Essays on Games with Incomplete Information

by

Ziwei Wang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2022

Date of final oral examination: 05/18/2022

The dissertation is approved by the following members of the Final Oral Committee:

Marzena J. Rostek, Professor, Economics

Daniel S. Quint, Associate Professor, Economics

Marek Weretka, Professor, Economics

Dmitry V. Orlov, Assistant Professor, Finance

To my parents and my wife,
for their unconditional love and support.

Acknowledgments

This has been a long and thorny journey. I am extremely appreciative of the help and guidance from many kind and thoughtful people along the way. They enlightened me about the importance of being independent, grateful, content, and optimistic during hard times.

Bill Sandholm, Antonio Penta, and Xiang Sun taught me game theory. Their wisdom and insights shaped my view of this field. Marzena Rostek has guided and influenced me with her immense passion for economics. I am also grateful to Dan Quint, Marek Weretka, Dmitry Orlov, David Johnson, and Betsy Stovall, from whom I learned a lot as a researcher and an educator. Ignacio Monzón offered invaluable and much-needed support both academically and emotionally, although we have never met in person. Kim Grocholski and Becca George have provided assistance with great patience for us graduate students at Wisconsin.

I will cherish the friendships that gave me the strength to keep fighting over these years. I truly enjoyed spending time with Wanjia Zhu, Chen Zheng, Yue Li, Yixi Yang, Jian Zhang, Xiaoye Tian, Yizhou Kuang, Hoyoung Yoo, Shilong Sun, and Xinrong Zhu. I am also thankful for the endless discussions about research with Xiao Lin and Kun Zhang, who helped me stay focused and triggered my curiosity with a lot of cool things in economic theory.

Finally, I wish to thank my parents Yuanlan Shao and Shilong Wang, whose continuous love and encouragement have made this dissertation possible. My beloved wife, Chenxi Yang, has constantly tolerated my stubbornness and motivated me to change for the better. No words can express my appreciation and gratitude to my family members. I dedicate this dissertation to them.

Contents

List of Figures

Abstract

The study of game theory has advanced our understanding of strategic interactions and economic behaviors. In applications, we economists often use parsimonious game-theoretic models to help us make sharp predictions. However, these models are associated with strong, sometimes unwarranted, common knowledge assumptions about players' payoffs and information. In order to make our predictions realistic and reliable, we need to embed these models into larger and more comprehensive ones, and then perform analysis that are robust to the relaxation of common knowledge assumptions. This dissertation contains three chapters that study various game-theoretic frameworks with incomplete information and investigate the implications of weakened assumptions.

The first chapter proposes a new notion of stability to study matching markets with one-sided incomplete information. A key contribution is to formulate a proper definition of uninformed agents' endogenous beliefs and a self-consistency condition on those beliefs. We define a criterion of stability for a given set of outcomes, and then iteratively apply this criterion to remove outcomes that cannot be deemed stable. Our solution concept, the set of rationalizable stable outcomes, is the limit of this procedure. We prove the existence of rationalizable stable outcomes using a fixed-point characterization. We then provide two additional characterizations of our solution concept. The first characterization links the non-equilibrium approach we pursue to the equilibrium approach pioneered by Liu (2020). The second one reveals the epistemic assumptions implicit in the iterative definition.

In the second chapter, we study standard auctions and compare their minimum expected revenues across all information structures. We show that, for a given symmetric common prior of values among bidders, if the seller is uncertain about the correct model of bidders' interim beliefs and evaluates her expected revenue by the worst-case scenario, the all-pay auction performs weakly worse than does the first-price auction. Specifically, we first provide a revenue equivalence result of standard auction formats under the "worst-case" information structure constructed in Bergemann et al. (2017a), which

implies that the minimum expected revenue of the all-pay auction never exceeds that of a first-price auction. We then construct an example to illustrate that the all-pay auction can generate strictly lower expected revenue in some cases.

The third chapter studies predictions that are robust against higher order payoff uncertainty in dynamic games. Common knowledge among players is captured by a preference-information structure, while a type space is used as a concise model of players' initial beliefs. We formulate an interim version of extensive form rationalizability (EFR) and use this solution concept as the starting point of our robustness analysis. Employing a collection-based approach, we provide conditions that fully characterize (i) what refinements of EFR are robust, (ii) when a Structure Theorem (Weinstein and Yildiz, 2007) of EFR holds, and (iii) when the prediction of EFR is generically unique. We then apply these results to study robust refinements of EFR when there is higher order uncertainty about privacy of information or about observability of actions. These applications demonstrate the power of our results and generate interesting observations in dynamic environments.

Chapter 1

Rationalizable Stability in Matching with Incomplete Information

1.1 Introduction

The theory of stability in two-sided matching problems has burgeoned since the seminal work of Gale and Shapley (1962) and Shapley and Shubik (1971). The literature has mostly focused on the case of complete information, where payoffs of all agents are common knowledge.[1] An allocation consists of a matching, which describes who is matched with whom, and a payment scheme, which specifies a payment among each matched pair. An allocation is said to be stable if no agent wants to unilaterally leave his or her partner, and no unmatched pair prefers to be rematched at some payment. The study of stable allocations has provided valuable insights about the patterns of marriage, labor markets, and other environments. Moreover, the theoretical development of matching has informed the design of many real-world markets such as the assignment of American medical students to residency training programs and the reform of school choice programs in major cities across the US.

Incomplete information is pervasive in many matching markets. For example, a firm may not know the productivity of a worker, or whether the worker is a good culture fit. This uncertainty can affect how a firm assesses the status quo or rematching opportunities, which is crucial in formalizing the criterion of stability. Moreover, *asymmetric* information plays an important role in the process of belief formation.

[1] See Roth and Sotomayor (1990) for a survey.

Consider a situation where firm 1 knows something about a worker that firm 2 does not know, and this is common knowledge. Then, firm 2 can potentially make some inferences from either (i) the fact that firm 1 is matched with that worker at a certain payment, or (ii) the fact that firm 1 is matched with someone else and is not willing to be rematched with the worker in question. These inferences then put restrictions on firm 2's beliefs and thus have implications on what allocations can be deemed stable. Notice that the persistence of a stable allocation conveys information to the uninformed agents. Thus, their beliefs are *endogenous* due to the dependence on allocation, which is an endogenous variable in the model.

In this paper, we focus on matching markets with one-sided incomplete information. We consider a worker-firm environment where firms are uncertain about the quality of workers. In the same framework, Liu et al. (2014) employ a non-equilibrium approach to study the stability of outcomes, i.e. conjunctions of a state and an allocation. They formulate a procedure that interatively removes blocked outcomes, resembling the notion of rationalizability in the game-theoretic literature (Bernheim, 1984; Pearce, 1984). In that paper, firms know the quality of their own partners, but their probabilistic beliefs about other workers' quality are left unspecified; payoffs from blocking pairs are thus evaluated by the worst possible case. In contrast, our goal is to examine how firms form and update their beliefs in the decision making process, which is crucial in an economic environment with uncertainty and often leads to sharper predictions.[2] Liu (2017b) points out that beliefs in a stable matching should be endogenously defined, together with the matching itself. Based on this insight, Liu (2020) pioneers an equilibrium approach and studies stable matching *functions* which are mappings from the space of uncertainty to the space of allocations.[3] In particular, firms understand the matching function, and their endogenous beliefs are updated from a prior belief given the observation of the realized allocation. Our approach is different from Liu (2020) in that we introduce endogenous beliefs into a *non-equilibrium* analysis. The motivation is twofold. First, a non-equilibrium analysis does not require firms correctly understand the underlying matching function as it imposes much weaker assumptions on beliefs. Second, even if

[2]See Example 1.2 and the discussion that follows.

[3]The notion of matching function belongs to an equilibrium *approach*, because it describes a relationship between the underlying uncertainty and the consequence of the game; just like in non-cooperative games, a strategy profile serves as a mapping from agents' private information to their final actions. However, there is an important distinction between a stable matching function and an equilibrium strategy profile: The former is defined with respect to the coalitional behavior among agents, while the latter is defined on the premise of individual optimization.

we are committed to making predictions using stable matching functions, we should be concerned about additional signals received by firms that are beyond the primitives of our model. The non-equilibrium solution concept we propose ensures informational robustness — It contains exactly those outcomes that can be realized by stable matching functions when we vary across *all* possible additional signal structures.

An important step of our analysis is the formulation of firms' endogenous beliefs, which we call conjectures. We adopt the notion of conditional probability systems because it allows us to model how firms react to potential pairwise deviations, especially when they are *surprised* by such events.[4] Moreover, conjectures should not be arbitrary in a stable outcome. This is because the stability of an outcome is tested against *all* counterfactual pairwise deviations. If a firm is willing to participate in a particular pairwise deviation but the worker refuses to do so, she can make inferences from such a failure. We formalize this reasoning by a *self-consistency* condition on conjectures, which requires each firm to believe, in a strong sense, that she cannot successfully block the allocation with some worker. This condition is key to ensure the logical consistency of our non-equilibrium analysis in a cooperative setting. We then propose a criterion of stability for a given set of outcomes (Definition 1.4). Specifically, an outcome in this set is said to be stable if we can find a profile of firm conjectures such that individual rationality, no blocking, and self-consistency are satisfied.

Firms can also make inferences from the lack of *other firms'* viable objections. To capture this, we start from the set of all outcomes, and iteratively apply the criterion of stability proposed above. In each round, we remove outcomes that are not stable, which in turn imposes a stronger restriction on firm conjectures in the next round.[5] This procedure converges in finitely many steps, and its limit constitutes our leading solution concept — the set of rationalizable stable outcomes (Definition 1.5).

We define a useful notion of self-stabilizing set, similar to the "best-reply set" in non-cooperative games, and show that the set of rationalizable stable outcomes is the largest self-stabilizing set. Moreover, we show that all complete information stable outcomes

[4] An event inferred from a proposed pairwise deviation contains the states at which the worker is willing to participate. If a firm initially attaches *zero* probability to such an event, we say the firm is "surprised." It is those surprises that we want to deal with by the notion of conditional probability systems.

[5] We assume that the realized allocation is observable to all firms in the market. So the restriction on firm conjectures is just the answer to the following question: Given the allocation I observe, what are the possible states such that the implied outcome (i.e. the conjunction of the state and the allocation) is not removed yet?

are rationalizable stable, which implies existence of rationalizable stable outcomes across states. These basic properties indicate that our solution concept is well-behaved, but we aim to take one step further and justify its formulation as an "appropriate" one. We achieve this by providing two characterizations of rationalizable stable outcomes that reveal (i) its relation to the equilibrium notion of stable matching functions and (ii) the implicit epistemic assumptions underlying its iterative definition.

The first characterization (Proposition 1.3) originates from the insight that, in non-cooperative games, the strategies that are rationalizable are exactly those that can be realized by a subjective correlated equilibrium (see Brandenburger and Dekel, 1987; Battigalli and Siniscalchi, 2003; Bergemann and Morris, 2017). To establish an analogy, we first extend the notion of stable matching functions to a subjective correlated version. This amounts to defining matching functions on an *expansion* which captures an additional signal received by each firm. We then formally show that an outcome is rationalizable stable if and only if there exists a subjective correlated stable matching function that realizes this outcome. To the best of our knowledge, this is the first time such outcome equivalence result has appeared in a cooperative setting. An immediate consequence is that, when we are agnostic about additional signals firms may receive, we should appeal to our solution concept if we want to make an informationally robust equilibrium prediction. In non-cooperative games with incomplete information, different assumptions on the informativeness of expansions lead to different notions of rationalizability.[6] One may wonder which assumption our solution concept corresponds to. Proposition 1.3 implies that the assumption on informativeness is *irrelevant* in our setting, which highlights a striking distinction between cooperative and non-cooperative games. The driving force of this informational irrelevance is firms' ability to observe the realized allocation and update their beliefs based on it. Thus, any additional signal received by a firm can be replicated by the observation of the allocation generated by a particular matching function.

The iterative definition of our solution concept imposes implicit assumptions on firm conjectures. Our second characterization (Proposition 1.5) is aimed at revealing these assumptions. We explicitly model an *epistemic* state as the conjunction of an

[6]See Bergemann and Morris (2017) for a summary in the case of static games. To be more specific, suppose we want to identify the outcomes that can be realized by a subjective correlated equilibrium defined on an expansion. If we vary across all possible expansions, we obtain the solution concept *belief-free rationalizability* (Battigalli and Siniscalchi, 1999; Bergemann and Morris, 2017); however, if we restrict to the expansions that are individually uninformative, we instead obtain the *interim correlated rationalizability* (Dekel et al., 2007) which is a refinement of the former.

outcome and firms' conditional belief hierarchies about the underlying uncertainty. This way, an assumption can be expressed as a measurable subset of an epistemic state space. We say an epistemic state is pairwise rational if no pair of agents would be better off blocking the specified outcome;[7] this cooperative assumption plays the same role as "rationality" in non-cooperative games which requires agents be subjective expected utility-maximizers. We then introduce the strong belief operator (Battigalli and Siniscalchi, 2002), which imposes restrictions on the conditional beliefs of firms. Intuitively, a firm strongly believes an event if she always attaches probability one to it as long as it is not falsified.[8] The reason we employ the *strong* belief operator, instead of the plain belief operator, is that we need to restrict firm conjectures upon being surprised by a pairwise deviation. Finally, we express our assumptions by an event that corresponds to "pairwise rationality and common strong belief in pairwise rationality (PRCSBPR)" in the universal epistemic state space (i.e. the space that contains all possible conditional belief hierarchies). Proposition 1.5 shows the projection of PRCSBPR on the outcome space is exactly the set of rationalizable stable outcomes. At first glance, the requirement of common strong belief may seem at odds with the informational robustness of our solution concept due to the "non-monotonicity" of strong belief.[9] This is not a contradiction because, in our model, conditional beliefs that are relevant for *prediction* are always monotone. Therefore, our focus on the universal epistemic state space ensures robustness.

The remainder of this paper is organized as follows. The rest of this section reviews the related literature. Section 1.2 introduces the model and discusses two examples which illustrate our notion of stability and distinguish it from the existing concepts. Section 1.3 defines our main solution concept and establishes some basic properties. Section 1.4 is the core of this paper, which provides two characterizations and a related result on the common prior assumption. Finally, Section 1.5 concludes with a discussion on allocative (in)efficiency and some directions to extend our analysis.

[7]We interpret individual rationality as a special case of pairwise rationality because "leaving the match alone" can be seen as "forming a blocking pair with oneself."

[8]Stalnaker (1998) independently introduces a similar notion called "absolutely robust belief."

[9]See Battigalli and Siniscalchi (2002) and Battigalli and Friedenberg (2012).

1.1.1 Related Literature

This paper contributes to several strands of the game-theoretic literature. In this subsection, we briefly discuss the more closely related ones.

Matching with Incomplete Information. Roth (1989) first studies the impact of incomplete information on agents' reporting behavior in matching mechanisms. However, he maintains the notion of complete information stability. Chakraborty et al. (2010) focus on the case of interdependent values and study the existence of *stable mechanisms*. Their notion of stability is with respect to the mechanism, because different mechanisms can reveal different information to the agents. Another stream of literature generates insights by studying the assortativity of matching outcomes in the presence of information frictions; see Chade (2006), Chade et al. (2014), and Hoppe et al. (2009) among others. These papers model the matching process as non-cooperative games and use the usual equilibrium concepts to make predictions.

Liu et al. (2014) make the first attempt to formalize a cooperative notion of stability with one-sided incomplete information. Their non-equilibrium solution concept does not model firms' endogenous beliefs in the reasoning process. Chen and Hu (2020) establishes a learning foundation for this stability notion, and Pomatto (2021) offers a forward-induction interpretation using a version of rationalizability in dynamic games. Bikhchandani (2017) introduces a belief restriction in this approach and investigates its implications on efficiency. However, such a belief restriction may result in non-existence of stable outcome; Alston (2020) shows that this non-existence issue is generic. Our paper provides a logical way to introduce endogenous beliefs into a non-equilibrium analysis and develops a well-behaved solution concept.

To formulate a Bayesian theory of stability, Liu (2020) proposes the notion of stable matching functions which can be viewed as an equilibrium concept for matching games. We establish a formal connection between our non-equilibrium solution concept and this equilibrium one.

Liu (2017a) and Chen and Hu (2021) study environments where incomplete information is two-sided.

Informationally Robust Solution Concepts. In non-cooperative games, if we hold the view that agents' behaviors are driven by a Nash equilibrium but want to take into account additional signals agents may receive, then we need to employ solution concepts that are informationally robust. One branch of the literature imposes a common

prior assumption on agents' beliefs. In the case of complete information games, the outcomes that can be achieved as we vary across all signal structures are exactly the set of correlated equilibria, as argued in the groundbreaking work of Aumann (1987). With incomplete information, versions of correlated equilibrium proposed by Liu (2015) and Bergemann and Morris (2016) serve the purpose of informational robustness, depending on the informativeness of extra signals we want to consider.[10]

On the other hand, we can relax the common prior assumption and assume that agents hold subjective views about the underlying uncertainty. For complete information games, Brandenburger and Dekel (1987) show that (correlated) rationalizability (Bernheim, 1984; Pearce, 1984) characterizes all strategies that can be played in a subjective correlated equilibrium. Under incomplete information, if we assume agents only receive individually uninformative signals, then informational robustness is ensured by the interim correlated rationalizability studied in Dekel et al. (2007); however, if we allow additional signals to convey payoff-relevant information to the agents, versions of belief-free rationalizability proposed by Battigalli and Siniscalchi (1999, 2003) and Bergemann and Morris (2017) can be applied to make robust predictions.[11] Propositions 1.3 and 1.4 in our paper show that informativeness of the additional signal structures is *irrelevant* in a cooperative setting, which helps us understand an important distinction between cooperative and non-cooperative games.

Epistemic Game Theory. When we propose a solution concept, we usually make implicit assumptions on agents' rationality and, more importantly, their interactive beliefs. Epistemic game theory explicitly models these assumptions in a formal language, and then obtains solution concepts as their behavioral implications. This framework has proved useful in clarifying different solution concepts in non-cooperative game theory. Dekel and Siniscalchi (2015) offers a review of this literature. Our analysis in Section 1.4.3 illustrates that the same technique applies seamlessly in cooperative games.

For static (non-cooperative) games, Tan and Werlang (1988) characterizes rationalizability using the assumption "rationality and common belief in rationality" under complete information; Battigalli et al. (2011) provide a unifying epistemic analysis of the existing notions of rationalizability under incomplete information. Analyzing dynamic

[10]Bergemann and Morris (2013) illustrate how to use Bayes correlated equilibrium (Bergemann and Morris, 2016) to make sharp informationally robust predictions. They also reverse the perspective and study the identification of parameters of the game under concerns for informational robustness.

[11]See Bergemann and Morris (2009, 2011), and more recently Penta and Ollár (2017, 2021), among others, for applications of belief-free rationalizability in the theory of robust mechanism design.

games is more involved, because we need to deal with surprises agents face at unexpected histories; see Battigalli and Siniscalchi (1999, 2002, 2003). Our epistemic analysis is closer to the one for dynamic games as we employ the strong belief operator to handle a similar notion of surprise. Interestingly, we demonstrate an instance where common strong belief does *not* lead to non-robustness in terms of prediction.

Several recent papers also take an epistemic approach (different from ours) to study matching games with incomplete information; see Pomatto (2021) and Chen and Hu (2021).

Cooperative vs. Non-cooperative Game Theory. Both cooperative and non-cooperative concepts can be used to study economic interactions among a group of agents. The non-cooperative approach has the advantage of delivering sharp predictions due to its assumptions on the structural details of the game. However, those predictions may be sensitive to the assumptions we make. The cooperative approach avoids this by abstracting away from the fine details of strategic interactions. Nash (1953) makes the first attempt to bridge the gap between these two approaches and initiates the so-called "Nash program," intended to establish non-cooperative foundations of cooperative solution concepts. We do not list any papers in this vast literature, but refer the interested readers to a brief introduction by Serrano (2008).

After developing a Bayesian theory of stability with incomplete information, Liu (2020) proposes a new agenda called the "Kreps–Wilson program," which is aimed at deriving novel concepts in cooperative games under incomplete information by borrowing tools and insights from the literature of non-cooperative games. We emphasize that its objective is different from the earlier Nash program in that it focuses on the implications of incomplete information *without* compromising the cooperative features of the environment.[12] Our paper is another step forward, and we believe pursuing this agenda can bring us a deeper understanding of cooperative games and its relation to non-cooperative ones.

[12]For example, in our paper, we do not impose ad hoc assumptions about strategic interactions when negotiating a pairwise deviation, and the interactive beliefs in the epistemic analysis are purely based on the cooperative behavior among pairs of agents.

1.2 The Model

We consider a two-sided one-to-one matching market with transferable utility as in Crawford and Knoer (1981). Following the literature, we call agents on one side of the market workers, and those on the other side firms. The primitives of our model are a *matching game*, which specifies agents' payoffs from matching, and an *information structure*, which describes the uninformed agents' private information about the uncertain state.

1.2.1 Matching Game

Consider a finite set of workers I and a finite set of firms J. We write $i \in I$ for a typical worker (he) and $j \in J$ for a typical firm (she). Let T_i denote the set of payoff types for each worker i, and let $T = \times_{i \in I} T_i$. When worker i of type t_i and firm j are matched, they obtain *matching values* $a_{ij}(t_i)$ and $b_{ij}(t_i)$, respectively. With a slight abuse of notation, we write $a_{ij}(t) \equiv a_{ij}(t_i)$ and $b_{ij}(t) \equiv b_{ij}(t_i)$ where $t = (t_i, t_{-i})$. Moreover, let $a_{ii}(t) = b_{jj}(t) = 0$ for all t, so unmatched agents obtain no value. A *matching game* is formally defined as the tuple

$$\mathcal{M} = \{I, J, T, (a, b)\},$$

where $(a, b) : I \times J \times T \to \mathbb{R}^2$.

