(i, j, q)\}$ be the corresponding history. If $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^{n-2}$ then $h \in H_j^*$ so $s_j(h) = r$. If $(\mathbf{w}, s_i, a_{-i}) \notin \Sigma_I^{n-2}$ then $s_j(h) = s_j^*(h) = r$. To conclude, the strategy profile $(s_j)_{j \in J}$ satisfies properties (1) and (2) in the statement of Lemma 8. Because $(\mathbf{w}, a_I) \in \Sigma_I^{n-1}$, then the same Lemma implies $(\mathbf{w}, a_I) \in \Sigma_I^n$. ■

The following result implies that the procedure of definition 6 does not rule out offers that under *complete* information would lead to a blocking pair.

Lemma 10 *Let $(\mathbf{w}, a_I) \in \Sigma_I^n$. If the worker i and the offer $s_i = (j, q)$ are such that*

$$\nu(\mathbf{w}_i, \mathbf{f}_j) + q \quad > \quad \nu\left(\mathbf{w}_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}, \text{ and} \tag{16}$$

$$\phi(\mathbf{w}_i, \mathbf{f}_j) - q \quad \geq \quad \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j} \tag{17}$$

then $(\mathbf{w}, s_i) \in \Sigma_i^n$.

Proof. Because $(\mathbf{w}, a_I) \in \Sigma_I^n$, we can apply Lemma 8. Let $\left(s_j^*\right)_{j \in J}$ 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 (16) and (17) hold. Let $h = \{(i, j, q)\}$. Define s_j as the strategy such that $s_j(h) = i$ and $s_j(h') = s_j^*(h')$ for every $h' \in H_j, h' \neq h$. So, the strategy s_j accepts the offer (j, q) and rejects any other offer that would improve i's payoff strictly above the status quo. We now show, inductively, that $(\mathbf{w}, s_i) \in \Sigma_i^m$ and $(\mathbf{w}, s_j) \in \Sigma_j^{m-1}$ for every $m \in \{1, ..., n\}$.

Given (16), the claim is easily seen to hold for $m = 1$. Suppose it is true for $m \in \{1, ..., n-1\}$. We now show that $(\mathbf{w}, s_i) \in \Sigma_i^{m+1}$ and $(\mathbf{w}, s_j) \in \Sigma_j^m$. Let b_j^* be a CPS such that s_j^* and b_j^* satisfy (P1-$(n-1)$)-(P5-$(n-1)$). Define a new CPS b_j as follows: if $h = \{(i, j, q)\}$ then $b_{j,h}$ assigns probability 1 to

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

and if $h' \neq h$ then $b_{j,h'} = b_{j,h'}^*$. Inequality (17) implies that $s_j(h)$ is optimal with respect to $b_{j,h}$. It follows that s_j and b_j satisfy (P1-m). It is immediate to verify that they also satisfy (P2-m)-(P4-m). To verify (P5-m), consider first the history $h = \{(i, j, q)\}$. Because $(\mathbf{w}, s_i) \in \Sigma_i^m \subseteq \Sigma_i^{m-1}$ by the inductive hypothesis and $(\mathbf{w}, a_I) \in \Sigma_I^n \subseteq \Sigma_I^{m-1}$ by assumption, then, by Lemma 5, we have $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^{m-1}$. Similarly, because $(\mathbf{w}, s_j) \in \Sigma_j^{m-1}$ and $\left(\mathbf{w}, s_{\hat{j}}^*\right) \in \Sigma_{\hat{j}}^{n-1}$ for every $\hat{j} \neq j$, we have $\left(\mathbf{w}, s_j, \left(s_{\hat{j}}^*\right)_{\hat{j} \in J - \{j\}}\right) \in \Sigma_J^{m-1}$. Hence

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

so $(\mathbf{W} \times S_{-j}(h)) \cap \langle \mathbf{w} \rangle_j \cap \Sigma_{-j}^{m-1} \neq \emptyset$ and $b_{j,h}\left(\Sigma_{-j}^{m-1}\right) = 1$. It follows from the definition of b_j^* and the fact that $\left(\mathbf{w}, s_j^*\right) \in \Sigma_j^{n-1}$ that property (P5-m) is verified with respect to any other history $h' \neq h$. We can conclude that $(\mathbf{w}, s_j) \in \Sigma_j^m$. Define a probability $\rho_i \in \Delta^H(\mathbf{W} \times S_{-i})$ assigning probability 1 to