1.2.2 Information Structure

All observable payoff relevant characteristics are captured by the identities of the agents. However, the payoff type of each worker is his private information, which may not be observed by firms. Thus, there is incomplete information on one side of the market, as considered in Liu et al. (2014) and Liu (2020). An information structure is formalized by a partition model, which describes hard pieces of information privately received by firms.[13]

[13]See, for example, Aumann (1987). We interpret "information" as what agents "know" about payoffs, instead of what they "believe." Since workers know their own payoffs and this is common knowledge, we do not explicitly include workers' information in the information structure. However, we note that workers' information plays a role in our analyses, and their participation is a crucial part of the cooperative game.

Specifically, let Ω be a finite *state space*. An element $\omega \in \Omega$ is called a *state*, and a subset of Ω is called an *event*. There is a mapping $\tau : \Omega \to T$ that specifies a worker type profile for each state. The information of firm j is described by an *information partition* $\mathcal{P}_j$ of Ω, which consists of a collection of disjoint cells that covers Ω. We can think of a cell $\Pi_j \in \mathcal{P}_j$ as a private signal received by firm j, so she knows that a state in Π_j has realized but cannot distinguish between the states in the same cell. We write $\Pi_j(\omega)$ for the cell that includes the state ω. To summarize, an *information structure* is a tuple

$$\mathcal{I} = \{\Omega, \tau, (\mathcal{P}_j)_{j \in J}\}.$$

We may append a profile of *exogenous beliefs*, $(\beta_j)_{j \in J}$, to the information structure $\mathcal{I}$. They describe firms' subjective assessments about the states, which should be treated as primitives of the model. Each firm j's exogenous belief is captured by a *conditional probability system* β_j on the state space Ω.[14] We shall postpone the definition of conditional probability systems to Section 1.3.1. In fact, these beliefs do not play a role in our main solution concept, but they are essential in an equilibrium analysis and will be used when we establish the connection between our non-equilibrium approach and the equilibrium one (Section 1.4.1). We call the combination of an information structure and a profile of exogenous beliefs, $\{\mathcal{I}, (\beta_j)_{j \in J}\}$, an *information environment*.

1.2.3 Allocation and Outcome

A matching $\mu : I \cup J \to I \cup J$ is a one-to-one mapping that satisfies: (i) $\mu(i) \in J \cup \{i\}$ for all $i \in I$, (ii) $\mu(j) \in I \cup \{j\}$ for all $j \in J$, and (iii) $\mu(i) = j$ if and only if $\mu(j) = i$. A payment scheme $\mathbf{p}$ associated with a matching μ is a vector of real numbers that specifies a payment $p_{i\mu(i)}$ received by worker i and $p_{\mu(j)j}$ paid by firm j. Unmatched agents make no payment, i.e. $p_{ii} = p_{jj} = 0$. Therefore, if worker i is matched with firm j, i.e. $\mu(i) = j$, at a payment p_{ij}, worker i's and firm j's ex post payoffs are $a_{ij}(t) + p_{ij}$ and $b_{ij}(t) - p_{ij}$, respectively.

Definition 1.1. An *allocation* $(\mu, \mathbf{p})$ is a matching μ together with a payment scheme $\mathbf{p}$ associated with μ. An *outcome* $(\omega, \mu, \mathbf{p})$ consists of a state and an allocation.

[14] The conditional probability system β_j on Ω cannot be replaced by a profile of probability distributions $(\beta_j[\cdot \,|\, \Pi_j])_{\Pi_j \in \mathcal{P}_j}$ as in Brandenburger and Dekel (1987). The reason is that a firm can infer new information from either observing the realized allocation or facing a potential pairwise deviation, and this piece of information may not be measurable to the information partition $\mathcal{P}_j$. Hence, a conditional probability system is necessary to ensure that firm j is prepared for any surprises she may face.

Let $\bar{p} = \max_{i,j,t} b_{ij}(t)$, $\underline{p} = \min_{i,j,t} a_{ij}(t)$, and define the *restricted* space of outcomes as

$$O = \{(\omega, \mu, \mathbf{p}) : p_{i\mu(i)} \in [\underline{p}, \bar{p}] \text{ for all } i \in I\}.$$

Any outcome $(\omega, \mu, \mathbf{p}) \notin O$ cannot sustain for any reasonable notion of individual rationality, so we can restrict our attention to this maximal set O. Write A for the projection of O onto the space of allocations, i.e.

$$A = \{(\mu, \mathbf{p}) : (\omega, \mu, \mathbf{p}) \in O \text{ for some } \omega \in \Omega\}.$$

Note that both O and A are compact spaces.

1.2.4 Examples

In the remainder of this section, we present two simple examples to illustrate our notion of stability and its relation to the existing solution concepts in the literature.

Example 1.1. Suppose there are two workers $I = \{i_1, i_2\}$ and one firm $J = \{j\}$. Worker i_1 only has one type, $T_{i_1} = \{m\}$, and worker i_2 can either be a low type or a high type, $T_{i_2} = \{\ell, h\}$. The matching values are given by the following table

	$i = i_1$	$i = i_2$	
t_i	m	ℓ	h
$a_{ij}(t_i)$	-2	-1	-3
$b_{ij}(t_i)$	5	3	10

A state is identified with a worker type profile, and the firm does not know the type of worker i_2. More precisely, we have $\Omega = \{m\ell, mh\}$, τ is the identity map, and $\mathcal{P}_j = \{\Omega\}$. Moreover, let the firm's exogenous belief about worker i_2's type be uniform, i.e. $\beta_j[m\ell] = \beta_j[mh] = \frac{1}{2}$. We now ask the following question: Can the firm be matched with worker i_1 in some (rationalizable) stable outcome?

We claim that, at state $\omega = m\ell$, if the firm is matched with worker i_1, the payment $p_{i_1 j}$ must be in the range $[2, 3]$. To see this, first note that $p_{i_1 j} \in [2, 5]$ because otherwise individual rationality is violated — one of the agents would be better off leaving the match. We next argue that, for any $p_{i_1 j} \in [2, 3]$, the allocation can be supported by a

conjecture ν_j of the firm such that $\nu_j[m\ell] \in (\frac{5}{7}, 1]$.[15] In this case, the firm gets a payoff weakly higher than 2. Suppose the firm and worker i_2 attempt to block the allocation at a payment $q > 1$ (otherwise no type of worker i_2 has incentives to deviate). If $1 < q \leq 3$, then only type ℓ would participate, which means the firm can *infer* a deviating payoff of $3 - q < 2$; if $q > 3$, then both types would participate, which means the firm makes no inference and anticipates an expected payoff of

$$3\nu_j[m\ell] + 10(1 - \nu_j[m\ell]) - q < 5 - q < 2.$$

On the other hand, if $p_{i_1 j} > 3$, the firm gets a payoff strictly less than 2. But then worker i_2 (of type ℓ) can block this allocation with the firm at a payment slightly higher than 1 (so the firm's payoff from the deviation is only slightly lower than 2).

Next consider the state $\omega = mh$. It turns out that, if the firm is matched with worker i_1, the payment $p_{i_1 j}$ must also be in the range $[2, 3]$. For $p_{i_1 j} \in [2, 3]$, the allocation is again supported by a conjecture ν_j such that $\nu_j[m\ell] \in (\frac{5}{7}, 1]$ as argued above. We next suppose $p_{i_1 j} > 3$ towards a contradiction. In this case, the firm *must* attach zero probability to the state $m\ell$, i.e. $\nu_j[m\ell] = 0$. This is because, if the state *were* $m\ell$, worker i_2 *would have agreed* to deviate with the firm at a payment slightly higher than 1. But at the true state $\omega = mh$, such a deviation is not viable (since worker i_2 refuses it). Thus, the firm should believe that worker i_2 is not of type ℓ. This is a self-consistency condition on ν_j which reflects information inferred by the firm from the lack of viable pairwise deviation (see Definition 1.4). Finally, if the firm attaches probability one to the true state, she will block the allocation with worker i_2 at a payment 4. So we conclude that $p_{i_1 j}$ cannot exceed 3.

Therefore, at either state, an allocation that matches the firm with worker i_1 is stable if and only if the payment $p_{i_1 j} \in [2, 3]$. We point out that all reasoning above is independent of the exogenous belief β_j.

As a comparison, if we employ the solution concept *Bayesian stability* proposed by Bikhchandani (2017), the conclusion will be very different:[16] at either state, an allocation that matches the firm with worker i_1 *cannot* be stable. This results from the assumption that the firm evaluates a pairwise deviation according to the exogenous

[15] This is *not* a complete description of the firm's conjecture. For an intuitive illustration, we avoid dealing with conditional probability systems in the examples.

[16] Bikhchandani (2017) assumes that a firm perfectly observes the type of the worker she is matched with, which is not the case in this paper. However, in this example, we only consider allocations where the firm is matched with worker i_1, whose type is common knowledge. So the comparison is valid.

β_j and the worker's willingness to participate (and nothing else). The argument is as follows. If the firm is matched with worker i_1, the highest payoff she obtains is 3 (at payment 2). Regardless of the state, worker i_2 can propose a pairwise deviation with the firm at a payment $q \in (3, \frac{7}{2})$. Then both worker types would be willing to participate, and the firm's expected payoff is

$$\beta_j[m\ell] \cdot 3 + \beta_j[mh] \cdot 10 - q = \frac{13}{2} - q > 3.$$

This means any allocation that matches the firm with worker i_1 is blocked under Bayesian stability. ♢

The main distinction between rationalizable stability and Bayesian stability is that, for the former, the exogenous belief β_j and the endogenous one (i.e. conjecture) ν_j are treated separately and can be different. One may wonder how such divergence can arise. Here we use the example above and offer two distinct interpretations, both of which involve a belief updating process. The first one is that, even though the firm initially assigns equal probabilities to two types of worker i_2, she may receive some *additional* signal (an interview, for example) indicating it is much more likely that worker i_2 is of low type. This helps the firm update her belief to ν_j. Note that such a signal is outside our model, and we implicitly allow for all possible (payoff-relevant) additional signals observed by the firm. This interpretation highlights the informational robustness property of our solution concept.

The second interpretation does not involve any additional signal. Instead, it relies on the firm's observation of the allocation and understanding of a "matching function" which relates the underlying states and the allocations (Liu, 2020). Moreover, we assume that the matching function is defined on an *expanded* state space, while the firm has a more elaborate probabilistic assessment about the auxiliary states. In particular, we let the expanded state space be $\{\phi_*^{m\ell}, \phi_*^{mh}, \phi_\circ^{mh}\}$, with the interpretation that the superscript of an auxiliary state indicates the type profile of the workers. The firm's subjective assessment is represented by a probability distribution ξ_j as follows

$$\xi_j[\phi_*^{m\ell}] = \frac{1}{2}, \quad \xi_j[\phi_*^{mh}] = \frac{\varepsilon}{2}, \quad \text{and} \quad \xi_j[\phi_\circ^{mh}] = \frac{1-\varepsilon}{2},$$

where $\varepsilon > 0$ is sufficiently small. Note that ξ_j is consistent with the firm's exogenous belief β_j, in the sense that $\xi_j[\phi_*^{m\ell}] = \beta_j[m\ell]$ and $\xi_j[\{\phi_*^{mh}, \phi_\circ^{mh}\}] = \beta_j[mh]$. The matching

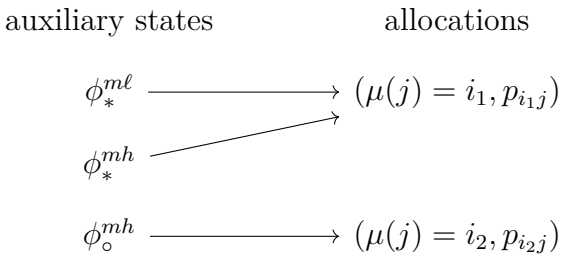


Figure 1.1: A matching function on an expanded environment of Example 1.1

function is defined as in Figure 1.1, where $(\mu(j) = i_1, p_{i_1 j})$ with $p_{i_1 j} \in [2, 3]$ is the allocation to be justified, and $(\mu(j) = i_2, p_{i_2 j}) \in A$ is another allocation that assigns worker i_2 to firm j.

Now given the firm's understanding of the matching function, her subjective assessment ξ_j, and the fact that she observes the realized allocation $(\mu(j) = i_1, p_{i_1 j})$, the firm entertains an updated (endogenous) belief ν_j such that $\nu_j[m\ell] = \frac{1}{1+\varepsilon}$. Since $\nu_j[m\ell] \in (\frac{5}{7}, 1]$ when ε is small enough, it then justifies the stability of $(\mu(j) = i_1, p_{i_1 j})$ when worker i_2 is of either type, as argued in the example. We emphasize that the two interpretations above are conceptually different since no additional signal is observed in the latter one. However, they both lead to the same solution concept in our setting exactly because "observing an allocation" can serve as a belief updating device which has the same effect as "receiving an extra signal."[17]

Due to the flexibility of conjectures, it is tempting to think that our notion of stability is equivalent to the one where firms always evaluate their payoffs by a worst-case reasoning as in Liu et al. (2014). The next example shows this is not the case.

Example 1.2. Suppose there are two workers $I = \{i_1, i_2\}$ and one firm $J = \{j\}$. Each worker can be either a low type or a high type, i.e. $T_{i_1} = T_{i_2} = \{\ell, h\}$. The firm's payoff only depends on the type of the worker she is matched with, but not the identity of the worker. Both workers have identical preferences. The matching values are given by the following table

t_i	ℓ	h
$a_{ij}(t_i)$	-2	-5
$b_{ij}(t_i)$	1	16

[17]The first interpretation reflects the equivalence $(i) \Leftrightarrow (iii)$ in Proposition 1.3, while the second one exhibits the equivalence $(i) \Leftrightarrow (ii)$ in the same result.

Assume that a state is identified with a worker type profile, and the firm knows whether there is *at least one* high type worker. More precisely, we have $\Omega = \{\ell\ell, \ell h, h\ell, hh\}$, τ is the identity map, and $\mathcal{P}_j = \{\{\ell\ell\}, \{\ell h, h\ell, hh\}\}$.[18] We do not specify the firm's exogenous belief in this example. The question we ask here is: Can the firm stay unmatched when at least one worker is of high type?

Let us first consider the stability notion from Liu et al. (2014), which predicts that the firm can be unmatched in a stable allocation at any state $\omega \in \{\ell h, h\ell, hh\}$. The argument is as follows. If the firm stays unmatched, she obtains a payoff of 0. Now suppose worker i_1 proposes to rematch with the firm at a payment $q > 2$. No matter what q is, the firm believes that worker i_1 is of the *worst* type that would participate in such deviation, i.e. type ℓ. This means the firm evaluates the payoff from deviating with worker i_1 by $1 - q < 0$, so she will not accept the rematch. The same argument applies to any potential deviation with worker i_2.

However, if we assume the firm is Bayesian (i.e. forms a subjective belief and then makes decisions based on expected payoffs), she *cannot* stay unmatched in any stable allocation at $\omega \in \{\ell h, h\ell, hh\}$. We prove this by contradiction. Suppose the stable allocation that leaves the firm unmatched is supported by a conjecture $\nu_j \in \Delta(\{\ell h, h\ell, hh\})$. Therefore, the allocation cannot be blocked by the firm and worker i_1 at a payment of 6, which means

$$\nu_j[\ell h] + 16(1 - \nu_j[\ell h]) - 6 \leq 0 \quad \Leftrightarrow \quad \nu_j[\ell h] \geq \frac{2}{3}.$$

Notice that both types of worker i_1 would be willing to participate, so the firm makes no inference from this pairwise deviation. At the same time, the allocation also cannot be blocked by the firm and worker i_2 at a payment of 6, i.e.

$$\nu_j[h\ell] + 16(1 - \nu_j[h\ell]) - 6 \leq 0 \quad \Leftrightarrow \quad \nu_j[h\ell] \geq \frac{2}{3},$$

which leads to a contradiction. In other words, such an allocation will always be blocked by the firm and one of the workers depending on the firm's conjecture, and hence it cannot be (rationalizable) stable at state $\omega \in \{\ell h, h\ell, hh\}$. $\diamondsuit$

[18] As a (somewhat extreme) example, this partition structure can arise if the firm is an economics department, and two job candidates have only one coauthored paper which is to be examined by the department. If the quality of the paper is high, then the department knows *at least* one of the coauthors is a well-trained economist.

As we can see from Example 1.2, both our notion of stability and the one proposed by Liu et al. (2014) are *belief-free* (or exogenous-belief-free), in the sense that exogenous beliefs $(\beta_j)_{j\in J}$ do not play a role in the analyses. However, Liu et al. (2014) further assume that firms are non-Bayesian and use a worst-case reasoning, which means their solution concept is *conjecture-free* (or endogenous-belief-free).[19] Instead, we explicitly model the conjectures of firms in their reasoning process before any pairwise deviation arises, and require these conjectures be updated and used to make decisions across all possible unexpected events in a consistent way.[20] Therefore, our solution concept is more demanding and delivers sharper predictions.

In both examples above, there is only a single firm, so we can make our final predictions after the first round of reasoning. In general, when there are many firms and the information partitions have a non-product structure, more rounds of removal can be performed as firms infer more information from previous rounds and hence refine their beliefs. We formally define our solution concept for general environments in the next section.

1.3 Rationalizable Stable Outcomes

The solution concept we shall formalize is an iterative removal procedure which intuitively captures common (strong) belief, among firms, in individual rationality and no mutually profitable pairwise deviation.[21] In this section, we first describe the criterion of stability for a given set of outcomes. Then we start from the set O and iteratively apply this criterion. The limit of this process is our final prediction, the set of rationalizable stable outcomes.

[19] Here we follow the terminologies from the non-cooperative game literature. When addressing a similar point, Liu (2017b) refers to what we call belief-free as "prior-free," and what we call conjecture-free as "belief-free."

[20] Within a blocking pair, a firm rejecting a rematch according to the worst case is equivalent to her rejecting against all justifiable endogenous beliefs. However, this can lead to inconsistency of beliefs *across* multiple blocking pairs involving the same firm. To see this, notice that in Example 1.2, the unmatched firm does not want to rematch with either worker because of the following reasoning: "If either worker is willing to rematch with me, I believe he is of low type." But it is conceivable that *both* workers can be willing to form a blocking pair with the firm. In that case, the worst-case reasoning seems to suggest that the firm believes both workers are of low type, which contradicts her information indicating at least one worker is of high type. Such inconsistency does not arise in our analysis.

[21] We make this statement formal in Section 1.4.3.

1.3.1 Conjectures and Stability

We assume that firms observe the current allocation. Since firms are uncertain about their payoffs, they entertain endogenous beliefs over states which govern their decisions about whether to maintain the status quo or reject it. We call these beliefs *conjectures* and denote them by $(\nu_j)_{j\in J}$. Fix a set $S \subseteq O$, which is understood as the set of "candidate" outcomes that firms do not yet rule out.[22] Suppose the realized outcome is $(\omega, \mu, \mathbf{p}) \in S$. Upon observing the allocation $(\mu, \mathbf{p})$, firms can infer that the state is an element in

$$S_{(\mu,\mathbf{p})} = \{\omega' \in \Omega : (\omega', \mu, \mathbf{p}) \in S\}.$$

Moreover, for each firm $j \in J$, her private information about the state is given by $\Pi_j(\omega)$. Therefore, she believes a state in $\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$ has realized and conceives a conjecture on this event. When specifying such a conjecture of firm j, one may attempt to define it as a probability distribution over $\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$. However, this is not enough. To see this, suppose that firm j's conjecture attaches zero probability to some states, including the true state ω, in $\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$. When a pairwise deviation is to be carried out by firm j and some worker i, it is possible that only those states to which firm j attaches zero probability can "rationalize" worker i's incentives to participate in the deviation. In this case, firm j is *surprised* (because her belief is falsified), and her updated conjecture conditional on a zero probability event is undefined. In order to resolve this issue, we employ the notion of conditional probability systems introduced by Rényi (1955) (see also Myerson, 1986).

Definition 1.2. Let X be a Polish space with Borel sigma-algebra $\mathcal{B}$. Let $\mathcal{C} \subseteq \mathcal{B}$ be a nonempty countable collection of "conditioning events" such that $\varnothing \notin \mathcal{C}$. A *conditional probability system* (hereafter CPS) on $(X, \mathcal{B}, \mathcal{C})$ is a mapping $\nu[\cdot \,|\, \cdot] : \mathcal{B} \times \mathcal{C} \to [0, 1]$ such that the following two conditions are satisfied:

(i) For every $C \in \mathcal{C}$, $\nu[\cdot \,|\, C]$ is a probability measure on $(X, \mathcal{B})$ and $\nu[C \,|\, C] = 1$;

(ii) If $E_1 \in \mathcal{B}$, $E_2 \in \mathcal{C}$, $C \in \mathcal{C}$, and $E_1 \subseteq E_2 \subseteq C$, then $\nu[E_1 \,|\, E_2] \cdot \nu[E_2 \,|\, C] = \nu[E_1 \,|\, C]$.

Let $\Delta^{\mathcal{C}}(X)$ denote the set of CPSs on $(X, \mathcal{B}, \mathcal{C})$. Moreover, when X is a finite set, we write $\Delta^*(X)$ for the set of CPSs on $(X, 2^X, 2^X \backslash \{\varnothing\})$.[23]

[22] This set will be the input of our iterative definition and will be repeatedly refined in each round (see Section 1.3.2).

[23] The space of uncertainty is a finite set for most of our analysis (because Ω is assumed to be finite),

Remark. What makes the analysis of a cooperative game special is that any sensible solution concept relies on the *true* state, which is usually irrelevant in non-cooperative games. Note that the generality of CPS has a bite only when a firm attaches zero probability to the true state. To avoid this possibility, we could alternatively make the assumption of *grains of truth*: When forming conjectures, firms always attach positive probability to the true state. But this assumption is hard to justify, especially when we take the viewpoint that conjectures are purely subjective. Moreover, we do not wish to rule out situations where a firm assigns zero probability to the true state, when this conjecture cannot be falsified by any pairwise deviation.[24]

Given an outcome $(\omega, \mu, \mathbf{p})$, a *pairwise deviation* (i, j, q) consists of a worker i, a firm $j \neq \mu(i)$, and a payment $q \in \mathbb{R}$. The notation (i, j, q) omits the outcome for brevity. Each firm j entertains a conjecture defined as a CPS on $\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$; if a pairwise deviation (i, j, q) is carried out, an additional piece of information $D_{(i,j,q)}$ is revealed to firm j, which captures worker i's willingness to participate in this pairwise deviation. A firm's conjecture implicitly describes her belief about whether she can successfully deviate with some worker. The following definition is a key component that helps us formalize this description.