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

Because $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^m$ and $\left(\mathbf{w}, s_j, \left(s_{\hat{j}}^*\right)_{\hat{j} \in J - \{j\}}\right) \in \Sigma_J^m$, it follows that ρ_i satisfies (P2-$(m+1)$). Moreover, offer $s_i = (j, q)$ is the unique optimal strategy with respect to the probability ρ_i. By applying Lemma 3, we can define a CPS b_i such that $b_{i,\varnothing} = \rho_i$ and such that s_i and b_i satisfy (P1-$(m+1)$)-(P5-$(m+1)$). Therefore $(\mathbf{w}, s_i) \in \Sigma_i^{m+1}$. This concludes the proof of the inductive step. We conclude that $(\mathbf{w}, s_i) \in \Sigma_I^n$. ■

If (i, j, q) is a combination that satisfies (5)-(9) in the definition of Λ^n, then the outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ is said to be Λ^n-*blocked* by (i, j, q). The next two lemmas are the main steps in the proof of Theorem 2.

Lemma 11 *Given $(\mathbf{f}, \mu, \mathbf{p})$ there exists an $\bar{\varepsilon} > 0$ such that for every $\varepsilon \in (0, \bar{\varepsilon})$, every $n \geq 0$, and every $\mathbf{w} \in \mathbf{W}$ such that $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{ir}$, if $(\mathbf{w}, a_I) \in \Sigma_I^{3n}$ then $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n$.*

Proof. We first define $\bar{\varepsilon}$. For every $\mathbf{w} \in \mathbf{W}$ such that $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^\infty$, let $n_\mathbf{w} \geq 0$ satisfy $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{n_\mathbf{w}} - \Lambda^{n_\mathbf{w}+1}$ and select a combination (i, j, q) that $\Lambda^{n_\mathbf{w}}$-blocks $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$. By replicating the argument for every $\mathbf{w} \in \mathbf{W}$, we obtain a finite set B of combinations. It is easy to see that for every $(i, j, q) \in B$, the corresponding payment q must belong to the interval $[q_*, q^*]$. Because definition 5 only involves strict inequalities, for each $(i, j, b) \in B$ that $\Lambda^{n_\mathbf{w}}$-blocks an outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$, there exists a small enough $\varepsilon^{(i,j,b)} > 0$ such that for every $\varepsilon \in \left(0, \varepsilon^{(i,j,b)}\right)$ there exists a payment q' belonging to the grid Q_ε such that the combination (i, j, q') also $\Lambda^{n_\mathbf{w}}$-blocks the same outcome. Let $\bar{\varepsilon} = \min\left\{\varepsilon^{(i,j,b)} : (i, j, b) \in B\right\}$. Because B is finite, then $\bar{\varepsilon} > 0$.

When $n = 0$ the result follows by the fact that $\Lambda^{ir} \subseteq \Lambda^0$ and $\Sigma_I^0 = \Sigma_I$. Proceding inductively, assume that the result is true for $n \geq 0$. Let $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^{n+1}$. We now show that $(\mathbf{w}, a_I) \notin \Sigma^{3n+3}$. Assume that $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n - \Lambda^{n+1}$. This assumption is without loss of generality since, if $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^n$ then $(\mathbf{w}, a_I) \notin \Sigma_I^{3n}$ by the inductive hypothesis. From the definition of $\bar{\varepsilon}$ it follows that if $\varepsilon \in (0, \bar{\varepsilon})$ then we can find a tuple (i, j, q) that Λ^n-blocks $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ and such that $q \in Q_\varepsilon$. Hence, (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 $\phi(\mathbf{w}_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. Suppose, by way of contradiction, that $(\mathbf{w}, a_I) \in \Sigma_I^{3n+3}$. We now show that Lemma 9 leads to a contradiction. Let $s_i = (j, q)$. 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 10 implies $(\mathbf{w}, s_i, a_{-i}) \in \Sigma_I^{3n+3}$. Consider now any profile $\mathbf{w}'$ that, as $\mathbf{w}$, satisfies

$$\begin{aligned} \mathbf{w}'_{\mu^{-1}(j)} &= \mathbf{w}_{\mu^{-1}(j)}, \text{ and} \\ (\mathbf{w}', s_i, a_{-i}) &\in \Sigma_I^{3n+1}. \end{aligned}$$