Definition 1.3 (No-blocking Sets)**.** Let $S \subseteq O$ be a subset of outcomes and take $(\omega, \mu, \mathbf{p}) \in S$. The *no-blocking set* of firm j with respect to her conjecture $\nu_j \in \Delta^*(\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})})$ is

$$NB_j(\nu_j) = \left\{ \begin{array}{l} \omega' \in \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})} : \\ \qquad \text{There does not exist a pairwise deviation } (i,j,q) \text{ s.t.} \\ \qquad a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}, \text{ and} \\ \qquad \mathbb{E}_{\nu_j}[b_{ij} \mid D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] - q > \mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \end{array} \right\},$$

where $D_{(i,j,q)} = \{\omega'' \in \Pi_j(\omega) : a_{ij}(\tau(\omega'')) + q > a_{i\mu(i)}(\tau(\omega'')) + p_{i\mu(i)}\}$.

so the notation $\Delta^*(X)$ is heavily used. In Section 1.4.3, more complicated spaces of uncertainty are considered, and we shall invoke the more general definition of $\Delta^{\mathcal{C}}(X)$. When the collections $\mathcal{B}$ and $\mathcal{C}$ are understood, we simply say "ν is a CPS on X" without any confusion.

[24] One may attempt to make a stronger (but seemingly more justifiable) assumption to avoid using CPS: Firms are cautious so their conjectures always have full support. Unfortunately, this assumption is logically inconsistent with our iterative removal procedure; a similar point has been made against iterative removal of weakly dominated strategies in non-cooperative games (see Samuelson, 1992; Börgers, 1994).

In the definition above, $\mathbb{E}_{\nu_j}[b_{ij} \mid D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}]$ denotes firm j's expected matching value conditional on the event $D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}$, i.e.

$$\mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] = \sum_{\omega'' \in D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}} b_{\mu(j)j}(\tau(\omega''))\nu_j[\omega'' \mid D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}].$$

Intuitively, the event $NB_j(\nu_j)$ contains the states at which, given firm j's conjecture ν_j, she cannot benefit from participating in any pairwise deviation with some worker.

We now define the notion of stability with respect to a fixed set of outcomes $S \subseteq O$.

Definition 1.4 (S-Stability)**.** Fix a set of outcomes $S \subseteq O$. An outcome $(\omega, \mu, \mathbf{p}) \in S$ is *S-stable* if there exists a profile of conjectures $(\nu_j)_{j \in J}$, where $\nu_j \in \Delta^*(\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})})$ for each $j \in J$, such that the following three conditions are satisfied:

(i) (Individual rationality) The allocation $(\mu, \mathbf{p})$ is individually rational at ω, i.e.

$$\begin{aligned} a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} &\geq 0, \text{ for all } i \in I \text{ and} \\ \mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}] - p_{\mu(j)j} &\geq 0, \text{ for all } j \in J. \end{aligned}$$

(ii) (No blocking) The allocation $(\mu, \mathbf{p})$ is not blocked at ω, i.e. $\omega \in NB_j(\nu_j)$ for all $j \in J$.

(iii) (Self-consistency) Conjectures $(\nu_j)_{j \in J}$ are self-consistent: For every $j \in J$ and conditioning event $C \subseteq \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$ such that $C \cap NB_j(\nu_j) \neq \varnothing$, we have $\nu_j[NB_j(\nu_j) \mid C] = 1$.

Part (i) of Definition 1.4 is a basic requirement for stability: No agent should prefer to leave the match unilaterally. Since incomplete information is one-sided, workers know their ex post payoffs, while firms compute their expected payoffs using their conjectures. Part (ii) says that, at the true state ω, there does not exist a *viable* pairwise deviation that benefits some unmatched pair.

Part (iii) of the definition is more subtle. In words, it requires that every firm j believe the event $NB_j(\nu_j)$ in a strong sense: as long as the conditioning event does not falsify $NB_j(\nu_j)$, firms j assigns probability one to it. This is based on the premise that, in a cooperative setting, a firm should be able to make whatever proposals she likes, and learn from the failure of those proposals. Therefore, each firm j should attach zero probability to any state $\omega' \notin NB_j(\nu_j)$ whenever possible, because if the state *were*

ω', there exists a worker who *would have agreed* to accept some payment and rematch with firm j. This condition is key to ensure the logical consistency of belief formation process in our cooperative environment. Lemma 1.2 in the Appendix ensures that a self-consistent conjecture always exists.

1.3.2 Solution Concept: Rationalizable Stable Outcomes

Given the notion of S-stability, the set of *rationalizable stable outcomes* is computed by iteratively applying this criterion until a limit is reached. The idea is that, if a stable outcome persists, firms can refine their conjectures based on the fact that other firms do not want to block the allocation with some worker, the fact that other firms believe so, and so on.

Definition 1.5 (Rationalizable Stability)**.** Let $S^0 = O$. For each $n \geq 0$, let

$$S^{n+1} = \{(\omega, \mu, \mathbf{p}) \in S^n : (\omega, \mu, \mathbf{p}) \text{ is } S^n\text{-stable}\}.$$

Define $S^\infty = \bigcap_{n \geq 0} S^n$ as the set of *rationalizable stable outcomes*. If $(\omega, \mu, \mathbf{p}) \in S^\infty$, we call $(\mu, \mathbf{p})$ a *rationalizable stable allocation* at ω.

This solution concept has a formal resemblance to the non-cooperative notions of rationalizability. The closest one in spirit is the belief-free rationalizability proposed by Battigalli and Siniscalchi (2003).[25] The definition of S^∞ is also belief-free, in the sense that in each round of removal, the only relevant inputs are the inferred set of states $S^n_{(\mu,\mathbf{p})}$ and the information partitions $(\mathcal{P}_j)_{j \in J}$. However, unlike the ones formulated in Liu et al. (2014) and Chen and Hu (2020), our notion is not conjecture-free — Importantly, we require that all firms entertain conjectures in each round and Bayes' update across different "off-path" scenarios in a consistent way.

In the rest of this section, we establish some basic properties of S^∞, showing that it is a well-behaved solution concept.

Definition 1.6. A nonempty set of outcomes $F \subseteq O$ is *self-stabilizing* if every $(\omega, \mu, \mathbf{p}) \in F$ is F-stable.

The notion of self-stabilizing set is proposed by Liu et al. (2014). It is analogous to the best-reply set in non-cooperative games (Pearce, 1984; Dekel et al., 2007). The

[25] See also Bergemann and Morris (2009, 2017).

following proposition offers a fixed-point characterization of S^∞, which can be used as an equivalent definition.

Proposition 1.1 (Fixed-Point Characterization)**.**

(i) Let L be some index set. If F^ℓ is a self-stabilizing set for every $\ell \in L$, then $F \equiv \bigcup_{\ell \in L} F^\ell$ is also self-stabilizing.

(ii) The set of rationalizable stable outcomes S^∞ is the largest self-stabilizing set.

Proof. See Appendix 1.6.1. □

Bikhchandani (2017) and Alston (2020) demonstrate that their way of imposing belief restrictions in a non-equilibrium analysis leads to non-existence of stable allocations at some states. It is natural to ask whether S^∞ has the same drawback. The next proposition shows that a rationalizable stable allocation exists at every $\omega \in \Omega$.

Proposition 1.2 (Existence)**.** *For every $\omega \in \Omega$, there exists an allocation $(\mu, \mathbf{p})$ such that $(\omega, \mu, \mathbf{p}) \in S^\infty$.*

Proof. Fix an $\omega \in \Omega$. Let $(\mu^\omega, \mathbf{p}^\omega)$ be a complete information stable allocation when ω is commonly known. The existence of such an allocation is established by Shapley and Shubik (1971) and Crawford and Knoer (1981). When firms' conjectures are restricted by the singleton set $\{(\omega, \mu^\omega, \mathbf{p}^\omega)\}$, it is as if there is complete information among firms and workers. Therefore, Definition 1.6 reduces to individual rationality and no blocking of $(\mu^\omega, \mathbf{p}^\omega)$ at ω with complete information, which implies that $\{(\omega, \mu^\omega, \mathbf{p}^\omega)\}$ is a self-stabilizing set. By Proposition 1.1, we have $(\omega, \mu^\omega, \mathbf{p}^\omega) \in S^\infty$. The same argument holds for all $\omega \in \Omega$. □

The proof of Proposition 1.2 shows that the set of all complete information stable outcomes is a subset of S^∞. This is a common feature of the solution concepts formalized in the recent literature (Liu et al., 2014; Liu, 2020; Chen and Hu, 2021). Intuitively, introducing incomplete information does not eliminate agents' incentives to maintain a complete information stable outcome; however, the set of rationalizable stable outcomes can be a proper superset due to the information friction which leads to firms' reluctance to carry out a pairwise deviation, even if it would be profitable to them ex post.

1.4 Informational and Epistemic Characterizations

1.4.1 Relation to the Equilibrium Approach

To develop a Bayesian theory of stability in matching games, Liu (2020) employs an equilibrium approach reminiscent of the classic theory of rational expectations equilibrium (Radner, 1979). In particular, he studies the stability of matching *functions* which describe the relationship between the uncertain states and the observables (i.e. allocations). It is natural to explore the connection between the equilibrium approach of Liu (2020) and the non-equilibrium approach we pursue in this paper. To perform an equilibrium analysis, we now need to specify exogenous beliefs $(\beta_j)_{i\in J}$ as primitives of the model.[26] It turns out that, for any information environment $\{\mathcal{I}, (\beta_j)_{j\in J}\}$, our solution concept S^∞ is *outcome equivalent* to a subjective correlated version of stable matching functions. To formally state this result, we first describe a way to expand the information environment $\{\mathcal{I}, (\beta_j)_{j\in J}\}$, so that each firm can receive an extra signal about the worker type profile.

Definition 1.7 (Expansions)**.** An *expansion* of the information environment $\{\mathcal{I}, (\beta_j)_{j\in J}\}$ consists of the following elements:

(i) A finite set $\Phi \equiv \bigcup_{\omega\in\Omega} \Phi^\omega$, where each Φ^ω is nonempty and $\Phi^\omega \cap \Phi^{\omega'} = \varnothing$ for any $\omega \neq \omega'$;

(ii) A partition $\mathcal{Q}_j$ of Φ for each firm $j \in J$, where a typical cell is denoted by $\Lambda_j \in \mathcal{Q}_j$;

(iii) A CPS $\xi_j \in \Delta^*(\Phi)$ for each firm $j \in J$ that is consistent with β_j in the following sense: For any $E \subseteq \Omega$ and nonempty $C \subseteq \Omega$, we have

$$\xi_j[\Phi^E \,|\, \Phi^C] = \beta_j[E \,|\, C],$$

where we denote $\Phi^{\Omega'} = \bigcup_{\omega\in\Omega'} \Phi^\omega$ for any $\Omega' \subseteq \Omega$.

Let $\mathcal{E} = \{\Phi, (\mathcal{Q}_j)_{j\in J}, (\xi_j)_{j\in J}\}$ denote such an expansion. In addition, if $\mathcal{Q}_j = \{\Phi\}$ for all $j \in J$, we say that $\mathcal{E}$ is an *expansion with trivial partitions*.

We make a few remarks on Definition 1.7 to facilitate the reader's understanding and pave the way for the result to follow:

[26] Recall that for each firm $j \in J$, her exogenous belief $\beta_j \in \Delta^*(\Omega)$ is a CPS defined on the state space Ω.

- Intuitively, when a state $\omega \in \Omega$ is realized, Nature spins a "roulette wheel" with outputs in Φ^ω. Each firm j may not observe the precise realization of the *auxiliary* state $\phi \in \Phi^\omega$, but is informed the cell $\Lambda_j(\phi)$ in her partition $\mathcal{Q}_j$ that contains ϕ. Therefore, we can interpret $\Lambda_j(\phi)$ as an extra signal that firm j receives in addition to the signal $\Pi_j(\omega)$ from $\mathcal{P}_j$.
- Each firm j has a subjective assessment, captured by the CPS $\xi_j \in \Delta^*(\Phi)$, about the auxiliary states in Φ. The consistency requirement in part (iii) ensures that the expansion $\mathcal{E}$ can be viewed as a richer information environment than the original $\{\mathcal{I}, (\beta_j)_{j\in J}\}$.
- Notice that we model the finite set $\Phi \equiv \bigcup_{\omega\in\Omega} \Phi^\omega$ as a *disjoint* union, so each $\phi \in \Phi$ uniquely determines the realized state $\omega \in \Omega$. As a consequence, we can conveniently work with the *expanded state space* Φ in the following analysis. With a slight abuse of notation, we extend the domain of τ to Φ and write $\tau(\phi) \equiv \tau(\omega)$ for all $\phi \in \Phi^\omega$. Moreover, for each firm j, we can envision the information partition $\mathcal{P}_j$ as one of Φ and denote $\Pi_j(\phi) \equiv \Phi^{\Pi_j(\omega)} = \bigcup_{\omega'\in\Pi_j(\omega)} \Phi^{\omega'}$ for all $\phi \in \Phi^\omega$. Hence, whenever we write $\Pi_j(\phi)$, we think of it as a subset of Φ.
- When an expansion $\mathcal{E}$ has trivial partitions, it only introduces more details to the original state space Ω and describes more elaborate subjective assessments $(\xi_j)_{j\in J}$. However, no new information is received by firms.

We now extend the notion of matching functions studied in Liu (2020) and define them on an arbitrary expansion $\mathcal{E}$ of the information environment $\{\mathcal{I}, (\beta_j)_{j\in J}\}$.[27] A *matching function* defined on $\mathcal{E}$,

$$M : \Phi \to A,$$

is a mapping that assigns an allocation $(\mu, \mathbf{p})$ to every auxiliary state $\phi \in \Phi$. As in the formulation of a rational expectations equilibrium, we assume that all agents have correct understanding of this matching function, and their information is updated accordingly after they observe the realized allocation. Next we define the stability of a matching function M.

[27] Although defined in a different language, this extension is in the same spirit as the "stability with private beliefs" in Liu (2017b) and the "correlated stability" in Liu (2020).

Definition 1.8 (Stable Matching Functions; cf. Liu (2020, Definition 5))**.** Given an expansion $\mathcal{E}$ of the information environment $\{\mathcal{I}, (\beta_j)_{j\in J}\}$, a matching function M defined on $\mathcal{E}$ is *stable* if the following two conditions are satisfied:

(i) (Individual rationality) For every $\phi \in \Phi$ and allocation $(\mu, \mathbf{p}) = M(\phi)$, we have

$$\begin{aligned} a_{i\mu(i)}(\tau(\phi)) + p_{i\mu(i)} \geq 0, \text{ for all } i \in I \text{ and} \\ \mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid \Pi_j(\phi) \cap \Lambda_j(\phi) \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j} \geq 0, \text{ for all } j \in J. \end{aligned}$$

(ii) (No blocking) For every $\phi \in \Phi$ and allocation $(\mu, \mathbf{p}) = M(\phi)$, there does not exist a pairwise deviation (i, j, q) such that

$$\begin{aligned} a_{ij}(\tau(\phi)) + q > a_{i\mu(i)}(\tau(\phi)) + p_{i\mu(i)}, \text{ and} \\ \mathbb{E}_{\xi_j}[b_{ij} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - q > \mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j}, \end{aligned}$$

where $D^{\Phi}_{(i,j,q)} = \{\phi' \in \Pi_j(\phi) \cap \Lambda_j(\phi) : a_{ij}(\tau(\phi')) + q > a_{i\mu(i)}(\tau(\phi')) + p_{i\mu(i)}\}$.

In the definition above, for any subset $\Phi' \subseteq \Phi$,

$$\mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid \Phi'] = \sum_{\phi' \in \Phi'} b_{\mu(j)j}(\tau(\phi'))\xi_j[\phi' \mid \Phi']$$

denotes firm j's expectation of her matching value conditional on the event Φ'. For individual rationality, the conditioning event is the join of three pieces of information that firm j receives: the signal $\Pi_j(\phi)$ from the original information partition $\mathcal{P}_j$, the extra signal $\Lambda_j(\phi)$ from the expansion $\mathcal{E}$, and the inference $M^{-1}(\mu, \mathbf{p})$ from understanding the function M and observing the allocation $(\mu, \mathbf{p})$. Similar to our previous analysis, when firm j evaluates a pairwise deviation (i, j, q), there is an additional piece of information revealed by worker i's willingness to participate; therefore, her expected payoff is conditional on a refined event $D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})$.[28]

Note that Definition 1.8 formalizes a subjective correlated version of the stable matching function, because M is defined on an expansion $\mathcal{E}$ which involves firms' additional and possibly correlated private signals. We are ready state the main result of this subsection.

[28]The inference set $D^{\Phi}_{(i,j,q)}$ is analogous to the $D_{(i,j,q)} \subseteq \Omega$ in Definition 1.3. We use a superscript to indicate that $D^{\Phi}_{(i,j,q)}$ is now a subset of Φ.

Proposition 1.3 (Informational Robustness)**.** *Fix an information environment* $\{\mathcal{I}, (\beta_j)_{j\in J}\}$*. The following statements are equivalent:*

(i) An allocation $(\mu, \mathbf{p})$ *is rationalizable stable at* ω*, i.e.* $(\omega, \mu, \mathbf{p}) \in S^\infty$*;*

(ii) There exist an expansion $\mathcal{E}$ *with trivial partitions and a stable matching function* M *defined on* $\mathcal{E}$*, such that* $(\mu, \mathbf{p}) = M(\phi)$ *for some* $\phi \in \Phi^\omega$*;*

(iii) There exist an expansion $\mathcal{E}$ *and a stable matching function* M *defined on* $\mathcal{E}$*, such that* $(\mu, \mathbf{p}) = M(\phi)$ *for some* $\phi \in \Phi^\omega$*.*

Proof. See Appendix 1.6.2. □

Proposition 1.3 deserves some discussion in detail. The equivalence between (i) and (iii) is the cooperative analog of the outcome equivalence between non-equilibrium and equilibrium solution concepts in non-cooperative games (Brandenburger and Dekel, 1987; Battigalli and Siniscalchi, 2003; Bergemann and Morris, 2017). In words, it says that, given an information environment, an outcome is rationalizable stable if and only if it can be realized by some subjective correlated stable matching function. This equivalence result provides an informational robustness foundation for our solution concept S^∞: If we know the baseline information environment and want to predict outcomes that can arise from a stable matching function, but we are agnostic about additional signals firms may receive, then S^∞ delivers a robust prediction.

Perhaps surprisingly, statement (iii) is also equivalent to (ii), which means the existence of unobserved randomness without any revelation is sufficient to generate all possible outcomes through stable matching functions; in other words, to realize the whole set of rationalizable stable outcomes, all we need is a richer state space Φ and more elaborate conditional probability systems $(\xi_j)_{j\in J}$, but *not* the additional information partitions $(\mathcal{Q}_j)_{j\in J}$. This is due to the cooperative nature of matching games and our assumption on the observability of realized allocation. First, an expansion with trivial partitions is by default one with *public* signals, and thus introducing private signals does not expand the set of outcomes that can arise. This is because in our model, the realized allocation is not determined by individual actions, and firms do not interact with each other when contemplating pairwise deviations.[29] On the other hand, an expansion with trivial partitions is *payoff-irrelevant*, in the sense that it does not contain informative

[29]Note that this redundancy of private signals is generally not true in non-cooperative games, where players can use them as a correlating device to coordinate on their actions.

signals that alter firms' beliefs and higher order beliefs about the payoff types of workers (see Liu, 2015; Bergemann and Morris, 2017, for rigorous formulations of this idea). In other words, (individual) informativeness is another property of expansions that is moot in our model. To see this, we note that firms can infer new information about the state from understanding the matching function and observing the realized allocation; therefore, any payoff-relevant information revealed by additional signals can also be revealed by the matching function itself.[30] This argument is the underlying logic of the proof of Proposition 1.3.

1.4.2 Imposing Common Prior

The outcome equivalence established in Proposition 1.3 relies on the non-common prior feature of expansions. In this subsection, we take a little detour to investigate the consequence of imposing a common prior assumption. Put another way, we want to identify a solution concept that corresponds to the *(common prior) correlated equilibrium* defined in non-cooperative games with incomplete information (see Forges, 1993, 2006; Liu, 2015; Bergemann and Morris, 2016). Fix an information environment $\{\mathcal{I}, (\beta_j)_{j\in J}\}$ that admits a common prior, i.e. $\beta_j = \beta$ for all $j \in J$. To simplify our analysis, we assume that $\beta \in \Delta(\Omega)$ has *full support*, so the CPS reduces to a standard probability distribution.[31] In addition, we say $\mathcal{E} = \{\Phi, (\mathcal{Q}_j)_{j\in J}, (\xi_j)_{j\in J}\}$ is a *common prior expansion* of the information environment $\{\mathcal{I}, \beta\}$ if there exists a probability distribution $\xi \in \Delta(\Phi)$ with *full support* such that

$$\xi_j = \xi, \quad \text{for all } j \in J.$$

As suggested by the discussion following Proposition 1.3, introducing unobserved randomness through expansions would be enough to characterize outcomes that can be realized by stable matching functions. In an objective world, this implies what we are after is essentially a stability notion of *stochastic matching functions* defined on the original information environment $\{\mathcal{I}, \beta\}$. We now make this observation formal. A stochastic matching function $\tilde{M}$ defined on $\{\mathcal{I}, \beta\}$,

$$\tilde{M} : \Omega \to \Delta_f(A),$$

is a mapping that specifies, for each state $\omega \in \Omega$, a random allocation defined by a proba-

[30] The discussion following Example 1.1 demonstrates this effect of a matching function.