Because $(\mathbf{w}', s_i, a_{-i}) \in \Sigma_I^{3n+1}$ then Lemma 7 implies $(\mathbf{w}', a_I) \in \Sigma_I^{3n}$. By the inductive hypothesis, we conclude that $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n$. Therefore $\phi(\mathbf{w}'_i, \mathbf{f}_j) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. Because this is true for any such profile $\mathbf{w}'$, the strategy $s_i = (j, q)$ satisfies all the conditions that by Lemma 9 imply $(\mathbf{w}, a_I) \notin \Sigma_I^{3n+3}$. A contradiction. Thus, $(\mathbf{w}, a_I) \notin \Sigma_I^{3n+3}$. ■

Lemma 12 *For every n and every $\mathbf{w}$, if $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{ir} \cap \Lambda^n$ then $(\mathbf{w}, a_I) \in \Sigma_I^{2n}$.*

Proof. Consider first the case where $n = 1$. Let $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{ir}$. Suppose $(\mathbf{w}, a_I) \notin \Sigma_I^2$. We now show $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^1$. It is immediate to verify that $(\mathbf{w}, a_I) \in \Sigma_I^1$. This follows from the fact that abstaining is a best response to the belief that all offers are rejected. Because $(\mathbf{w}, a_I) \in \Sigma_I^1 - \Sigma_I^2$, Lemma 9 implies we can find 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 all $\mathbf{w}' \in \mathbf{W}$ such that $\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)}$ and $(\mathbf{w}', s_i, a_{-i}) \in \Sigma_I^0 = \Sigma_I$. Because each $\mathbf{w}'$ also satisfies $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^0$, then $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^1$.

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

By Lemma 9 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

$$\mathbf{w}'_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } (\mathbf{w}', s_i, a_{-i}) \in \Sigma_I^{m-1}. \tag{18}$$

Consider now a profile $\mathbf{w}''$ that, as $\mathbf{w}$, satisfies

$$(\mathbf{w}'', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n,\ \mathbf{w}''_{\mu^{-1}(j)} = \mathbf{w}_{\mu^{-1}(j)} \text{ and } \nu(\mathbf{w}''_i, \mathbf{f}_j) + q > \nu\left(\mathbf{w}''_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}. \tag{19}$$

By the inductive hypothesis, we know that $(\mathbf{w}'', a_I) \in \Sigma_I^{2n}$. Because $m \leq 2n+1$, then $\Sigma^{2n+1} \subseteq \Sigma^m$ so $\Sigma^{2n} \subseteq \Sigma^{m-1}$. Thus, $(\mathbf{w}'', a_I) \in \Sigma_I^{m-1}$.

We now show that $(\mathbf{w}'', s_i, a_{-i}) \in \Sigma_I^{m-1}$. This conclusion is reached in three steps. First, fix a profile $\mathbf{w}'$ that satisfies (18) and notice that because $(\mathbf{w}', s_i, a_{-i}) \in \Sigma_I^{m-1}$ there must exists a strategy s'_j such that $\left(\mathbf{w}', s'_j\right) \in \Sigma_j^{m-2}$ and s'_j accepts the offer $s_i = (j, q)$, i.e. $s'_j(\{i, j, q\}) = i$. (If not, then offer s_i could not be a best reply to a CPS b_i such that $b_{i,\varnothing}$ puts probability 1 on Σ_{-i}^{m-2}). In addition, because $\mathbf{w}''_{\mu^{-1}(j)} = \mathbf{w}'_{\mu^{-1}(j)}$, Lemma 5 implies that $\left(\mathbf{w}'', s'_j\right)$ belongs to Σ_j^{m-2}.

Second, because $(\mathbf{w}'', a_I) \in \Sigma_I^{m-1}$ we can apply Lemma 8 and obtain a profile $\left(\mathbf{w}'', \left(s^*_{\hat{\jmath}}\right)_{\hat{\jmath} \in J}\right)$ in Σ_J^{m-2} with the property that for every offer $\hat{s}_i = (\hat{\jmath}, \hat{q})$ such that $\nu(\mathbf{w}''_i, \mathbf{f}_{\hat{\jmath}}) + \hat{q} > \nu\left(\mathbf{w}''_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}$, the strategy $s^*_{\hat{\jmath}}$ satisfies $s^*_{\hat{\jmath}}(\{(i, \hat{\jmath}, \hat{q})\}) = r$.