[31] The generalization to a common CPS $\beta \in \Delta^*(\Omega)$ is straightforward.

bility distribution on A with finite support. For fixed $\{\mathcal{I}, \beta\}$ and $\tilde{M}$, if $\tilde{M}(\omega)[(\mu, \mathbf{p})] > 0$, we let $\beta_j^{(\omega,\mu,\mathbf{p})} \in \Delta(\Pi_j(\omega))$ denote the Bayes' updated belief of firm j when she observes the realized allocation $(\mu, \mathbf{p})$ at state ω, i.e. for each $\omega' \in \Pi_j(\omega)$,

$$\beta_j^{(\omega,\mu,\mathbf{p})}[\omega'] = \frac{\beta[\omega']\tilde{M}(\omega')[(\mu, \mathbf{p})]}{\sum_{\omega'' \in \Pi_j(\omega)} \beta[\omega'']\tilde{M}(\omega'')[(\mu, \mathbf{p})]}.$$

We can now define stable stochastic matching functions as follows.

Definition 1.9 (Stable Stochastic Matching Functions)**.** A stochastic matching function $\tilde{M}$ defined on $\{\mathcal{I}, \beta\}$ is *stable* if the following two conditions are satisfied:

(i) (Individual rationality) For every $\omega \in \Omega$ and allocation $(\mu, \mathbf{p})$ such that $\tilde{M}(\omega)[(\mu, \mathbf{p})] > 0$, we have

$$a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} \geq 0, \text{ for all } i \in I \text{ and}$$
$$\mathbb{E}_{\beta_j^{(\omega,\mu,\mathbf{p})}}[b_{\mu(j)j}] - p_{\mu(j)j} \geq 0, \text{ for all } j \in J.$$

(ii) (No blocking) For every $\omega \in \Omega$ and allocation $(\mu, \mathbf{p})$ such that $\tilde{M}(\omega)[(\mu, \mathbf{p})] > 0$, there does not exist a pairwise deviation (i, j, q) such that

$$a_{ij}(\tau(\omega)) + q > a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)}, \text{ and} \tag{1.1}$$
$$\mathbb{E}_{\beta_j^{(\omega,\mu,\mathbf{p})}}[b_{ij} \mid D_{(i,j,q)}] - q > \mathbb{E}_{\beta_j^{(\omega,\mu,\mathbf{p})}}[b_{\mu(j)j} \mid D_{(i,j,q)}] - p_{\mu(j)j}, \tag{1.2}$$

where $D_{(i,j,q)} = \{\omega' \in \Pi_j(\omega) : a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}\}$.

Note that whenever (1.1) in condition (ii) is satisfied, we have $\omega \in D_{(i,j,q)}$ so the conditional expectation in (1.2) is well-defined. Given a common prior information environment $\{\mathcal{I}, \beta\}$, we say a common prior expansion $\mathcal{E}$ and a matching function M defined on $\mathcal{E}$ *induce* a stochastic matching function $\tilde{M}$ defined on $\{\mathcal{I}, \beta\}$ such that

$$\tilde{M}(\omega)[(\mu, \mathbf{p})] = \sum_{\phi \in \Phi^\omega \cap M^{-1}(\mu,\mathbf{p})} \frac{\xi[\phi]}{\xi[\Phi^\omega]}.$$

We are now ready to state the common prior analogue of Proposition 1.3.

Proposition 1.4 (Informational Robustness with Common Prior)**.** *Fix an information environment $\{\mathcal{I}, \beta\}$ that admits a common prior. The following statements are equivalent:*

(i) $\tilde{M}$ *is a stable stochastic matching function defined on* $\{\mathcal{I}, \beta\}$*;*

(ii) There exist a common prior expansion $\mathcal{E}$ *with trivial partitions and a stable matching function* M *defined on* $\mathcal{E}$ *that induce* $\tilde{M}$*;*

(iii) There exist a common prior expansion $\mathcal{E}$ *and a stable matching function* M *defined on* $\mathcal{E}$ *that induce* $\tilde{M}$*;*

Proof. See Appendix 1.6.3. □

1.4.3 An Epistemic Characterization

In this subsection, we provide a more explicit characterization of S^∞ and show that it characterizes the allocative implications of a cooperative notion of rationality and common strong belief thereof. This axiomatic approach has proved useful in revealing underlying assumptions defining a solution concept and clarifying different versions of rationalizability in the game-theoretic literature (see, for example, Tan and Werlang, 1988; Battigalli et al., 2011). For the sake of simplicity, we assume there are only two firms, $|J| = 2$, but our analysis carries over to more general cases.

In Appendix 1.6.4, we construct a space $H_j(\Pi_j)$ for each firm $j = 1, 2$ that collects all conditional beliefs and higher order conditional beliefs consistent with her information Π_j. An infinite hierarchy of conditional beliefs of firm j is denoted by $\delta_j = (\delta_j^1, \delta_j^2, \ldots) \in H_j(\Pi_j)$. Adapting the analysis from Battigalli and Siniscalchi (1999), it can be shown that there exists a belief-preserving homeomorphism

$$g_{\Pi_j} : H_j(\Pi_j) \to \Delta^{\mathcal{C}_j^\infty(\Pi_j)}\left(Y_j^\infty(\Pi_j)\right),$$

where $Y_j^\infty(\Pi_j) = \{(\omega, \delta_k) : \omega \in \Pi_j \text{ and } \delta_k \in H_k(\Pi_k(\omega))\}$ and $\mathcal{C}_j^\infty(\Pi_j)$ contains the cylinders in $Y_j^\infty(\Pi_j)$ of all nonempty $C \subseteq \Pi_j$; see Appendix 1.6.4 for more details.

Define the *universal epistemic state space* as

$$W = \bigcup_{\omega \in \Omega} \left(\{\omega\} \times A \times H_1(\Pi_1(\omega)) \times H_2(\Pi_2(\omega))\right),$$

which is a compact space. An *epistemic state* $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in W$ is a complete description of the world, consisting of an outcome $(\omega, \mu, \mathbf{p})$ and both firms' conditional belief hierarchies (δ_1, δ_2). A measurable subset $E \subseteq W$ is also called an *event*.

Definition 1.10 (Pairwise Rationality)**.** Agents are *pairwise rational* at an epistemic state $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2)$ if

(i) There does not exist a worker $i \in I$ such that $0 > a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)}$;

(ii) There does not exist a firm $j \in \{1, 2\}$ such that $0 > \mathbb{E}_{\delta_j^1}[b_{\mu(j)j} \mid \Pi_j(\omega)] - p_{\mu(j)j}$;

(iii) There does not exist a pairwise deviation (i, j, q) such that

$$a_{ij}(\tau(\omega)) + q > a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)},$$
$$\mathbb{E}_{\delta_j^1}[b_{ij} \mid D_{(i,j,q)}] - q > \mathbb{E}_{\delta_j^1}[b_{\mu(j)j} \mid D_{(i,j,q)}] - p_{\mu(j)j},$$

where $D_{(i,j,q)} = \{\omega' \in \Pi_j(\omega) : a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}\}$.

Let $PR \subseteq W$ denote the set of all epistemic states where agents are pairwise rational.

In the definition above, $\delta_j^1 \in \Delta^*(\Pi_j(\omega))$ is firm j's *first-order* CPS. Conditions (i) and (ii) are a restatement of "individual rationality" in our previous analysis. We choose this slightly different angle to highlight the interpretation that an agent unilaterally leaving a match can be seen as forming a blocking pair with him/herself. Therefore, the term "pairwise rationality" summarizes these assumptions. Also notice that we no longer require self-consistency as in Definition 1.4. This restriction on beliefs will be incorporated into the belief operator introduced shortly.

Remark. We point out that condition (iii) in Definition 1.10 still embeds an assumption that "firms believe workers are rational when accepting a pairwise deviation." Describing such an assumption formally may be insightful in future research, especially when incomplete information is two-sided. However, it inevitably involves modeling pairwise deviations as primitives of the setting, which we refrain from in this paper.

Lemma 1.1. *$PR \subseteq W$ is an event.*

Proof. See Appendix 1.6.5. □

To reveal the implicit assumptions in S^∞, we employ the *strong belief operator*, which puts restrictions on a firm's conditional beliefs so that she is always certain of a "working hypothesis" whenever it is not falsified (Battigalli and Siniscalchi, 2002; Stalnaker, 1998).

Formally, for an event $E \subseteq W$, the event "firm j strongly believes E" is

$$\mathrm{SB}_j(E) = \left\{ (\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in W : \begin{array}{l} \forall C \subseteq \Pi_j(\omega) \text{ s.t. } C \cap \mathrm{proj}_{\Pi_j(\omega)} E_{(\mu,\mathbf{p},\delta_j)} \neq \varnothing, \\ g_{\Pi_j(\omega)}(\delta_j)\left[E_{(\mu,\mathbf{p},\delta_j)} \,\middle|\, \mathrm{cyl}_j^\infty(C) \right] = 1 \end{array} \right\},$$

where $\mathrm{cyl}_j^\infty(C) = \{(\omega, \delta_k) \in Y_j^\infty(\Pi_j) : \omega \in C\}$ is the cylinder of C in $Y_j^\infty(\Pi_j)$ and

$$E_{(\mu,\mathbf{p},\delta_j)} = \left\{ (\omega, \delta_k) \in Y_j^\infty(\Pi_j) : (\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in E \right\},$$

which reflects the fact that firm j observes the allocation and knows her own beliefs. Moreover, the event "both firms strongly believe E" is

$$\mathrm{SB}(E) = \mathrm{SB}_1(E) \cap \mathrm{SB}_2(E).$$

Note that whenever E is an event, so is $\mathrm{SB}(E)$. With all these definitions in place, we can now define the event "pairwise rationality and common strong belief in pairwise rationality" (PRCSBPR) by the following iteration: Let $PR^1 = PR$; for $n \geq 2$,

$$PR^n = PR^{n-1} \cap \mathrm{SB}(PR^{n-1}).$$

Finally, we let

$$\mathrm{PRCSBPR} = \bigcap_{n \in \mathbb{N}} PR^n.$$

Proposition 1.5 (Epistemic Characterization)**.** *The set of rationalizable stable outcomes characterizes the allocative implications of pairwise rationality and common strong belief in pairwise rationality, i.e.*

$$S^\infty = \mathrm{proj}_O \mathrm{PRCSBPR}.$$

Proof. See Appendix 1.6.6. □

To prove this proposition, we show by induction that $\mathrm{proj}_O PR^{n+1} \subseteq S^n \subseteq \mathrm{proj}_O PR^n$, for all $n \in \mathbb{N}$. This relationship can be easily seen in the first round: Intuitively, PR^1 means agents are pairwise rational, and PR^2 says "PR^1 and each firm strongly believes so;" however, when we compute S^1, we require pairwise rationality, but each firm also strongly believes pairwise rationality holds for *herself* (which is captured by the self-consistency of her conjecture). Therefore, S^1 is in $\mathrm{proj}_O PR^1$ but only "halfway into" $\mathrm{proj}_O PR^2$.

In the following, we make two closely related comments on the epistemic characterization:

- Battigalli and Siniscalchi (2002) argue that common strong belief intuitively captures a "best-rationalization principle," which means agents' beliefs are always consistent with the highest possible degree of sophistication.[32] As we can see from the proof of Proposition 1.5, this principle actually imposes stronger restrictions on conditional beliefs than does the solution concept S^∞ in an inessential way: Only *the highest* degree of sophistication is relevant in computing S^∞, because firms can only be surprised by an event that includes the true state, which (by default) never gets ruled out in the iterative procedure.
- Since the strong belief operator is not monotone,[33] applying common strong belief in the universal epistemic state space can potentially leave out predictions that arise in smaller spaces; see Battigalli and Siniscalchi (2002) and Battigalli and Friedenberg (2012) for a discussion of this issue in the analysis of dynamic games. However, this is *not* the case in our setting as suggested by Proposition 1.3 (informational robustness), which may seem a bit counter-intuitive. A brief comparison can help us understand the underlying logic. In dynamic games, the behavior of a player upon reaching a zero-probability history (i.e. being surprised) depends on her belief about the opponent's behavior at such history, so the non-monotonicity of strong belief has a bite through such interaction. In our cooperative setting, however, any relevant surprise is caused by the formation of a blocking pair with a worker, which always corresponds to an event containing the true state. Therefore, the highest degree of sophistication among firms is never discarded, and hence the non-monotonicity of strong belief does not result in a non-monotonicity of predictions implied by PRCSBPR.

[32]This implies that S^∞ involves a *forward-induction* type of reasoning. Pomatto (2021) makes a related observation about the solution concept proposed by Liu et al. (2014).

[33]That is, even if $E_1 \subset E_2 \subseteq W$, we do *not* have $\text{SB}(E_1) \subseteq \text{SB}(E_2)$.

1.5 Discussion and Extensions

1.5.1 Allocative (In)efficiency

The efficiency property of stable outcomes has been a focus of discussion in the literature. Since our model does not admit a common prior among firms (except Section 1.4.2), we only comment on the *ex post* (in)efficiency of rationalizable stable outcomes.[34]

After imposing suitable assumptions of monotonicity and supermodularity on matching values, Liu et al. (2014) show that *all* outcomes contained in their solution concept are efficient. This is because the assumptions ensure that, by a pairwise deviation, a worker can "signal" his true type to a firm better than his current match. It then leads to assortativity of matching outcomes, which implies efficiency. However, an important assumption in their model, which is key to the proof, is that a firm can perfectly observe the type of the worker that she is matched with. If, instead, matching with a worker does not reveal his type to the firm, as we assume here, inefficiency can easily arise. The reason is that it may be impossible to correct a firm's conjecture about her *own* worker through pairwise deviations. We illustrate this source of inefficiency in an example.

Example 1.3. Suppose there are two workers $I = \{i_1, i_2\}$ and one firm $J = \{j\}$. A worker can be either of a low type, a medium type, or a high type, i.e. $T_{i_1} = T_{i_2} = \{\ell, m, h\}$. The firm does not care about the identity of the worker. The matching values are given by the following table

t_i	ℓ	m	h
$a_{ij}(t_i)$	-3	-2	-1
$b_{ij}(t_i)$	2	4	10

Assume worker i_1 is of low type and worker i_2 is of medium type. However, the firm's information only tells her that i_2 is a medium type worker (and nothing more). We can model this situation as $\Omega = T_{i_1} \times T_{i_2}$, τ is the identity map, the true state $\omega = \ell m$, and $\Pi_j(\omega) = \{\ell m, mm, hm\}$.

[34] Specifically, an outcome $(\omega, \mu, \mathbf{p})$ is (ex post) efficient if μ maximizes

$$\sum_{i=1}^{|I|} a_{i\mu'(i)}(\tau(\omega)) + \sum_{j=1}^{|J|} b_{\mu'(j)j}(\tau(\omega))$$

over all matchings $\mu' : I \cup J \to I \cup J$.

Let $\omega' = hm$ and consider the set of outcomes

$$F = \{(\omega, \mu(j) = i_1, p_{i_1 j} = 5), (\omega', \mu(j) = i_1, p_{i_1 j} = 5)\}.$$

We now argue that F is a self-stabilizing set. First observe that $(\omega', \mu(j) = i_1, p_{i_1 j} = 5)$ is a complete information stable outcome; therefore, it is justified by a conjecture of the firm that attaches probability one to ω' whenever possible. For $(\omega, \mu(j) = i_1, p_{i_1 j} = 5)$, define a conjecture $\nu_j \in \Delta^*(\{\omega, \omega'\}))$ as follows:

$$\nu_j[\omega \,|\, \{\omega, \omega'\}] = \varepsilon, \text{ and } \quad \nu_j[\omega' \,|\, \{\omega, \omega'\}] = 1 - \varepsilon,$$

where $\varepsilon > 0$ is sufficiently small. Hence, the firm's expected payoff from matching with worker i_1 at the payment 5 is $2\varepsilon + 10(1 - \varepsilon) - 5 = 5 - 8\varepsilon$, regardless of the pairwise deviation she is offered to participate. But because worker i_2 is of medium type (which is known by the firm), the firm's payoff from deviating with worker i_2 cannot exceed $2 < 5 - 8\varepsilon$. This means $(\omega, \mu(j) = i_1, p_{i_1 j} = 5)$ is also F-stable and thus F is a self-stabilizing set. Finally, we note that $(\omega, \mu(j) = i_1, p_{i_1 j} = 5)$ is apparently inefficient, but it is a rationalizable stable outcome because $F \subseteq S^\infty$ by Proposition 1.1. $\diamondsuit$

Notice that in Example 1.3, matching values are strictly monotone in worker's type, and supermodularity is vacuously satisfied because there is only a single firm (cf. Liu et al. (2014)). This suggests there is little hope to ensure efficiency across all rationalizable stable outcomes. However, such a negative result should not be surprising, because (ex post) efficiency is usually hindered by incomplete information and thus intrinsically hard to achieve.[35,36]

We can also make the assumption that firms learn the types of their partners once being matched, which can be thought of as receiving an extra piece of signal in addition to the information structure. Once we make this assumption, S^∞ becomes a *refinement* of (a generalization of) the solution concept developed in Liu et al. (2014), so their efficiency result carries over to our setting.

[35] Bikhchandani (2017) shows that the allocative efficiency result established in Liu et al. (2014) fails if utilities are not transferable.

[36] With a common prior assumption, efficiency *can* be achieved in a weaker sense. Developments in this direction can be found in Bikhchandani (2017) and Liu (2020).

1.5.2 Rationalizable Core

We can generalize our approach to develop a theory for the core.[37] Without belaboring the details, we only give a brief description of how to extend the definitions. Given an outcome $(\omega, \mu, \mathbf{p})$, a *coalitional deviation* $(I', J', \mu', \mathbf{q})$ consists of (i) a subset of workers $I' \subseteq I$ and a subset of firms $J' \subseteq J$; (ii) a matching among the deviating agents $\mu' : I' \cup J' \to I' \cup J'$, which is different from the original matching μ restricted to $I' \cup J'$; and (iii) a payment scheme $\mathbf{q}$ associated with the rematch μ'.

As before, fixing a set of outcomes, the conjecture of a firm should be a CPS defined on the intersection of her information cell and the possible states implied by this set. We say an outcome $(\omega, \mu, \mathbf{p})$ is in the core of this set if there exists a profile of firm conjectures such that the following three conditions are satisfied:

(i) (Individual rationality) The allocation $(\mu, \mathbf{p})$ is individually rational at ω;

(ii) (No blocking) There does not exist a viable coalitional deviation at ω;

(iii) (Self-consistency) Whenever possible, each firm j attaches zero probability to the states at which, *regardless* of the conjectures of other firms, there does not exist a viable coalitional deviation $(I', J', \mu', \mathbf{q})$ such that $j \in J'$.

Note that when a coalitional deviation only contains one firm, condition (iii) reduces to the one in Definition 1.4. When the size of coalitions can be larger, self-consistency becomes more restrictive.

With this definition, we can then start from O and iteratively remove outcomes that are not in the core using the criteria above. We may call the limit of this process the set of *rationalizable core outcomes*. Similar results as those in Section 1.3.2 can be established to show that a rationalizable core allocation exists at every state. Since the possibility of deviations is richer for coalitional blocking, the set of rationalizable core outcomes may become a proper refinement of S^∞; Example 8 of Liu (2020) illustrates this strict inclusion in his setting.

Liu (2020) defines a notion of *core matching function* which is parallel to the generalization we laid out above. Reworking the analysis in Section 1.4.1, we can establish an outcome equivalence result à la Proposition 1.3.[38] The epistemic characterization in

[37] Wilson (1978) initiates a discussion on the core of an exchange economy with incomplete information; see Forges et al. (2002) for a survey of the earlier literature.

[38] However, it is unclear whether the equivalence (*ii*) $\Leftrightarrow$ (*iii*) in Proposition 1.3 still holds with coalitional deviations. The reason is that firms' participation in a coalitional deviation can reveal

Section 1.4.3 can also be adapted, and we conjecture that the set of rationalizable core outcomes characterizes the allocative implications of a form of coalitional rationality and common strong belief thereof.

1.5.3 Two-Sided Incomplete Information

A natural direction for future investigation is to extend our framework to the case of two-sided incomplete information. Liu (2017a) illustrates how to extend the equilibrium approach of Liu (2020) to such environments. The main difficulty lies in the modeling of how agents infer new information from a viable pairwise deviation. A fixed-point type of inference is proposed to deal with this: Given the information revealed by the worker, the firm's willingness to participate in the pairwise deviation exactly captures the information revealed by the firm, and vise versa.[39] Chen and Hu (2021) extend the belief-free framework of Chen and Hu (2020) and propose a novel notion of pairwise blocking when incomplete information is two-sided: a worker and a firm block the current allocation if and only if it is "common knowledge" that both parties will benefit from being rematched at some payment. Their approach is conjecture-free, which does not address how agents form and refine their beliefs in the reasoning process. How to perform a non-equilibrium analysis with endogenous beliefs remain an open question, but we believe our paper provides some useful tools and paves the way for future explorations in this direction.

1.6 Appendix

1.6.1 Proof of Proposition 1.1

We first prove a useful lemma. Part (i) of the lemma establishes existence of self-consistent conjectures. Part (ii) provides a condition that can replace self-consistency in the definition of S-stability.

Lemma 1.2. *Let $S \subseteq O$ and take $(\omega, \mu, \mathbf{p}) \in S$.*

(i) A self-consistent conjecture in $\Delta^*(\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})})$ exists for each firm $j \in J$;

their private information, so private signals introduced by expansions may play a role. We leave this interesting question for future research.

[39] A similar argument is used in Dutta and Vohra (2005) to define an interim notion of the core under incomplete information.

(ii) Suppose there is a profile of conjectures $(\nu_j)_{j\in J}$ such that $(\mu, \mathbf{p})$ is individually rational at ω. Moreover, suppose there exists an event $E_j \subseteq \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$ for each $j \in J$ such that $\omega \in E_j \subseteq NB_j(\nu_j)$, and $\nu_j[E_j \,|\, C] = 1$ whenever $C \cap E_j \neq \varnothing$. Then $(\omega, \mu, \mathbf{p})$ is S-stable.

Proof of Lemma 1.2. For part (i), we use an iterative construction to ensure self-consistency.

Take an arbitrary conjecture $\nu_j^0 \in \Delta^*(\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})})$ for firm j. If $NB_j(\nu_j^0) = \varnothing$ or $NB_j(\nu_j^0) = \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$, we are done. Otherwise, define a conjecture $\nu_j^1 \in \Delta^*(\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})})$ as follows: For each conditioning event $C \subseteq \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$, if $C \cap NB_j(\nu_j^0) = \varnothing$, let $\nu_j^1[\cdot \,|\, C] = \nu_j^0[\cdot \,|\, C]$; if $C \cap NB_j(\nu_j^0) \neq \varnothing$, let $\nu_j^1[\cdot \,|\, C] = \nu_j^0[\cdot \,|\, C \cap NB_j(\nu_j^0)]$. This pins down a CPS ν_j^1 defined on $\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}$. By construction, we have $NB(\nu_j^1) \subseteq NB_j(\nu_j^0)$.

Iteratively, for every $n \in \mathbb{N}$, if $NB_j(\nu_j^n) = \varnothing$ or $NB_j(\nu_j^n) = NB_j(\nu_j^{n-1})$, we are done. Otherwise, define a conjecture $\nu_j^{n+1} \in \Delta^*(\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})})$ as follows: For each $C \subseteq \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$, if $C \cap NB_j(\nu_j^0) = \varnothing$, let $\nu_j^{n+1}[\cdot \,|\, C] = \nu_j^0[\cdot \,|\, C]$; otherwise, let

$$m_j(C) = \max\{n' \leq n : C \cap NB_j(\nu_j^{n'}) \neq \varnothing\},$$

and $\nu_j^{n+1}[\cdot \,|\, C] = \nu_j^0[\cdot \,|\, C \cap NB_j(\nu_j^{m_j(C)})]$. This again pins down a CPS ν_j^{n+1} defined on $\Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}$. By construction, we have $NB_j(\nu_j^{n+1}) \subseteq NB_j(\nu_j^n)$.

Since $NB_j(\nu_j^0)$ is finite, this iterative process stops at a finite step $\bar{n}$, i.e. either $NB_j(\nu_j^{\bar{n}}) = \varnothing$ or $NB_j(\nu_j^{\bar{n}}) = NB_j(\nu_j^{\bar{n}-1})$. By definition, $\nu_j^{\bar{n}}$ is a self-consistent conjecture.

For part (ii), we set $\nu_j^0 = \nu_j$ and start the iterative construction in part (i). We claim that for all $0 \leq n \leq \bar{n}$, we have $E_j \subseteq NB_j(\nu_j^n)$ for each firm $j \in J$. By assumption, this is true for $n = 0$. Suppose the claim holds for n, and consider the profile $(\nu_j^{n+1})_{j\in J}$. For each firm $j \in J$ and state $\omega' \in E_j$, take any pairwise deviation (i, j, q) such that

$$a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}.$$

This implies $\omega' \in D_{(i,j,q)} \cap E_j \neq \varnothing$ and thus $m_j(D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}) = n$. Then we have

$$\nu_j^{n+1}[\cdot \,|\, D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] = \nu_j[\cdot \,|\, D_{(i,j,q)} \cap NB_j(\nu_j^n)] = \nu_j[\cdot \,|\, D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}],$$

where the first equality is by construction of ν_j^{n+1}, and the second one is due to the

assumption on ν_j and the induction hypothesis $E_j \subseteq NB_j(\nu_j^n)$. Therefore, we must have

$$\begin{aligned}\mathbb{E}_{\nu_j^{n+1}}[b_{ij} \,|\, D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] - q &= \mathbb{E}_{\nu_j}[b_{ij} \,|\, D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] - q \\ &\leq \mathbb{E}_{\nu_j}[b_{\mu(j)j} \,|\, D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \\ &= \mathbb{E}_{\nu_j^{n+1}}[b_{\mu(j)j} \,|\, D_{(i,j,q)} \cap S_{(\mu,\mathbf{p})}] - p_{\mu(j)j}.\end{aligned}$$

This means $\omega' \in NB_j(\nu_j^{n+1})$. Since $\omega' \in E_j$ was arbitrary, we have $E_j \subseteq NB_j(\nu_j^{n+1})$. Hence, induction implies that $E_j \subseteq NB_j(\nu_j^{\bar{n}})$. Finally, we argue that the profile of self-consistent conjectures $(\nu_j^{\bar{n}})_{j\in J}$ supports the S-stability of $(\omega, \mu, \mathbf{p})$.

(Individual rationality.) By assumption, $a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} \geq 0$ for all $i \in I$. For each firm $j \in J$, by construction of $\nu_j^{\bar{n}}$ and the fact that $E_j \subseteq NB_j(\nu_j^{\bar{n}})$, we have $\nu_j^{\bar{n}}[\cdot \,|\, \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}] = \nu_j[\cdot \,|\, \Pi_j(\omega) \cap NB_j(\nu_j^{\bar{n}})] = \nu_j[\cdot \,|\, \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}]$. Therefore,

$$\mathbb{E}_{\nu_j^{\bar{n}}}[b_{\mu(j)j} \,|\, \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}] - p_{\mu(j)j} = \mathbb{E}_{\nu_j}[b_{\mu(j)j} \,|\, \Pi_j(\omega) \cap S_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0$$

(No blocking.) This is implied by the fact that $\omega \in E_j \subseteq NB_j(\nu_j^{\bar{n}})$ for all $j \in J$. $\square$

We can now prove Proposition 1.1.

(i) Take any outcome $(\omega, \mu, \mathbf{p}) \in F$. It must belong to a self-stabilizing set F^ℓ. For each firm $j \in J$, let $\nu_j^\ell \in \Delta^*(\Pi_j(\omega) \cap F^\ell_{(\mu,\mathbf{p})})$ denote the conjecture that supports the F^ℓ-stability of $(\omega, \mu, \mathbf{p})$. We can extend these conjectures to a profile $(\nu_j)_{j\in J}$, where $\nu_j \in \Delta^*(\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})})$ for each $j \in J$, as follows: For each conditioning event $C \subseteq \Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}$, if $C \cap F^\ell_{(\mu,\mathbf{p})} \neq \varnothing$, let $\nu_j[B \,|\, C] = \nu_j^\ell[B \cap F^\ell_{(\mu,\mathbf{p})} \,|\, C \cap F^\ell_{(\mu,\mathbf{p})}]$ for all $B \subseteq C$; if $C \cap F^\ell_{(\mu,\mathbf{p})} = \varnothing$, let $\nu_j[\cdot \,|\, C] = \text{Unif}(C)$.[40] This construction pins down a CPS ν_j on $\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}$ for each $j \in J$.

Since $(\nu_j^\ell)_{j\in J}$ supports the F^ℓ-stability of $(\omega, \mu, \mathbf{p})$, we have $a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} \geq 0$ for all $i \in I$ and

$$\mathbb{E}_{\nu_j}[b_{\mu(j)j} \,|\, \Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j} = \mathbb{E}_{\nu_j^\ell}[b_{\mu(j)j} \,|\, \Pi_j(\omega) \cap F^\ell_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0, \quad \forall j \in J.$$

Moreover, $\omega \in NB_j(\nu_j^\ell) \subseteq NB_j(\nu_j)$ for each $j \in J$ by construction of ν_j. Also notice that $\nu_j[NB_j(\nu_j^\ell) \,|\, C] = 1$ whenever $C \cap NB_j(\nu_j^\ell) \neq \varnothing$ by the self-consistency of ν_j^ℓ, we can invoke Lemma 1.2 and conclude that $(\omega, \mu, \mathbf{p})$ is F-stable.

[40] For a finite set X, let $\text{Unif}(X)$ denote the distribution that assigns uniform mass on all elements in X.

(ii) We first show that any self-stabilizing set F is a subset of S^∞. Since any outcome not in O cannot be individually rational for any profile of conjectures, we have $F \subseteq S^0 = O$. Suppose $F \subseteq S^n$. Then for any $(\omega, \mu, \mathbf{p}) \in F$, there exists a profile of conjectures $(\nu_j)_{j\in J}$, where $\nu_j \in \Delta^*(\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})})$ for each $j \in J$, such that the three conditions in Definition 1.4 are satisfied. Since $F_{(\mu,\mathbf{p})} \subseteq S^n_{(\mu,\mathbf{p})}$ by the induction hypothesis, the same extension of conjectures in part (i) combined with Lemma 1.2 can be used to show that $(\omega, \mu, \mathbf{p}) \in S^{n+1}$ by Definition 1.5. This means $F \subseteq S^{n+1}$, and induction shows that $F \subseteq \bigcap_{n\geq 0} S^n = S^\infty$.

For the converse, we argue that S^∞ is a self-stabilizing set. Take any outcome $(\omega, \mu, \mathbf{p}) \in S^\infty$. This implies $\omega \in S^\infty_{(\mu,\mathbf{p})}$. Note that $S^0_{(\mu,\mathbf{p})} \subseteq \Omega$ is a finite set, and $(S^n_{(\mu,\mathbf{p})})_{n\geq 0}$ is a decreasing sequence of sets. Therefore, there must exist a finite number n' such that $S^n_{(\mu,\mathbf{p})} = S^\infty_{(\mu,\mathbf{p})}$ for all $n \geq n'$. By Definition 1.5, since $(\omega, \mu, \mathbf{p}) \in S^\infty \subseteq S^{n'+1}$, there exists a profile of conjectures $(\nu_j)_{j\in J}$, where $\nu_j \in \Delta^*(\Pi_j(\omega) \cap S^{n'}_{(\mu,\mathbf{p})})$ for each $j \in J$, such that the three conditions in Definition 1.4 are satisfied. Replacing $S^{n'}_{(\mu,\mathbf{p})}$ with $S^\infty_{(\mu,\mathbf{p})}$ in this statement shows that $(\omega, \mu, \mathbf{p})$ is S^∞-stable. Thus, S^∞ is a self-stabilizing set.

In view of part (i), we conclude that S^∞ is the largest self-stabilizing set.

1.6.2 Proof of Proposition 1.3

We first prove the following lemma. The content of this lemma is that, if F is a self-stabilizing set, then for each $j \in J$ and $\Pi_j \in \mathcal{P}_j$, the F-stability of $(\omega, \mu, \mathbf{p})$ across *all* $\omega \in \Pi_j \cap F_{(\mu,\mathbf{p})}$ can be supported by the *same* CPS on $\Pi_j \cap F_{(\mu,\mathbf{p})}$.

Lemma 1.3. *Suppose F is a self-stabilizing set and $(\mu, \mathbf{p})$ is some allocation. Then for every $j \in J$ and $\Pi_j \in \mathcal{P}_j$ such that $\Pi_j \cap F_{(\mu,\mathbf{p})} \neq \varnothing$, there exists $\nu_{\Pi_j} \in \Delta^*(\Pi_j \cap F_{(\mu,\mathbf{p})})$ that satisfies*

(i) $\mathbb{E}_{\nu_{\Pi_j}}[b_{\mu(j)j} \,|\, \Pi_j \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0$, *and*

(ii) $NB_j(\nu_{\Pi_j}) = \Pi_j \cap F_{(\mu,\mathbf{p})}$ *for all* $j \in J$.

Proof of Lemma 1.3. Fix $j \in J$ and $\Pi_j \in \mathcal{P}_j$ such that $\Pi_j \cap F_{(\mu,\mathbf{p})} \neq \varnothing$. Since F is self-stabilizing, for every $\omega \in \Pi_j \cap F_{(\mu,\mathbf{p})}$, by Definitions 1.4 and 1.6, there exists a self-consistent conjecture $\nu^\omega_j \in \Delta^*(\Pi_j \cap F_{(\mu,\mathbf{p})})$ such that

(i) $\mathbb{E}_{\nu^\omega_j}[b_{\mu(j)j} \,|\, \Pi_j \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0$, and

(ii) $\omega \in NB_j(\nu^\omega_j)$ for all $j \in J$.

Now construct a new conjecture $\nu_{\Pi_j} \in \Delta^*(\Pi_j \cap F_{(\mu,\mathbf{p})})$ as follows: First, write $R_0 = \Pi_j \cap F_{(\mu,\mathbf{p})}$ and let

$$\nu_{\Pi_j}[\cdot \mid R_0] = \frac{1}{|R_0|} \sum_{\omega \in R_0} \nu_j^\omega[\cdot \mid R_0].$$

Iteratively, for every $n \in \mathbb{N}$, if $R_n = R_{n-1} \backslash \text{supp}\nu_{\Pi_j}[\cdot \mid R_{n-1}] \neq \varnothing$, let

$$\nu_{\Pi_j}[\cdot \mid R_n] = \frac{1}{|R_n|} \sum_{\omega \in R_n} \nu_j^\omega[\cdot \mid R_n].$$

Since $\Pi_j \cap F_{(\mu,\mathbf{p})}$ is finite, the iteration stops in finitely many steps. This definition pins down a CPS $\nu_{\Pi_j} \in \Delta^*(\Pi_j \cap F_{(\mu,\mathbf{p})})$.

Since $\mathbb{E}_{\nu_j^\omega}[b_{\mu(j)j} \mid \Pi_j \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0$ for every $\omega \in \Pi_j \cap F_{(\mu,\mathbf{p})}$, and $\nu_{\Pi_j}[\cdot \mid \Pi_j \cap F_{(\mu,\mathbf{p})}]$ is a linear combination of $\nu_j^\omega[\cdot \mid \Pi_j \cap F_{(\mu,\mathbf{p})}]$'s by construction, we have

$$\mathbb{E}_{\nu_{\Pi_j}}[b_{\mu(j)j} \mid \Pi_j \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0.$$

By way of contradiction, suppose at some state $\omega \in \Pi_j \cap F_{(\mu,\mathbf{p})}$ there exists a viable pairwise deviation (i,j,q). Note that $\omega \in D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})} \neq \varnothing$. Let

$$n_j(D_{(i,j,q)}) = \min\{n \in \mathbb{N} : D_{(i,j,q)} \cap R_n \neq \varnothing\}.$$

By construction of $\nu_{\Pi_j}[\cdot \mid R_{n_j(D_{(i,j,q)})}]$, there must exist a state $\omega' \in R_{n_j(D_{(i,j,q)})}$ such that $\nu_j^{\omega'}[D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})} \mid R_{n_j(D_{(i,j,q)})}] > 0$ and

$$\mathbb{E}_{\nu_j^{\omega'}}[b_{ij} \mid D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})}] - q > \mathbb{E}_{\nu_j^{\omega'}}[b_{\mu(j)j} \mid D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j}.$$

The latter inequality implies $D_{(i,j,q)} \cap NB_j(\nu_j^{\omega'}) = \varnothing$. But since $\omega' \in NB_j(\nu_j^{\omega'})$, we know that $R_{n_j(D_{(i,j,q)})} \cap NB_j(\nu_j^{\omega'}) \neq \varnothing$. These facts combined with the self-consistency of $\nu_j^{\omega'}$ indicate that $\nu_j^{\omega'}[D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})} \mid R_{n_j(D_{(i,j,q)})}] = 0$, a contradiction. We have proved the Lemma. □

We now turn to the proof of Proposition 1.3.

$(i) \Rightarrow (ii)$. Suppose $(\omega, \mu, \mathbf{p}) \in S^\infty$ under information structure $\mathcal{I}$. We first construct an expansion $\mathcal{E}$, with trivial partitions, of $\{\mathcal{I}, (\beta_j)_{j \in J}\}$. For each $\omega' \in S^\infty_{(\mu,\mathbf{p})}$, let

$$\Phi^{\omega'} = \{\phi_*^{\omega'}, \phi_\circ^{\omega'}\},$$

and if there is any $\omega' \in \Omega \backslash S^{\infty}_{(\mu,\mathbf{p})}$, simply let $\Phi^{\omega'} = \{\phi^{\omega'}_{\circ}\}$. Define the expanded state space $\Phi = \bigcup_{\omega' \in \Omega} \Phi^{\omega'}$ as the disjoint union of the $\Phi^{\omega'}$'s specified above. For every firm $j \in J$ and every cell $\Pi_j \in \mathcal{P}_j$ such that $\Pi_j \cap S^{\infty}_{(\mu,\mathbf{p})} \neq \varnothing$, let ν_{Π_j} denote the CPS on $\Pi_j \cap S^{\infty}_{(\mu,\mathbf{p})}$ specified in Lemma 1.3. We construct firm j's subjective assessment $\xi_j \in \Delta^*(\Phi)$ as follows: First, for any $A \subseteq \Phi$ and $B \subseteq \Phi$ such that $\phi^{\omega'}_{\circ} \in B$ for some $\omega' \in \Omega$, let

$$\xi_j[A \,|\, B] = \beta_j[\{\omega' \in \Omega : \phi^{\omega'}_{\circ} \in A\} \,|\, \{\omega' \in \Omega : \phi^{\omega'}_{\circ} \in B\}], \tag{1.3}$$

which automatically ensures consistency (see part (iii) of Definition 1.7)). For any $\Pi_j \in \mathcal{P}_j$, if $A \subseteq \Phi$ and $B \subseteq \{\phi \in \Phi : \phi = \phi^{\omega'}_{*} \text{ for some } \omega' \in \Pi_j \cap S^{\infty}_{(\mu,\mathbf{p})}\}$, let

$$\xi_j[A \,|\, B] = \nu_{\Pi_j}[\{\omega' \in \Omega : \phi^{\omega'}_{*} \in A \text{ for some } \omega' \in \Pi_j \cap S^{\infty}_{(\mu,\mathbf{p})}\} \,|\, \{\omega' \in \Omega : \phi^{\omega'}_{*} \in B\}]. \tag{1.4}$$

The CPS $\xi_j \in \Delta^*(\Phi)$ is not uniquely pinned down but we have specified all relevant conditional probability measures. The idea is that, whenever some $\phi^{\omega'}_{\circ}$ is in the conditioning event, the conjecture is in agreement with the exogenous β_j; but when the conditioning event only consists of $\phi^{\omega'}_{*}$'s, the conjecture then follows from the ν_{Π_j}'s. Finally, let $\mathcal{Q}_j = \{\Phi\}$ for all $j \in J$, so we have constructed an expansion $\mathcal{E} = \{\Phi, (\mathcal{Q}_j)_{j \in J}, (\xi_j)_{j \in J}\}$ with trivial partitions.

For each $\omega' \in \Omega$, let $(\mu^{\omega'}, \mathbf{p}^{\omega'})$ be a complete information stable allocation at ω'. Now consider a matching function $M : \Phi \to A$ defined as follows:

$$M(\phi) = \begin{cases} (\mu, \mathbf{p}) & \text{if } \phi = \phi^{\omega'}_{*} \text{ for some } \omega' \in S^{\infty}_{(\mu,\mathbf{p})}, \\ (\mu^{\omega'}, \mathbf{p}^{\omega'}) & \text{if } \phi = \phi^{\omega'}_{\circ} \text{ for some } \omega' \in \Omega \end{cases}$$

Thus, we have $(\mu, \mathbf{p}) = M(\phi^{\omega}_{*})$ where $\phi^{\omega}_{*} \in \Phi^{\omega}$ as desired. It is left to show that M is a stable matching function defined on $\mathcal{E}$. We need to consider two cases.

Case 1: $M(\phi) = (\mu, \mathbf{p})$.

(Individual rationality.) By construction of M, we have

$$a_{i\mu(i)}(\tau(\phi)) + p_{i\mu(i)} \geq 0, \text{ for all } i \in I.$$

Moreover, all firms can infer the following event from observing the allocation $(\mu, \mathbf{p})$,

$$M^{-1}(\mu, \mathbf{p}) = \{\phi^{\omega'}_{*} : \omega' \in S^{\infty}_{(\mu,\mathbf{p})}\} \cup \{\phi^{\omega'}_{\circ} : (\mu^{\omega'}, \mathbf{p}^{\omega'}) = (\mu, \mathbf{p})\},$$

which is a union of two disjoint sets. To reduce notation, we let

$$G_1 = \{\phi_*^{\omega'} : \omega' \in S^{\infty}_{(\mu,\mathbf{p})}\}, \quad \text{and} \quad G_2 = \{\phi_\circ^{\omega'} : (\mu^{\omega'}, \mathbf{p}^{\omega'}) = (\mu, \mathbf{p})\}.$$

Notice that one of $\Pi_j(\phi) \cap G_1$ and $\Pi_j(\phi) \cap G_2$ may be empty (but not both). If $\Pi_j(\phi) \cap G_2 \neq \varnothing$, then

$$\begin{aligned}
&\mathbb{E}_{\xi_j}[b_{\mu(j)j} \,|\, \Pi_j(\phi) \cap (G_1 \cup G_2)] - p_{\mu(j)j} \\
&= \sum_{\phi' \in \Pi_j(\phi) \cap G_2} \xi_j[\phi' \,|\, \Pi_j(\phi) \cap (G_1 \cup G_2)] \cdot [b_{\mu(j)j}(\tau(\phi')) - p_{\mu(j)j}] \\
&\geq 0,
\end{aligned}$$

where the equality is by construction of ξ_j (equation (1.3)), and the inequality comes from the fact that $(\mu, \mathbf{p})$ is a complete information stable allocation at ϕ' for all $\phi' \in G_2$. On the other hand, if $\Pi_j(\phi) \cap G_2 = \varnothing$, then

$$\begin{aligned}
&\mathbb{E}_{\xi_j}[b_{\mu(j)j} \,|\, \Pi_j(\phi) \cap (G_1 \cup G_2)] - p_{\mu(j)j} \\
&= \mathbb{E}_{\nu_{\Pi_j(\phi)}}[b_{\mu(j)j} \,|\, \{\omega' \in \Omega : \phi_*^{\omega'} \in \Pi_j(\phi) \cap G_1\}] - p_{\mu(j)j} \\
&\geq 0,
\end{aligned}$$

where the equality is again by construction of ξ_j (equation (1.4)), and the inequality is a consequence of Lemma 1.3. Therefore, we conclude that

$$\mathbb{E}_{\xi_j}[b_{\mu(j)j} \,|\, \Pi_j(\phi) \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j} \geq 0, \text{ for all } j \in J,$$

and hence the matching function M satisfies individual rationality (for those ϕ in Case 1).