Third, let s_j be the strategy defined as $s_j(\{i, j, q\}) = i$ and $s_j(h) = s^*_j(h)$ for every $h \in H_j$ different from $\{(i, j, q)\}$. So, s_j is a composition of s'_j and s^*_j. By Lemma 6, $(\mathbf{w}'', s_j) \in \Sigma_j^{m-2}$. Now consider a CPS $b_i \in \Delta^H(\Sigma_{-i})$ such that $b_{i,\varnothing}$ assigns probability

1 to

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

By construction, it satisfies $b_{i,\varnothing}\left(\Sigma^{m-2}_{-i}\right)$. The strategy $s_i = (j, q)$ is a (strict) best reply to b_i. By Lemma 3, b_i can be chosen such that (P1-$(m-1)$)-(P5-$(m-1)$) are satisfied. Hence $(\mathbf{w}'', s_i, a_{-i}) \in \Sigma^{m-1}_I$.

Therefore $\mathbf{w}''$ satisfies (18). Thus $\phi\left(\mathbf{w}''_i, \mathbf{f}_j\right) - q > \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}$. Because this is true for every $\mathbf{w}''$ that satisfies (19), we conclude that (i, j, q) does Λ^n-block the outcome $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$. So $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \notin \Lambda^{n+1}$. ■

To conclude the proof of Theorem 2, let $\bar{\varepsilon}$ be as defined in Lemma 11, and fix n large enough such that $\Sigma^\infty = \Sigma^{3n} = \Sigma^{2n}$ and $\Lambda^\infty = \Lambda^n$. Fix $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^{ir}$. If $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^\infty$, then Lemma 12 implies $(\mathbf{w}, a_I) \in \Sigma^{2n}_I = \Sigma^\infty_I$. Conversely, let $(\mathbf{w}, a_I) \in \Sigma^\infty_I = \Sigma^{3n}_I$. By Lemma 11 we have $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^n = \Lambda^\infty$. So, $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^\infty$ if and only if $(\mathbf{w}, a_I) \in \Sigma^\infty_I$.

9.5 Proofs of Other Results

Proof of Theorem 3. Let $(\mathbf{w}, \mathbf{f}, \mu, \mathbf{p})$ be stable under forward induction. Given the blocking game $\Gamma(\mu, \mathbf{p}, \mathbf{f}, \varepsilon)$, the proof of Theorem 5 implies that $(\mathbf{w}, a_I) \in \Sigma^\infty_I$. Let n be such that $\Sigma^{n-1} = \Sigma^\infty$. Since $(\mathbf{w}, a_I) \in \Sigma^n_I$, by Lemma 9 there exists a strategy profile $\left(s^*_j\right)_{j\in J}$ such that $\left(\mathbf{w}, s^*_j\right) \in \Sigma^{n-1}_j = \Sigma^\infty_j$ 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 5 implies

$$\left(\mathbf{w}, a_I, \left(s^*_j\right)_{j\in J}\right) \in \Sigma^{n-1} = \Sigma^\infty.$$

By Theorem 5 we can find a type profile $t \in T$ such that the state of the world $\omega = \left(\mathbf{w}, a_I, \left(s^*_j\right)_{j\in J}, t\right)$ belongs to $A \cap R \cap B_\varnothing(A) \cap C \cap SB^\infty(R \cap B_\varnothing(A) \cap C)$. ■

Proof of Theorem 4. Suppose (2) is violated. Then, 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

$$\begin{aligned} \phi\left(\mathbf{w}'_i, \mathbf{f}_j\right) - q &\geq \phi\left(\mathbf{w}_{\mu^{-1}(j)}, \mathbf{f}_j\right) - \mathbf{p}_{\mu^{-1}(j),j}, && (20)\\ \nu\left(\mathbf{w}'_i, \mathbf{f}_j\right) + q &> \nu\left(\mathbf{w}'_i, \mathbf{f}_{\mu(i)}\right) + \mathbf{p}_{i,\mu(i)}, \text{ and} && (21)\\ (\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) &\in \Lambda^\infty. \end{aligned}$$