(No blocking.) For any pairwise deviation (i, j, q) such that

$$a_{ij}(\tau(\phi)) + q > a_{i\mu(i)}(\tau(\phi)) + p_{i\mu(i)},$$

let $D^{\Phi}_{(i,j,q)} = \{\phi' \in \Pi_j(\phi) : a_{ij}(\tau(\phi')) + q > a_{i\mu(i)}(\tau(\phi')) + p_{i\mu(i)}\}$. We continue to use the

notation $M^{-1}(\mu, \mathbf{p}) = G_1 \cup G_2$. If $D^{\Phi}_{(i,j,q)} \cap G_2 \neq \varnothing$, then

$$\begin{aligned}\mathbb{E}_{\xi_j}[b_{ij} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - q &= \sum_{\phi' \in G_2} \xi_j[\phi' \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] \cdot [b_{ij}(\tau(\phi')) - q] \\ &\leq \sum_{\phi' \in G_2} \xi_j[\phi' \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] \cdot [b_{\mu(j)j}(\tau(\phi')) - p_{\mu(j)j}] \\ &= \mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j},\end{aligned}$$

where the two equalities are implied by construction of ξ_j (equation (1.3)), and the inequality comes from the fact that $(\mu, \mathbf{p})$ is a complete information stable allocation at ϕ' for all $\phi' \in G_2$. On the other hand, if $D^{\Phi}_{(i,j,q)} \cap G_2 = \varnothing$, then

$$\begin{aligned}\mathbb{E}_{\xi_j}[b_{ij} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - q &= \mathbb{E}_{\nu_{\Pi_j(\phi)}}[b_{ij} \mid \{\omega' \in S^{\infty}_{(\mu,\mathbf{p})} : \phi^{\omega'}_{*} \in D^{\Phi}_{(i,j,q)}\}] - q \\ &\leq \mathbb{E}_{\nu_{\Pi_j(\phi)}}[b_{\mu(j)j} \mid \{\omega' \in S^{\infty}_{(\mu,\mathbf{p})} : \phi^{\omega'}_{*} \in D^{\Phi}_{(i,j,q)}\}] - p_{\mu(j)j} \\ &= \mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j},\end{aligned}$$

where the two equalities are again by construction of ξ_j (equation (1.4)), and the inequality is a consequence of Lemma 1.3. The argument above implies that

$$\mathbb{E}_{\xi_j}[b_{ij} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - q \leq \mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j},$$

which means the matching function M is not blocked (for those ϕ in Case 1).

Case 2: $M(\phi) = (\mu^{\omega'}, \mathbf{p}^{\omega'}) \neq (\mu, \mathbf{p})$.

In this case, the inference all firms can make from observing the allocation is

$$M^{-1}(\mu^{\omega'}, \mathbf{p}^{\omega'}) = \{\phi^{\omega''}_{\circ} : (\mu^{\omega''}, \mathbf{p}^{\omega''}) = (\mu^{\omega'}, \mathbf{p}^{\omega'})\},$$

which is analogous to the event G_2 defined above. It can then be shown that the matching function M is individually rational and not blocked for those ϕ in Case 2. The argument closely follows that in the previous case and is hence omitted.

We conclude that the matching function M defined on $\mathcal{E}$ is stable.

$(ii) \Rightarrow (iii)$. Trivial.

$(iii) \Rightarrow (i)$. Let $\mathcal{E}$ be an arbitrary expansion of the information environment $\{\mathcal{I}, (\beta_j)_{j \in J}\}$. Suppose M is a stable matching function defined on $\mathcal{E}$. Define a set of

outcomes as

$$F = \{(\omega, \mu, \mathbf{p}) \in O : \exists \phi \in \Phi^\omega \text{ such that } M(\phi) = (\mu, \mathbf{p})\}.$$

We now show that F is a self-stabilizing set. Fixing $(\omega, \mu, \mathbf{p}) \in F$, let $\phi \in \Phi$ be such that $\phi \in \Phi^\omega$ and $M(\phi) = (\mu, \mathbf{p})$. For every firm $j \in J$, define a conjecture $\nu_j \in \Delta^*(\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})})$ as follows: Denote $E_j^\Phi = \Pi_j(\phi) \cap \Lambda_j(\phi) \cap M^{-1}(\mu, \mathbf{p})$ and $E_j = \{\omega' \in \Pi_j(\omega) \cap F_{(\mu,\mathbf{p})} : \exists \phi' \in E_j^\Phi \text{ such that } \phi' \in \Phi^{\omega'}\}$. For every conditioning event $C \subseteq \Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}$, if $C \cap E_j \neq \varnothing$, let

$$\nu_j[B \,|\, C] = \xi_j[\{\phi' \in E_j^\Phi : \exists \omega' \in B \text{ s.t. } \phi' \in \Phi^{\omega'}\} \,|\, \{\phi' \in E_j^\Phi : \exists \omega' \in C \text{ s.t. } \phi' \in \Phi^{\omega'}\}] \tag{1.5}$$

for all $B \subseteq C$; if $C \cap E_j = \varnothing$, let $\nu_j[\cdot \,|\, C] = \text{Unif}(C)$. This pins down a CPS ν_j defined on $\Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}$. Note that $\nu_j[E_j \,|\, C] = 1$ whenever $C \cap E_j \neq \varnothing$. Take any state $\omega' \in E_j$ and a pairwise deviation (i, j, q) such that

$$a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}.$$

This means $\omega' \in D_{(i,j,q)} \cap E_j \neq \varnothing$. But then

$$\begin{aligned}
\mathbb{E}_{\nu_j}[b_{ij} \,|\, D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})}] - q &= \mathbb{E}_{\xi_j}[b_{ij} \,|\, D^\Phi_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - q \\
&> \mathbb{E}_{\xi_j}[b_{\mu(j)j} \,|\, D^\Phi_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p})] - p_{\mu(j)j} \\
&= \mathbb{E}_{\nu_j}[b_{\mu(j)j} \,|\, D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j},
\end{aligned} \tag{1.6}$$

where

$$\begin{aligned}
D^\Phi_{(i,j,q)} \cap M^{-1}(\mu, \mathbf{p}) &= \{\phi' \in E_j^\Phi : a_{ij}(\tau(\phi')) + q > a_{i\mu(i)}(\tau(\phi')) + p_{i\mu(i)}\} \\
&= \{\phi' \in E_j^\Phi : \exists \omega' \in D_{(i,j,q)} \cap F_{(\mu,\mathbf{p})} \text{ s.t. } \phi' \in \Phi^{\omega'}\}.
\end{aligned}$$

The two equalities in (1.6) come from the definition of ν_j (equation (1.5)), while the inequality is due to the stability of matching function M. This implies $\omega' \in NB_j(\nu_j)$. Since $\omega' \in E_j$ was arbitrary, we have $\omega \in E_j \subseteq NB_j(\nu_j)$.

Finally, we need to verify individual rationality of the allocation $(\mu, \mathbf{p})$ at ω. Since

M is a stable matching function and $M(\phi) = (\mu, \mathbf{p})$ where $\phi \in \Phi^\omega$, we have

$$\begin{aligned} a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} &\geq 0, \text{ for all } i \in I \text{ and} \\ \mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid \Pi_j(\omega) \cap F_{(\mu,\mathbf{p})}] - p_{\mu(j)j} &= \mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid E_j^\Phi] - p_{\mu(j)j} \\ &\geq 0, \text{ for all } j \in J, \end{aligned}$$

where the equality above is by definition of ν_j, and the inequalities are due to the stability of matching function M. In view of Lemma 1.2, the outcome $(\omega, \mu, \mathbf{p})$ is F-stable. Since $(\omega, \mu, \mathbf{p}) \in F$ was arbitrary, this implies F is a self-stablizing set. By Proposition 1.1, we have $F \subseteq S^\infty$, which completes the proof.

1.6.3 Proof of Proposition 1.4

$(i) \Rightarrow (ii)$. Suppose $\tilde{M}$ is a stable stochastic matching function defined on $\{\mathcal{I}, \beta\}$. For each $\omega \in \Omega$, let

$$\Phi^\omega = \{(\mu, \mathbf{p}) : \tilde{M}(\omega)[(\mu, \mathbf{p})] > 0\},$$

and write $\phi^\omega_{(\mu,\mathbf{p})}$ for the element in Φ^ω that corresponds to $(\mu, \mathbf{p})$. For each $\phi^\omega_{(\mu,\mathbf{p})} \in \Phi^\omega$ and each $j \in J$, we define

$$\xi_j[\phi^\omega_{(\mu,\mathbf{p})}] = \xi[\phi^\omega_{(\mu,\mathbf{p})}] = \beta[\omega] \cdot \tilde{M}(\omega)[(\mu, \mathbf{p})].$$

Finally, let $\mathcal{Q}_j = \{\Phi\}$ for all $j \in J$. So we have constructed a common prior expansion $\mathcal{E} = \{\Phi, (\mathcal{Q}_j)_{j\in J}, \xi\}$ with trivial partitions. Define a matching function M on $\mathcal{E}$ as follows:

$$M : \phi^\omega_{(\mu,\mathbf{p})} \mapsto (\mu, \mathbf{p}).$$

By construction, we have $\Phi^\omega \cap M^{-1}(\mu, \mathbf{p}) = \{\phi^\omega_{(\mu,\mathbf{p})}\}$, so the expansion $\mathcal{E}$ and the matching function M induce $\tilde{M}$. It remains to show that M is a stable matching function. But observe that for each $\omega' \in \Pi_j(\omega)$,

$$\begin{aligned} \xi_j[\phi^{\omega'}_{(\mu,\mathbf{p})} \mid \Pi_j(\phi^{\omega'}_{(\mu,\mathbf{p})}) \cap M^{-1}(\mu, \mathbf{p})] &= \frac{\xi_j[\phi^{\omega'}_{(\mu,\mathbf{p})}]}{\xi_j[\Pi_j(\phi^{\omega'}_{(\mu,\mathbf{p})}) \cap M^{-1}(\mu, \mathbf{p})]} \\ &= \frac{\beta_j[\omega']\tilde{M}(\omega')[(\mu, \mathbf{p})]}{\sum_{\omega'' \in \Pi_j(\omega)} \beta_j[\omega'']\tilde{M}(\omega'')[(\mu, \mathbf{p})]} \\ &= \beta_j^{(\omega,\mu,\mathbf{p})}[\omega']. \end{aligned}$$

This means that individual rationality of the stochastic matching function $\tilde{M}$ implies individual rationality of M. Similarly, for any pairwise deviation (i, j, q), we have

$$\begin{aligned}\xi_j[\phi^{\omega'}_{(\mu,\mathbf{p})} \mid D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu,\mathbf{p})] &= \frac{\xi_j[\phi^{\omega'}_{(\mu,\mathbf{p})}]}{\xi_j[D^{\Phi}_{(i,j,q)} \cap M^{-1}(\mu,\mathbf{p})]} \\ &= \frac{\beta_j[\omega']\tilde{M}(\omega')[(\mu,\mathbf{p})]}{\sum_{\omega'' \in D_{(i,j,q)}} \beta_j[\omega'']\tilde{M}(\omega'')[(\mu,\mathbf{p})]} \\ &= \beta_j^{(\omega,\mu,\mathbf{p})}[\omega' \mid D_{(i,j,q)}].\end{aligned}$$

Therefore, no blocking of $\tilde{M}$ also implies no blocking of M. Hence, we can conclude that M is a stable matching function defined on $\mathcal{E}$.

$(ii) \Rightarrow (iii)$. Trivial.

$(iii) \Rightarrow (i)$. Suppose $\mathcal{E} = \{\Phi, (\mathcal{Q}_j)_{j\in J}, \xi\}$ is a common prior expansion of $\{\mathcal{I}, \beta\}$ and M is a stable matching function defined on $\mathcal{E}$. Let $\tilde{M}$ be the stochastic matching function induced by $\mathcal{E}$ and M. We need to verify that $\tilde{M}$ is stable.

(Individual rationality.) For every $\omega \in \Omega$ and allocation $(\mu, \mathbf{p})$ such that $\tilde{M}(\omega)[(\mu, \mathbf{p})] > 0$, there exists $\phi \in \Phi^\omega \cap M^{-1}(\mu, \mathbf{p})$. Since M is stable, we have

$$a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} = a_{i\mu(i)}(\tau(\phi)) + p_{i\mu(i)} \geq 0$$

for all $i \in I$. Now fix a firm $j \in J$. Because M is stable, for every $\phi \in \Phi$ and allocation $(\mu, \mathbf{p}) = M(\phi)$, we have

$$\mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid \Pi_j(\phi) \cap \Lambda_j(\phi) \cap M^{-1}(\mu,\mathbf{p})] - p_{\mu(j)j} \geq 0.$$

By law of iterated expectations, this implies

$$\mathbb{E}_{\xi_j}[b_{\mu(j)j} \mid \Pi_j(\phi) \cap M^{-1}(\mu,\mathbf{p})] - p_{\mu(j)j} \geq 0.$$

Now observe that for each $\omega' \in \Pi_j(\omega)$,

$$\begin{aligned}\beta_j^{(\omega,\mu,\mathbf{p})}[\omega'] &= \frac{\beta_j[\omega']\tilde{M}(\omega')[(\mu,\mathbf{p})]}{\sum_{\omega''\in\Pi_j(\omega)}\beta_j[\omega'']\tilde{M}(\omega'')[(\mu,\mathbf{p})]} \\ &= \frac{\sum_{\phi'\in\Phi^{\omega'}\cap M^{-1}(\mu,\mathbf{p})}\xi[\phi']}{\sum_{\phi'\in\Pi_j(\phi)\cap M^{-1}(\mu,\mathbf{p})}\xi[\phi']} \\ &= \xi_j[\Phi^{\omega'}\cap M^{-1}(\mu,\mathbf{p}) \,|\, \Pi_j(\phi)\cap M^{-1}(\mu,\mathbf{p})].\end{aligned}$$

Using the fact that $b_{\mu(j)j}(\tau(\cdot))$ is constant on $\Phi^{\omega'}$ for each $\omega' \in \Omega$, we conclude

$$\mathbb{E}_{\beta_j^{(\omega,\mu,\mathbf{p})}}[b_{\mu(j)j}] - p_{\mu(j)j} = \mathbb{E}_{\xi_j}[b_{\mu(j)j} \,|\, \Pi_j(\phi)\cap M^{-1}(\mu,\mathbf{p})] - p_{\mu(j)j} \geq 0.$$

Therefore, the stochastic matching function $\tilde{M}$ satisfies individual rationality.

(No blocking.) The argument closely follows the one above for individual rationality, except that now we need to take care of the conditioning event $D_{(i,j,q)}$ for each firm j. We omit the details of this part.

1.6.4 The Universal Epistemic State Space

In this section, we first construct a space $H_j(\Pi_j)$ that collects all possible conditional beliefs and higher order conditional beliefs about the uncertain state consistent with firm j's information Π_j. Due to the non-product structure of information received by the firms, extra care needs to be taken when doing so.[41]

For each firm $j = 1, 2$, and each piece of information $\Pi_j \in \mathcal{P}_j$, we write

$$Y_j^0(\Pi_j) = \Pi_j \quad \text{and} \quad \mathcal{C}_j^0(\Pi_j) = 2^{\Pi_j}\backslash\{\varnothing\},$$

where $Y_j^0(\Pi_j)$ is the space of firm j's first-order uncertainty consistent with Π_j, and $\mathcal{C}_j^0(\Pi_j)$ is the set of conditioning events. Therefore,

$$Z_j^1(\Pi_j) = \Delta^{\mathcal{C}_j^0(\Pi_j)}(Y_j^0(\Pi_j)) = \Delta^*(\Pi_j)$$

is the set of firm j's *first-order CPSs consistent with* Π_j.

[41] Our construction is similar to the one in Penta and Zuazo-Garin (2022). They study the implication of higher order uncertainty about observability of the opponent's action in two-player static games, where information has a non-product partition structure.

Recursively, for $n \geq 1$, $j = 1, 2$, $\Pi_j \in \mathcal{P}_j$, and $k \neq j$, the space of firm j's $(n+1)$th-order uncertainty consistent with Π_j is

$$Y_j^n(\Pi_j) = \left\{(\omega, \delta_k^1, \ldots, \delta_k^n) : \omega \in \Pi_j \text{ and } \delta_k^\ell \in Z_k^\ell(\Pi_k(\omega)) \ \forall \ell = 1, \ldots, n\right\},$$

and the set of corresponding conditioning events is

$$\mathcal{C}_j^n(\Pi_j) = \left\{\text{cyl}^n(C) \subseteq Y_j^n(\Pi_j) : C \in 2^{\Pi_j} \backslash \{\varnothing\}\right\},$$

where $\text{cyl}_j^n(C) = \left\{(\omega, \delta_k^1, \ldots, \delta_k^n)) \in Y_j^n(\Pi_j) : \omega \in C\right\}$ is a cylinder in $Y_j^n(\Pi_j)$ generated by the event $C \in 2^{\Pi_j} \backslash \{\varnothing\}$. With these definitions, the set of firm j's *coherent* belief hierarchies up to order $n+1$ is

$$Z_j^{n+1}(\Pi_j) = \left\{ \begin{array}{l} (\delta_j^1, \ldots, \delta_j^{n+1}) : \\ \quad \text{(i) } (\delta_j^1, \ldots, \delta_j^n) \in Z_j^n(\Pi_j) \text{ and } \delta_j^{n+1} \in \Delta^{\mathcal{C}_j^n(\Pi_j)}(Y_j^n(\Pi_j)); \\ \quad \text{(ii) } \text{marg}_{Z_j^n(\Pi_j)} \delta_j^{n+1}[\cdot \,|\, \text{cyl}_j^n(C)] = \delta_j^n[\cdot \,|\, \text{cyl}_j^{n-1}(C)] \ \forall C \in 2^{\Pi_j} \backslash \{\varnothing\}. \end{array} \right\}$$

The set of *collectively coherent conditional belief hierarchies* of firm j consistent with information Π_j is therefore

$$H_j(\Pi_j) = \left\{ \begin{array}{c} \delta_j = (\delta_j^1, \delta_j^2, \ldots) \in Z_j^0(\Pi_j) \times \times_{n \in \mathbb{N}} \Delta^{\mathcal{C}_j^n(\Pi_j)}(Y_j^n(\Pi_j)) : \\ (\delta_j^1, \ldots, \delta_j^n) \in Z_j^n(\Pi_j), \ \forall n \in \mathbb{N}. \end{array} \right\}$$

For any Polish space X and a collection $\mathcal{C}$ of nonempty measurable subsets, the space $\Delta^{\mathcal{C}}(X)$ is endowed with the relative topology inherited from $[\Delta(X)]^{\mathcal{C}}$ with the product topology. Moreover, we endow $H_j(\Pi_j)$ with the product topology for each $j \in J$ and $\Pi_j \in \mathcal{P}_j$. Since Ω is finite, each $H_j(\Pi_j)$ is compact. Generalizing the analysis from Battigalli and Siniscalchi (1999), it can be shown that there exists a homeomorphism[42]

$$g_{\Pi_j} : H_j(\Pi_j) \to \Delta^{\mathcal{C}_j^\infty(\Pi_j)}\left(Y_j^\infty(\Pi_j)\right)$$

[42] Analogous to the previous definition, we write

$$Y_j^\infty(\Pi_j) = \left\{(\omega, \delta_k) : \omega \in \Pi_j \text{ and } \delta_k \in H_k(\Pi_k(\omega))\right\},$$

and define

$$\mathcal{C}_j^\infty(\Pi_j) = \left\{\text{cyl}^\infty(C) \subseteq Y_j^\infty(\Pi_j) : C \in 2^{\Pi_j} \backslash \{\varnothing\}\right\},$$

where $\text{cyl}^\infty(C) = \left\{(\omega, \delta_k) \in Y_j^\infty(\Pi_j) : \omega \in C\right\}$ is a cylinder of C in $Y_j^\infty(\Pi_j)$.

that preserves all conditional beliefs; that is, for every $\delta_j \in H_j(\Pi_j)$, we have

$$\delta_j^n \left[E \,\middle|\, \text{cyl}_j^{n-1}(C) \right] = g_{\Pi_j}(\delta_j) \left[(\text{proj})^{-1}_{Y_j^{n-1}(\Pi_j)}(E) \,\middle|\, \text{cyl}_j^{\infty}(C) \right]$$

for any measurable $E \subseteq Y_j^{n-1}(\Pi_j)$, $C \in 2^{\Pi_j} \backslash \{\varnothing\}$, and $n \in \mathbb{N}$.

Now define the *universal epistemic state space* as

$$W = \bigcup_{\omega \in \Omega} \left(\{\omega\} \times A \times H_1(\Pi_1(\omega)) \times H_2(\Pi_2(\omega)) \right).$$

1.6.5 Proof of Lemma 1.1

We claim that the set PR is closed in W, and hence measurable. To prove this, take any sequence $\{(\omega, \mu, \mathbf{p}^{(n)}, \delta_1^{(n)}, \delta_2^{(n)})\}_{n \in \mathbb{N}} \subseteq PR$ such that $\mathbf{p}^{(n)} \to \mathbf{p}$ and $\delta_j^{(n)} \to \delta_j$ for each $j = 1, 2$ as $n \to \infty$. It suffices to show that $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in PR$ since the state space Ω and the set of possible matchings are finite. Conditions (i) and (ii) in Definition 1.10 can be seen by simply taking limits, so we focus on condition (iii) in what follows.

Towards a contradiction, suppose condition (iii) is violated for $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2)$, which means there exists a pairwise deviation (i, j, q) such that

$$\begin{aligned} a_{ij}(\tau(\omega)) + q &> a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)}, \\ \mathbb{E}_{\delta_j^1}[b_{ij} \,|\, D_{(i,j,q)}] - q &> \mathbb{E}_{\delta_j^1}[b_{\mu(j)j} \,|\, D_{(i,j,q)}] - p_{\mu(j)j}, \end{aligned}$$

where $D_{(i,j,q)} = \{\omega' \in \Pi_j(\omega) : a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}\}$. Since $\Pi_j(\omega)$ is finite, the set $D_{(i,j,q)}$ and both inequalities above remain the same if we replace q with q' sufficiently close to q. Suppose $|q - q'| = \varepsilon > 0$. Let n' be such that

$$|p_{i\mu(i)} - p_{i\mu(i)}^{(n)}| < \varepsilon, \text{ for all } n \geq n'.$$

This means the set $D_{(i,j,q)}$ and both inequalities still remain the same if we replace $p_{i\mu(i)}$ with $p_{i\mu(i)}^{(n)}$ and δ_j^1 with $\delta_j^{(n),1}$ for all $n \geq n'$, where q is assumed to be the larger one of q and q'. Since each $(\omega, \mu, \mathbf{p}^{(n)}, \delta_1^{(n)}, \delta_2^{(n)}) \in PR$, we must have

$$\mathbb{E}_{\delta_j^{(n),1}}[b_{ij} \,|\, D_{(i,j,q)}] - q \leq \mathbb{E}_{\delta_j^{(n),1}}[b_{\mu(j)j} \,|\, D_{(i,j,q)}] - p_{\mu(j)j}^{(n)}, \quad \forall n \geq n'.$$

Now taking $n \to \infty$ yields a contradiction.

1.6.6 Proof of Proposition 1.5

We split the statement of the proposition into two lemmas and prove them in order.

Lemma 1.4. $S^\infty \supseteq \text{proj}_O\text{PRCSBPR}$.

Proof. We shall prove that $S^n \supseteq \text{proj}_O PR^{n+1}$ for all $n \geq 0$ by induction.

(Base step.) Since $S^0 = O$ by definition, we have $S^0 \supseteq \text{proj}_O PR^1$.

(Induction step.) Suppose $S^{n-1} \supseteq \text{proj}_O PR^n$. If $PR^n \cap \text{SB}(PR^n) = PR^{n+1} = \varnothing$, we are done. Otherwise, take any epistemic state $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in PR^{n+1}$. Now for each $j = 1, 2$, define a conjecture $\nu_j \in \Delta^*(\Pi_j(\omega) \cap S^{n-1}_{(\mu,\mathbf{p})})$ as follows: For every $C \subseteq \Pi_j(\omega) \cap S^{n-1}_{(\mu,\mathbf{p})}$ such that $C \cap \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \neq \varnothing$, let $\nu_j[\cdot \,|\, C] = \delta^1_j[\cdot \,|\, C]$. Note that in this case,

$$\delta^1_j \left[\text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \,\middle|\, C \right] = g_{\Pi_j(\omega)}(\delta_j) \left[PR^n_{(\mu,\mathbf{p},\delta_j)} \,\middle|\, \text{cyl}^\infty_j(C) \right] = 1$$

since $g_{\Pi_j(\omega)}$ is a belief-preserving mapping and $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in \text{SB}(PR^n)$. Moreover, if $C \cap \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} = \varnothing$, let $\nu_j[\cdot \,|\, C] = \text{Unif}(C)$. This construction ensures that $\nu_j \in \Delta^*(\Pi_j(\omega) \cap S^{n-1}_{(\mu,\mathbf{p})})$ is a well-defined CPS for each $j = 1, 2$.

Since $PR^{n+1} \subseteq PR$, we know from the definition of PR that for all $i \in I$ and $j = 1, 2$,

$$\begin{aligned} a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} &\geq 0, \text{ and} \\ \mathbb{E}_{\delta^1_j}[b_{\mu(j)j} \,|\, \Pi_j(\omega)] - p_{\mu(j)j} &\geq 0. \end{aligned}$$

But since $\omega \in \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \subseteq S^{n-1}_{(\mu,\mathbf{p})}$, by construction of ν_j, we have

$$\mathbb{E}_{\nu_j}[b_{\mu(j)j} \,|\, \Pi_j(\omega) \cap S^{n-1}_{(\mu,\mathbf{p})}] - p_{\mu(j)j} \geq 0.$$

This implies that the pair of conjectures (ν_1, ν_2) satisfies individual rationality in Definition 1.4.

We next show that $\text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \subseteq NB_j(\nu_j)$ for each $j = 1, 2$. To see this, take any $\omega' \in \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)}$ and suppose there exists a pairwise deviation (i, j, q) such that

$$a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}.$$

Since $PR^n \subseteq PR$, the first-order CPS δ_j^1 must satisfy

$$\mathbb{E}_{\delta_j^1}[b_{ij} \mid D_{(i,j,q)}] - q \leq \mathbb{E}_{\delta_j^1}[b_{\mu(j)j} \mid D_{(i,j,q)}] - p_{\mu(j)j}.$$

Because $\omega' \in D_{(i,j,q)} \cap \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \subseteq D_{(i,j,q)} \cap S^{n-1}_{(\mu,\mathbf{p})}$, we know that $\nu_j[\cdot \mid D_{(i,j,q)} \cap S^{n-1}_{(\mu,\mathbf{p})}] = \delta_j^1[\cdot \mid D_{(i,j,q)}]$ by construction of ν_j. Therefore, the inequality above becomes

$$\mathbb{E}_{\nu_j}[b_{ij} \mid D_{(i,j,q)} \cap S^{n-1}_{(\mu,\mathbf{p})}] - q \leq \mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid D_{(i,j,q)} \cap S^{n-1}_{(\mu,\mathbf{p})}] - p_{\mu(j)j},$$

which implies $\omega' \in NB_j(\nu_j)$. Since $\omega' \in \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)}$ was arbitrary, we have established that $\text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \subseteq NB_j(\nu_j)$.

Observe that (i) $\omega \in \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \subseteq NB_j(\nu_j)$, and (ii) $\nu_j[\text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \mid C] = 1$ whenever $C \cap \text{proj}_{\Pi_j(\omega)} PR^n_{(\mu,\mathbf{p},\delta_j)} \neq \varnothing$. We can invoke Lemma 1.2 and conclude that $(\omega, \mu, \mathbf{p})$ is S^{n-1}-stable. By definition, we have $(\omega, \mu, \mathbf{p}) \in S^n$. This in turn implies $S^n \supseteq \text{proj}_O PR^{n+1}$. If $(\omega, \mu, \mathbf{p}) \in \text{proj}_O \text{PRCSBPR}$, then $(\omega, \mu, \mathbf{p}) \in S^n$ for all $n \geq 0$, which means $(\omega, \mu, \mathbf{p}) \in S^\infty$. We have proved the lemma. □

Lemma 1.5. $S^\infty \subseteq \text{proj}_O \text{PRCSBPR}$.

Proof. We first prove that $S^n \subseteq \text{proj}_O PR^n$ for all $n \geq 1$ by induction.

(Base step.) Take any $(\omega, \mu, \mathbf{p}) \in S^1$. By definition, there exists a pair of conjectures (ν_1, ν_2), where $\nu_j \in \Delta^*(\Pi_j(\omega))$, that supports the S^0-stability of $(\omega, \mu, \mathbf{p})$. For each $j = 1, 2$, since the mapping $g_{\Pi_j(\omega)}$ is onto, there exists $\delta_j \in H_j(\Pi_j(\omega_j))$ such that $\delta_j^1 = \nu_j$. Comparing Definition 1.10 with Definition 1.4, we conclude that $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in PR$. Therefore, $S^1 \subseteq \text{proj}_O PR^1$.

(Induction step.) Suppose $S^\ell \subseteq \text{proj}_O PR^\ell$ for all $\ell = 1, \ldots, n-1$. Take any $(\omega, \mu, \mathbf{p}) \in S^n$. By definition, there exists a pair of self-consistent conjectures (ν_1, ν_2), where $\nu_j \in \Delta^*(\Pi_j(\omega) \cap S^{n-1}_{(\mu,\mathbf{p})})$ for each $j \in J$, such that the following two conditions are satisfied:

(i) For all $i \in I$ and $j = 1, 2$,

$$\begin{aligned} a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)} &\geq 0 \\ \mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid \Pi_j(\omega) \cap S^{n-1}_{(\mu,\mathbf{p})}] - p_{\mu(j)j} &\geq 0. \end{aligned}$$

(ii) There does not exist a pairwise deviation (i,j,q) such that

$$a_{ij}(\tau(\omega)) + q > a_{i\mu(i)}(\tau(\omega)) + p_{i\mu(i)},$$
$$\mathbb{E}_{\nu_j}[b_{ij} \mid D_{(i,j,q)} \cap S^{n-1}_{(\mu,\mathbf{p})}] - q > \mathbb{E}_{\nu_j}[b_{\mu(j)j} \mid D_{(i,j,q)} \cap S^{n-1}_{(\mu,\mathbf{p})}] - p_{\mu(j)j},$$

where $D_{(i,j,q)} = \{\omega' \in \Pi_j(\omega) : a_{ij}(\tau(\omega')) + q > a_{i\mu(i)}(\tau(\omega')) + p_{i\mu(i)}\}$.

For each $j = 1, 2$, extend ν_j to a CPS $\bar{\nu}_j \in \Delta^*(\Pi_j(\omega))$ as follows: For any conditioning event $C \subseteq \Pi_j(\omega)$, write

$$m_j(C) = \max\left\{\ell = 1, \ldots, n-1 : C \cap S^{\ell}_{(\mu,\mathbf{p})} \neq \varnothing\right\}.$$

If $m_j(C) = n-1$, let $\bar{\nu}_j[E \,|\, C] = \nu_j[E \cap S^{n-1}_{(\mu,\mathbf{p})} \,|\, C \cap S^{n-1}_{(\mu,\mathbf{p})}]$; if $m_j(C) < n-1$, let $\bar{\nu}_j[E \,|\, C] = \text{Unif}(C \cap S^{m_j(C)}_{(\mu,\mathbf{p})})$. Now by the induction hypothesis, we can construct from $\bar{\nu}_j$ a CPS

$$\eta_j \in \Delta^{\mathcal{C}^{\infty}_j(\Pi_j)}\left(Y^{\infty}_j(\Pi_j(\omega))\right)$$

that satisfies the following:

(a) For every $C \subseteq \Pi_j(\omega)$ such that $C \cap S^{n-1}_{(\mu,\mathbf{p})} \neq \varnothing$ and $E \subseteq \Pi_j(\omega)$, we have

$$\eta_j\left[\text{proj}^{-1}_{\Pi_j(\omega)} E \,\middle|\, \text{cyl}^{\infty}(C)\right] = \nu_j\left[E \cap S^{n-1}_{(\mu,\mathbf{p})} \,\middle|\, C \cap S^{n-1}_{(\mu,\mathbf{p})}\right];$$

(b) The support of η_j is consistent with the highest degree of PR^{ℓ}, i.e. whenever $C \cap \text{proj}_{\Pi_j(\omega)} PR^{\ell} \neq \varnothing$ for any $\ell = 1, \ldots, n-1$, we have

$$\text{supp}\eta_j[\cdot \,|\, \text{cyl}^{\infty}(C)] \subseteq \text{proj}_{Y^{\infty}_j(\Pi_j(\omega))} PR^{\ell}.$$

Because $g_{\Pi_j(\omega)}$ is onto, we can now define a conditional belief hierarchy $\delta_j = g^{-1}_{\Pi_j(\omega)}(\eta_j)$ for each $j = 1, 2$. Since $g_{\Pi_j(\omega)}$ is belief-preserving, property (a) combined with conditions (i) and (ii) of ν_j implies the epistemic state $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in PR$. Moreover, property (b) implies that for each j and any $\ell = 1, \ldots, n-1$, we have $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in \text{SB}_j(PR^{\ell})$. Combining these two implications, we obtain

$$(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in PR \cap \bigcap_{j=1,2}\left\{\bigcap_{\ell=1}^{n-1} \text{SB}_j(PR^{\ell})\right\} = PR^{n-1} \cap \text{SB}(PR^{n-1}) = PR^n.$$

Therefore, we get $S^n \subseteq \text{proj}_O PR^n$ as desired.

Now take any $(\omega, \mu, \mathbf{p}) \in S^\infty$. For each $n \in \mathbb{N}$, define

$$U^n_{(\omega,\mu,\mathbf{p})} = PR^n \cap (\{(\omega, \mu, \mathbf{p})\} \times H_1(\Pi_1(\omega)) \times H_2(\Pi_2(\omega))\}).$$

Since $(\omega, \mu, \mathbf{p}) \in \text{proj}_O PR^n$ for all $n \in \mathbb{N}$, we know that $\{U^n_{(\omega,\mu,\mathbf{p})}\}_{n\in\mathbb{N}}$ is a decreasing sequence of *nonempty* sets. This means the collection $\{U^n_{(\omega,\mu,\mathbf{p})}\}_{n\in\mathbb{N}}$ has the finite intersection property. But the space W is compact, which in turn implies that there exists some pair (δ_1, δ_2) such that $(\omega, \mu, \mathbf{p}, \delta_1, \delta_2) \in \bigcap_{n\in\mathbb{N}} U^n_{(\omega,\mu,\mathbf{p})} \subseteq \text{PRCSBPR}$. Hence, we obtain $S^\infty \subseteq \text{proj}_O\text{PRCSBPR}$. □

1.7 Bibliography

Alston, M. (2020): "On the Non-existence of Stable Matches with Incomplete Information," *Games and Economic Behavior*, 120, 336–344.

Aumann, R. J. (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55(1), 1–18.

Battigalli, P., A. Di Tillio, E. Grillo, and A. Penta (2011): "Interactive Epistemology and Solution Concepts for Games with Asymmetric Information," *The B.E. Journal of Theoretical Economics*, 11(1).

Battigalli, P. and A. Friedenberg (2012): "Forward Induction Reasoning Revisited," *Theoretical Economics*, 7(1), 57–98.

Battigalli, P. and M. Siniscalchi (1999): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory*, 88(1), 188–230.

——— (2002): "Strong Belief and Forward Induction Reasoning," *Journal of Economic Theory*, 106(2), 356–391.

——— (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3(1), Article 3.

Bergemann, D. and S. Morris (2009): "Robust Implementation and Direct Mechanisms," *Review of Economic Studies*, 76(4), 1175–1204.

——— (2011): "Robust Implementation and General Mechanisms," *Games and Economic Behavior*, 71(2), 261–281.

——— (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81(4), 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

——— (2017): "Belief-free Rationalizability and Informational Robustness," *Games and Economic Behavior*, 104, 744–759.

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

Bikhchandani, S. (2017): "Stability with One-sided Incomplete Information," *Journal of Economic Theory*, 168, 372–399.

Börgers, T. (1994): "Weak Dominance and Approximate Common Knowledge," *Journal of Economic Theory*, 64(1), 265–276.

Brandenburger, A. and E. Dekel (1987): "Rationalizability and Correlated Equilibrium," *Econometrica*, 55(6), 1391–1402.

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

Chade, H., G. Lewis, and L. Smith (2014): "Student Portfolios and the College Admissions Problem," *Review of Economic Studies*, 81(3), 971–1002.

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

Chen, Y.-C. and G. Hu (2020): "Learning by Matching," *Theoretical Economics*, 15(1), 29–56.

——— (2021): "A Theory of Stability in Matching with Incomplete Information," *Working paper*.

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

Dekel, E., D. Fudenberg, and S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2(1), 15–40.

Dekel, E. and M. Siniscalchi (2015): "Chapter 12 – Epistemic Game Theory," in *Handbook of Game Theory with Economic Applications*, ed. by H. P. Young and S. Zamir, Elsevier, vol. 4, 619–702.

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

Forges, F. (1993): "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, 35, 277–310.

——— (2006): "Correlated Equilibrium in Games with Incomplete Information Revisited," *Theory and Decision*, 61, 329–344.

Forges, F., E. Minelli, and V. Rajiv (2002): "Incentives and the Core of an Exchange Economy: a Survey," *Journal of Mathematical Economics*, 38(1–2), 1–41.

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

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

Liu, Q. (2015): "Correlation and Common Priors in Games with Incomplete Information," *Journal of Economic Theory*, 157, 49–75.

——— (2017a): "Rational Expectations Stability and Competitive Equilibrium in Matching with Incomplete Information," *Working paper*.

——— (2017b): "Stable Belief and Stable Matching," *Working paper*.

——— (2020): "Stability and Bayesian Consistency in Two-Sided Markets," *American Economic Review*, 110(8), 2625–66.

Liu, Q., G. J. Mailath, A. Postlewaite, and L. Samuelson (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82(2), 541–587.

Myerson, R. B. (1986): "Multistage Games with Communication," *Econometrica*, 54(2), 323–358.

Nash, J. (1953): "Two-Person Cooperative Games," *Econometrica*, 21(1), 128–140.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52(4), 1029–1050.

Penta, A. and M. Ollár (2017): "Full Implementation and Belief Restrictions," *American Economic Review*, 107(8), 2243–77.

——— (2021): "A Network Solution to Robust Implementation: the Case of Identical but Unknown Distributions," *Working paper*.

Penta, A. and P. Zuazo-Garin (2022): "Rationalizability, Observability and Common Knowledge," *Review of Economic Studies*, 89(2), 948–975.

Pomatto, L. (2021): "Stable Matching under Forward-Induction Reasoning," *Working paper*.

Radner, R. (1979): "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica*, 47(3), 655–678.

Rényi, A. (1955): "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungarica*, 6, 285–335.

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

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

Samuelson, L. (1992): "Dominated Strategies and Common Knowledge," *Games and Economic Behavior*, 4(2), 284–313.

Serrano, R. (2008): "Nash Program," in *The New Palgrave Dictionary of Economics (2nd Edition)*, ed. by S. N. Durlauf and L. E. Blume, Palgrave Macmillan.

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

Stalnaker, R. (1998): "Belief Revision in Games: Forward and Backward Induction," *Mathematical Social Sciences*, 36(1), 31–56.

Tan, T. C.-C. and S. R. d. C. Werlang (1988): "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45(2), 370–391.

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

Chapter 2

All-Pay Auctions with General Information Structures

2.1 Introduction

Equilibrium analysis of auctions typically assumes that the designer not only knows the common prior distribution of values, but also the information bidders may receive which induces their beliefs and higher order beliefs. Two prevailing models that theorists have been working with are the independent private values model in which bidders receive independent and identically distributed signals that are equal to their true values (Krishna, 2002), and the affiliated signals model where bidders' values and signals are affiliated random variables (Milgrom and Weber, 1982). Concerns have been raised about these models due to the uncertainty a designer may face regarding the correct model of beliefs, and there is an emerging literature on robust auction design embracing such uncertainty of beliefs. In particular, Bergemann et al. (2017a) study the first price auction with general information structures, under the assumption that given a symmetric common prior of values, bidders' beliefs can be induced by an arbitrary information structure. They find a lower bound on revenue, and then construct an information structure and an equilibrium such that this lower bound is attained. A natural follow-up question one may ask is: How is the performance of other standard auctions we are familiar with? We can easily answer this question for the second-price auction since it always admits a "bidding ring" equilibrium regardless of the information structure, so the lower bound on revenue of a second-price auction is zero. In this paper,

we focus on the all-pay auction and explore its robustness property in comparison with the first-price auction.

Under particular assumptions on beliefs, there exist some discussions of such revenue comparison in the literature. From the celebrated revenue equivalence theorem (Myerson, 1981; Riley and Samuelson, 1981), we know that in an independent private values model, the expected revenue generated by any standard auction format is the same. Amann and Leininger (1995) and Krishna and Morgan (1993) discover that, when signals are affiliated, the expected revenue of the all-pay auction is weakly higher than that of the first-price auction. When bidders have a pure common and publicly known value but are facing random budget constraints, Che and Gale (1996) show that the all-pay auction yields a strictly higher expected revenue than a first-price auction. In contrast, in this paper, we show that when the auction designer does not know the correct model of beliefs which is consistent with a symmetric common prior, and evaluates an auction at its minimum expected revenue across all information structures and all equilibria, the all-pay auction performs *weakly worse* than the first-price auction. To this end, we first establish a revenue equivalence result in the "worst case" information structure constructed in Bergemann et al. (2017a), which implies that the minimum expected revenue of the all-pay auction never exceeds that of the first-price auction, and then construct an environment in which the former is strictly lower than the latter. Therefore, if a seller is extremely ambiguity averse about the uncertain model of beliefs, the seller should choose a first-price auction over an all-pay auction.[1]

In the special case of pure common value, some progress has been made to construct a robust auction mechanism which performs strictly better than does a first-price auction. As an implication of our revenue equivalence result, the auction to be constructed must be a non-standard one. Bergemann et al. (2020) characterize the optimal auctions when the bidders are intermediaries who wish to resell the good. Bergemann et al. (2017b) construct an optimal auction which maximizes the minimum expected revenue of the seller. Du (2018) constructs an "exponential price auction" such that as the number of bidders gets large, the minimum expected revenue converges to the full surplus. How to extend these auctions to a more general model of values, and how to construct simpler

[1]As pointed out in Bergemann et al. (2017b), this claim is based on distinct assumptions about seller's attitudes towards different sources of ambiguity in this environment: Conditional on a model of beliefs, the seller is risk neutral about the realization of the outcomes, while the seller only cares about the minimum expected revenue regarding the uncertainty about the model of beliefs and equilibrium selection.

auctions that perform relatively well remain to be explored in future studies.

2.2 Model

A single unit of good is to be allocated by an auction. There are N risk neutral bidders with quasi-linear preferences over the allocation and payments. We denote the set of bidders by $\mathcal{N} = \{1, \ldots, N\}$. Each bidder $i \in \mathcal{N}$ has a value v_i which is randomly drawn from a compact interval $V = [\underline{v}, \bar{v}] \subset \mathbb{R}_+$, and the values $v \in V^N$ are jointly distributed according to a probability measure $\mu \in \Delta(V^N)$.[2] In this paper, we shall focus on the cases where the common prior μ is *symmetric*.[3] Notice that, although symmetry is a restrictive assumption, it encompasses both the independent private value case and the pure common value case, and more generally, values could be correlated in many different ways.