Because $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^\infty$ then $(\mathbf{w}', a_I) \in \Sigma_I^\infty$. Let $s_i = (j, q)$. Then, (20), (21) and Lemma 10 imply $(\mathbf{w}', s_i) \in \Sigma_i^\infty$. We now show that $(\mathbf{w}, s_i) \in \Sigma_i^\infty$, concluding that (1) must be violated. The proof is similar to the proof of Lemma 10. Because $(\mathbf{w}', s_i) \in \Sigma_i^\infty$, there must exists a strategy s_j such that s_j accepts the offer (j, q) and $(\mathbf{w}', s_j) \in \Sigma_j^\infty$. Because $(\mathbf{w}, a_I) \in \Sigma_I^\infty$, by Lemma 8 we can find a strategy profile $\left(s_j^*\right)_{j \in J}$ such that $\left(\mathbf{w}, s_j^*\right) \in \Sigma_j^\infty$ 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, $s_j'(h) = s_j(h) = i$. At every other history h, $s_j'(h) = s_j^*(h)$. By Lemma 6, $\left(\mathbf{w}, s_j'\right) \in \Sigma_j^\infty$. Let b_i' be a CPS 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)-(P-n), where $\Sigma^n = \Sigma^\infty$. Hence $(\mathbf{w}, s_i) \in \Sigma_i^\infty$.

Conversely, suppose (1) is violated. Let $(\mathbf{w}, s_i) \in \Sigma_i^\infty$, where $s_i = (j, q)$. We can choose $n \geq 0$ large enough so that $\Sigma^\infty = \Sigma^n = \Sigma^{n-2}$. There must exists a strategy s_j such that $(\mathbf{w}, s_j) \in \Sigma_j^\infty$ 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, b_j) satisfies (P1-n)-(P5-n). By (P5-n) it must be that $b_{j,h}\left(\Sigma_{-j}^\infty\right) = 1$. Hence, there must exists 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 $(\mathbf{w}', s_i) \in \Sigma_i^\infty$. By Lemma 5, $(\mathbf{w}', a_I) \in \Sigma_I^\infty$. Hence $(\mathbf{w}', \mathbf{f}, \mu, \mathbf{p}) \in \Lambda^\infty$. Hence i, j, q and $\mathbf{w}'$ lead to the conclusion that (2) is violated. ■

References

[1] Battigalli, P. (1997): "Strategic Rationality Orderings and the Best Rationalization Principle," *Games and Economic Behavior*, 13(2).

[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).

[3] Battigalli, P. and A. Friedenberg. (2013): "Forward Induction Reasoning Revisited," *Theoretical Economics*, 7(1).

[4] Battigalli, P. and M. Siniscalchi. (1999): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory*, 88(1).

[5] Battigalli, P. and M. Siniscalchi. (2002): "Strong belief and Forward Induction Reasoning," *Journal of Economic Theory*, 106(2).

[6] Battigalli, P. and M. Siniscalchi. (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3(1).

[7] Battigalli, P. and M. Siniscalchi. (2007): "Interactive Epistemology in Games with Payoff Uncertainty," *Research in Economics*, 61(4).

[8] Bikhchandani, S. (2014): "Two-sided Matching with Incomplete Information," *mimeo.*

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

[10] Chade, H. (2006): "Matching with Noise and the Acceptance Curse," *Journal of Economic Theory*, 129(1).

[11] Chade, H., G. Lewis, and L. Smith (2011): "Student Portfolios and the College Admission Problem," *mimeo.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[27] Roth, A. (1989): "Two-sided Matching with Incomplete Information about others' Preferences," *Games and Economic Behavior*, 1(2).

[28] Roth, A. (2002): "The Economist as an Engineer: Game theory, Experimentation, and Computation as Tools for Design Dconomics," *Econometrica*, 70(4).

[29] Roth, A. (2008): "What Have We Learned from Market Design?," *The Economic Journal*, 118.

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

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

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

[33] Van Damme, E. (1989): "Stable Equilibria and Forward Induction," *Journal of Economic Theory*, 48.

[34] Vohra, R. (1999): "Incomplete Information, Incentive Compatibility, and the Core," *Journal of Economic Theory*, 86(1).

[35] Wilson, R. (1978): "Information, Efficiency, and the Core of an Economy," *Econometrica*, 46(4).