The distribution of the average of the $N-1$ lowest values plays a central role in our analysis. Let $Q(\cdot)$ denote the cumulative distribution function (CDF) of this random variable, and we assume that $Q(\cdot)$ is continuous, i.e. this distribution is non-atomic.

We now describe the information that bidders may obtain in addition to the common prior. An *information structure* $\mathcal{S} = (S, \pi)$ contains a product space $S = \prod_{i=1}^N S_i$, where each S_i is a measurable space of signals that bidder i may receive, and a measurable mapping $\pi : V^N \to \Delta(S)$ which generates a probability measure over signal profiles given each realization of values. Therefore, the signals themselves may be correlated, and they are also correlated with the profile of values. The information structure $\mathcal{S}$, as well as the common prior μ, is assumed to be common knowledge. We then interpret each bidder's signal as her private information, and she can only infer her opponents' signals by Bayesian updating.

An *auction* mechanism (or auction, for simplicity) $\mathcal{A} = (q_i, t_i)_{i \in \mathcal{N}}$ is a set of allocation rules $q_i : B^N \to [0, 1]$ satisfying $\sum_{i=1}^N q_i(b) \leq 1$ for all $b \in B^N$ and a set of payment rules $t_i : B^N \to \mathbb{R}$, where $B = [0, \infty)$ is the set of possible bids that a bidder may submit.

[2] We consider the product topology of product spaces. All topological spaces are endowed with the Borel sigma-algebra. For any topological space X, let $\Delta(X)$ denote the set of Borel probability measures on X, endowed with the weak* topology.

[3] The precise definition of symmetry is stated as follows. Let Ξ denote the set of permutations of the index set $\mathcal{N}$. Each $\xi \in \Xi$ defines a (unique) mapping $\hat{\xi} : V^N \to V^N$ by $\left(\hat{\xi}(v)\right)_i = v_{\xi^{-1}(i)}, \forall i = 1, \ldots, N$. We say that the common prior μ is *symmetric* if for all measurable sets $E \subset V^N$ and all $\hat{\xi}$, we have $\mu(E) = \mu(\hat{\xi}(E))$.

Given a profile of bids $b \in B^N$, $q_i(b)$ is the probability that bidder i obtains the good, and $t_i(b)$ is the amount the payments that bidder i makes to the seller. We assume that a bidder can always opt out by submitting 0, i.e. $q_i(0, b_{-i}) = t_i(0, b_{-i}) = 0$ for all $b_{-i} \in B^{N-1}$. Therefore, an auction defines a (base) game in which the bidders submit their bids simultaneously, and obtain payoffs

$$U_i(v_i, b) = v_i q_i(b) - t_i(b). \tag{2.1}$$

As in Krishna (2002), we call an auction *standard* if the allocation rule says that the object is awarded to the bidder who submits the highest bid. In addition, we assume that if there are multiple winning bids, the winner is selected uniformly. Hence, if we let $W(b) = \{i : b_i \geq b_j, \forall j\}$ denote the set of high bidders, and χ_E be the indicator function of an event E, an auction is standard if and only if

$$q_i(b) = \frac{\chi_{i \in W(b)}}{|W(b)|}.$$

Most of the auction formats we are familiar with are standard. For example, the first-price auction is standard with $t_i(b) = b_i q_i(b)$; the second-price auction is standard with $t_i(b) = (\max_{j \neq i} b_j) q_i(b)$; and the all-pay auction is also standard with $t_i(b) = b_i$.

Given an information structure $\mathcal{S}$ and an auction $\mathcal{A}$, we have a *Bayesian game*. A strategy of a bidder i is represented by a measurable mapping $\sigma_i : S_i \to \Delta(B)$. A *Bayes-Nash Equilibrium* (BNE) is a strategy profile $\sigma = (\sigma_1, \ldots, \sigma_N)$ such that for all i and any σ_i', we have

$$\begin{aligned} &\int_{v \in V^N} \int_{s \in S} U_i(v_i, \sigma_i(s_i), \sigma_{-i}(s_{-i})) \pi(ds\,|v) \mu(dv) \\ &\qquad \geq \int_{v \in V^N} \int_{s \in S} U_i(v_i, \sigma_i'(s_i), \sigma_{-i}(s_{-i})) \pi(ds\,|v) \mu(dv), \end{aligned}$$

where $U_i(v_i, \sigma_i(s_i), \sigma_{-i}(s_{-i}))$ is the multi-linear extension of $U_i(v_i, b)$ in the equation (1). Fixing an information structure $\mathcal{S}$ and an auction $\mathcal{A}$, let $\Sigma^*(\mathcal{S}, \mathcal{A})$ denote the set of all possible BNE.

2.3 Revenue Comparison

In our model, we evaluate the performance of an auction by its lowest possible revenue across all information structures and all BNE. We call this tight lower bound the *robust revenue* of an auction $\mathcal{A}$, and it is calculated as

$$R^*(\mathcal{A}) = \inf_{\mathcal{S}} \inf_{\sigma \in \Sigma^*(\mathcal{S},\mathcal{A})} \int_{v \in V^N} \int_{s \in S} \sum_{i=1}^{N} t_i(\sigma_i(s_i), \sigma_{-i}(s_{-i})) \pi(ds\,|v) \mu(dv).$$

This measure of performance may appeal to a seller who is extremely ambiguity averse about bidders' information and equilibrium selection.

We now describe an information structure that is of particular interest in our study. Define

$$a(v) = \frac{1}{N-1}\left(\sum_i v_i - \max_i v_i\right)$$

as the average of the $N-1$ lowest values given $v \in V^N$, and recall that $Q(\cdot)$ is the CDF of the average of the $N-1$ lowest values, i.e.

$$Q(\omega) = \mu(\{v | a(v) \leq \omega\}).$$

Let $[\underline{\omega}, \bar{\omega}]$ be the support of $Q(\omega)$.

Definition 2.1. In an *independent lower average* information structure, each bidder's signal is an independent draw from the distribution $F(s) = Q(s)^{\frac{1}{N}}$, and the signals and values are correlated in a way such that (i) The highest signal is equal to the realized value of $a(v)$; and (ii) The bidder with the highest value receives the highest signal.

In the special case of pure common value, this information structure reduces to the "maximum game" studied in Bulow and Klemperer (2002) and Bergemann et al. (2020). In first-price auctions, Bergemann et al. (2017a) show that this is actually a "winning-bid-minimizing" information structure such that the robust revenue is achieved by a symmetric BNE.

Proposition 2.1 (Bergemann et al. (2017a))**.** *In the first-price auction, the strategy profile*

$$\sigma_i(s) = \frac{1}{Q^{(N-1)/N}(s)} \int_{\omega=\underline{\omega}}^{s} \omega \frac{N-1}{N} \frac{1}{Q^{1/N}(\omega)} Q(d\omega), \quad \forall i \in \mathcal{N}$$

constitutes a Bayes-Nash Equilibrium under the independent lower average information structure. Moreover, the robust revenue of first-price auction is achieved by this information structure and equilibrium.

If we define $G(\omega) = Q(\omega)^{(N-1)/N}$, the equilibrium strategy in Proposition 1 can be rewritten as

$$\sigma_i(s) = \frac{1}{G(s)} \int_{\omega=\underline{\omega}}^{s} \omega G(d\omega).$$

Notice that this is exactly the equilibrium strategy of a first-price auction with *independent private values* where $s_i = v_i$ (Krishna, 2002). Indeed, there is a surprising connection between the independent lower average information structure and the independent private value (IPV) model.

Proposition 2.2 (Revenue Equivalence)**.**

(i) Consider an IPV environment in which bidders' values are independent draws from $Q(s)^{\frac{1}{N}}$. For any standard auction which has a symmetric and increasing equilibrium σ^*_{IPV} in such IPV environment, where a bidder with the lowest possible signal obtains zero surplus, this equilibrium strategy σ^*_{IPV} still constitutes a Bayes-Nash equilibrium in the independent lower average information structure in our setting with common prior μ.

(ii) All these Bayes-Nash equilibria of standard auctions generate the same ex ante expected revenue for the seller.

Proof. According to the revenue equivalence theorem in the IPV model, in a symmetric and increasing equilibrium σ^*_{IPV} of any standard auction where a bidder with the lowest possible signal $\underline{\omega}$ obtains zero surplus, the expected payments of a bidder receiving signal s is

$$t_{\text{IPV}}(s) = \int_{\omega=\underline{\omega}}^{s} \omega G(d\omega).$$

We next argue that all bidders playing the strategy σ^*_{IPV} is also a Bayes-Nash equilibrium in the independent lower average information structure in our model. To this end, we only need to show that after receiving a signal s, it is not profitable for a bidder i to deviate to another bid $\sigma^*_{\text{IPV}}(s')$ if all others are following the strategy σ^*_{IPV}. If bidder i receives a signal s and submits $\sigma^*_{\text{IPV}}(s)$, the expected payoffs are

$$\tilde{v}(s)G(s) - \int_{\omega=\underline{\omega}}^{s} \omega G(d\omega),$$

where $\tilde{v}(s)$ is the expected highest value conditional on $\alpha(v) = s$, and the second part is just the expected payments in the IPV model.

If bidder i deviates to a bid $\sigma^*_{\text{IPV}}(s')$ with $s' < s$, she wins only if she was going to win if she did not deviate, i.e. s is the highest signal. In this case, bidder i's expected value is still $\tilde{v}(s)$. (This is because the calculation of $\tilde{v}(s)$ is based on the realized value of $\alpha(v)$, for which the highest signal is a sufficient statistic.) In addition, the expected payments now become $t_{\text{IPV}}(s')$. Hence, the expected payoffs can be written as

$$\begin{aligned}
\tilde{v}(s)G(s') - \int_{\omega=\underline{\omega}}^{s'} \omega G(d\omega) &= (\tilde{v}(s) - s)G(s') + \left(sG(s') - \int_{\omega=\underline{\omega}}^{s'} \omega G(d\omega) \right) \\
&\leq (\tilde{v}(s) - s)G(s) + \left(sG(s') - \int_{\omega=\underline{\omega}}^{s'} \omega G(d\omega) \right) \\
&\leq (\tilde{v}(s) - s)G(s) + \left(sG(s) - \int_{\omega=\underline{\omega}}^{s} \omega G(d\omega) \right) \\
&= \tilde{v}(s)G(s) - \int_{\omega=\underline{\omega}}^{s} \omega G(d\omega),
\end{aligned}$$

where the first inequality comes from the fact that $\tilde{v}(s) - s \geq 0$, and the second inequality is because that σ^*_{IPV} is an equilibrium strategy in the IPV environment. Therefore, downward deviations are not profitable.

On the other hand, if bidder i deviates to a bid $\sigma^*_{\text{IPV}}(s')$ with $s' > s$, she still wins if s is the highest signal, and such gain from allocation is $\tilde{v}(s)G(s)$. In addition, she can also win if the highest signal lies between s and s', and such gain from allocation is $\int_{\omega=s}^{s'} \omega G(d\omega)$. This is because conditional on opponents' highest signal being ω and $s < \omega$, bidder i's expected value is exactly ω — the average of the lower values. Again, the expected payments become $t_{\text{IPV}}(s')$. Therefore, the expected payoffs from upward deviation is

$$\tilde{v}(s)G(s) + \int_{\omega=s}^{s'} \omega G(d\omega) - \int_{\omega=\underline{\omega}}^{s'} \omega G(d\omega) = \tilde{v}(s)G(s) - \int_{\omega=\underline{\omega}}^{s} \omega G(d\omega),$$

meaning that bidder i is actually indifferent to any upward deviation.

Therefore, we have proved the first statement of Proposition 2. The second statement is a direct consequence of the revenue equivalence theorem in the IPV environment. □

A combination of Proposition 1 and Proposition 2 yields the following Corollary.

Corollary 2.1. *The robust revenue of a first-price auction is weakly higher than that*

of any standard auction formats which have a symmetric and increasing equilibrium. In particular, the all-pay auction performs weakly worse than the first-price auction in terms of robust revenue.

Proposition 2 and Corollary 1 can be viewed as an extension of the Proposition 3 and Corollary 2 in Bergemann et al. (2020), where they focus on the special case of pure common value. Our analysis shows that even in the more general environment where values are only assumed to be symmetric, the performance of first-price auction is still very robust. As long as we want to use an easily understood auction mechanism such that "the highest bid wins the auction," first-price auction ensures the highest robust revenue. This Corollary also suggests that, if we want to construct a robust mechanism which performs strictly better than the first-price auction, the resulting mechanism should be a "non-standard" one. Indeed, the "exponential price auction" constructed in Du (2018) and the optimal auctions designed in Bergemann et al. (2017b) and Brooks and Du (2021) are all non-standard. Nonetheless, all these papers mentioned above focus on the pure common value case, and searching for an optimal mechanism in the more general environment seems to be a promising direction for future research.

Remark. Even if we consider standard auctions with reserve price, the first-price auction is still superior to any other standard auction format. To see this, we only need to apply a variant of the revenue equivalence theorem for a fixed marginal signal $\hat{s}$ below which it is optimal for a bidder to opt out (Riley and Samuelson, 1981). Therefore, we can categorize the class of standard auctions with reserve price by the marginal bidder signal $\hat{s}$, i.e. the cutoff of the exclusion induced by the reserve price. By a similar argument as in the proof of Proposition 2, one can see that in every subclass of the standard auctions with reserve price, the strategic and revenue equivalence still holds in the independent lower average information structure, and hence the first-price auction ensures the highest robust revenue.

We have now established that the all-pay auction generates weakly less robust revenue than the first-price auction. The next natural question we want to understand is: Can an all-pay auction be strictly worse than a first-price auction in some situations? Before addressing this, it is worth mentioning that we already know the answer for the second-price auction. In a second-price auction, for any information structure, there *always* exists a bad equilibrium in which one bidder makes an extremely high bid, and

all others bid zero. Therefore, the robust revenue generated by a second-price auction is actually zero.

We now present a heuristic example which suggests that an all-pay auction could be strictly worse than a first-price auction. Let us assume that there are two bidders, one has a high value 2, and the other has a low value 1. For each bidder, it is equally likely to have either a high value or a low value, but the values are correlated such that whenever one bidder has a high value, her opponent must have a low value, and vice versa. In this environment, a straightforward lower bound on revenue of the first-price auction is 1. Indeed, this lower bound can be achieved by full disclosure of information, in which case the auction becomes a Bertrand competition. We next turn to the all-pay auction, and again assume that the information is fully disclosed by an information structure. In each realization of values, it is easy to check that the high value bidder submitting a bid uniformly drawn from $(0, 1)$, and low value bidder opting out with $\frac{1}{2}$ probability and submitting a bid uniformly drawn from $(0, 1)$ with complementary probability constitute a BNE. A simple calculation shows that the ex ante expected revenue generated from this specific information structure and equilibrium is strictly lower than 1 — the robust revenue in the first-price auction. Our assumption that $Q(\cdot)$ is a continuous CDF is not satisfied by the example above, but the following example confirms that this assumption does not preclude the possibility of strict order of the robust revenues.

Example 2.1. Suppose that there are two bidders, and θ is a random variable uniformly distributed on $[1, 2] \cup [3, 4]$. Values are determined by $v_1 = \theta$, and $v_2 = 5 - \theta$. Therefore, the common prior on V^2 is symmetric (and perfectly correlated).

We first consider the robust revenue of a first-price auction. Since the lower value is uniformly distributed on $[1, 2]$, the signals in the independent lower average information structure is drawn from $F(s) = \sqrt{s-1}$ for $s \in [1, 2]$. By Proposition 1, the equilibrium strategy

$$\sigma_i(s) = \frac{1}{\sqrt{s-1}} \int_{\omega=1}^{2} \frac{1}{2\sqrt{\omega-1}}\, d\omega = \frac{s+2}{3}, \quad i = 1, 2$$

achieves the robust revenue, which can therefore be calculated as

$$R^*_{\text{FPA}} = \int_1^2 \frac{s+2}{3}\, ds = \frac{7}{6}.$$

We now construct an information structure and an equilibrium in an all-pay auction

such that the expected revenue is strictly lower than R^*_{FPA}. Consider an information structure that fully reveals the realization of values. Therefore, for every realization of values, we have an all-pay auction with complete information. If $\theta \in [1, 2]$,

$$\sigma_1(\theta) = \frac{5 - 2\theta}{5 - \theta} \times \delta_0 + \frac{\theta}{5 - \theta} \times \text{Uniform}(0, \theta),$$

and

$$\sigma_2(\theta) = \text{Uniform}(0, \theta),$$

constitute an equilibrium, where δ_0 is the Dirac measure at 0 and the sum in $\sigma_1(\theta)$ is in the sense of convex combination over distributions. Let us briefly check that this strategy profile is an equilibrium: For bidder 1, any bid $b \in (0, \theta)$ gives her a surplus of $\theta \times \frac{b}{\theta} - b = 0$, and opting out also ensures a surplus of 0, so she is indifferent to any bid between 0 and θ; For bidder 2, by submitting a bid $b \in (0, \theta)$, she obtains a surplus of

$$\left(\frac{5 - 2\theta}{5 - \theta} + \frac{\theta}{5 - \theta} \times \frac{b}{\theta}\right) \times (5 - \theta) - b = 5 - 2\theta,$$

which does not depend on b, meaning that she is also indifferent to any bid between 0 and θ. Therefore, the strategy profile we specified is indeed an equilibrium. If $\theta \in [3, 4]$, due to the symmetry of values, we can easily retain the equilibrium by slightly modifying the strategies and relabeling the bidders. Hence, the ex ante expected revenue a seller can obtain from this equilibrium is

$$\int_1^2 \left(\frac{\theta}{2} + \frac{\theta}{5 - \theta} \times \frac{\theta}{2}\right) d\theta = \left(-\frac{5}{2}\theta - \frac{25}{2}\log(5 - \theta)\right)\Big|_1^2 = 1.096,$$

which is strictly less than $R^*_{\text{FPA}} = \frac{7}{6}$. ♢

This example provides a constructive proof of the following result.

Proposition 2.3. *For some symmetric common prior μ, the robust revenue of an all-pay auction is strictly less than that of a first-price auction.*

Notice that we are not claiming the robust revenue of an all-pay auction is always strictly less — it is true for *some* environments. But we believe that from the perspective of a seller, this result gives us enough reason to choose a first-price auction over an

all-pay auction in an effort to guarantee a higher robust revenue.[4]

2.4 Bibliography

Amann, E. and W. Leininger (1995): "Expected Revenue of All-Pay and First-Price Sealed-Bid Auctions with Affiliated Signals," *Journal of Economics*, 61(3), 273–279.

Bergemann, D., B. Brooks, and S. Morris (2017a): "First-Price Auctions With General Information Structures: Implications for Bidding and Revenue," *Econometrica*, 85(1), 107–143.

——— (2017b): "Informationally Robust Optimal Auction Design," *Working paper*.

——— (2020): "Countering the Winner's Curse: Optimal Auction Design in a Common Value Model," *Theoretical Economics*, 15(4), 1399–1434.

Brooks, B. and S. Du (2021): "Optimal Auction Design With Common Values: An Informationally Robust Approach," *Econometrica*, 89(3), 1313–1360.

Bulow, J. and P. Klemperer (2002): "Prices and the Winner's Curse," *RAND Journal of Economics*, 33(1), 1–21.

Che, Y.-K. and I. Gale (1996): "Expected Revenue of All-Pay Auctions and First-Price Sealed-Bid Auctions with Budget Constraints," *Economics Letters*, 50(3), 373–379.

Du, S. (2018): "Robust Mechanisms Under Common Valuation," *Econometrica*, 86(5), 1569–1588.

Krishna, V. (2002): *Auction Theory*, San Diego: Academic Press.

Krishna, V. and J. Morgan (1993): "An Analysis of the War of Attrition and the All-Pay Auction," *Journal of Economic Theory*, 72(2), 343–362.

Milgrom, P. R. and R. J. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5), 1089–1122.

[4]One can see that the robust revenue of the all-pay auction should be bounded below away from 0. This is because every bidder submitting zero is clearly not an equilibrium. Therefore, we have an idea about the range of this robust revenue but do not know exactly where it is. However, we consider this question less intriguing since we are interested in auctions that perform well in general.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operation Research*, 6(1), 58–73.

Riley, J. G. and W. F. Samuelson (1981): "Optimal Auctions," *American Economic Review*, 71(3), 381–392.

Chapter 3

Robust Predictions in Dynamic Games with Incomplete Information

3.1 Introduction

Models of strategic interactions often impose strong common knowledge assumptions on payoffs, and they usually have a large set of rationalizable or equilibrium strategies. Game theorists have observed that, by slightly relaxing the common knowledge assumption, we are able to select some equilibrium outcomes as a unique prediction (Rubinstein, 1989; Carlsson and van Damme, 1993; Morris and Shin, 2000). Weinstein and Yildiz (2007) substantially generalize this insight in a static environment where all common knowledge assumptions are relaxed (i.e. the so-called *richness condition* holds). In particular, they show that (i) (the *Structure Theorem*:) any interim correlated rationalizable (ICR) strategy of a type can be uniquely selected by perturbing higher order beliefs, and (ii) (*generic uniqueness*:) the prediction delivered by ICR is generically unique in the space of all belief hierarchies, i.e. the universal type space. Chen (2012) and Penta (2012) extend these results to dynamic games with suitably defined richness conditions.[1]

In this paper, we focus on a dynamic environment where common knowledge on

[1]The richness condition of Weinstein and Yildiz (2007) rules out all genuine dynamic games because a strategy cannot be strictly dominant when some information set may not be reached. To tackle this issue, Chen (2012) focuses on the normal form of dynamic games and extends the results under a weakened version of (extensive-form) richness. Penta (2012) studies a more general dynamic environment with an information structure. Based on sequential rationality, he proposes a solution concept *interim sequential rationalizability* and establishes its Structure Theorem and generic uniqueness under a richness condition in his setting.

payoffs is captured by a *payoff-information (PI) structure*. We dispense with any richness condition and study robust predictions under arbitrary PI structures. This direction of exploration is important for two reasons. First, although sequential rationality has a bite at all histories when a player has some information about payoffs, the richness condition implies all opponents' strategies can potentially be sequentially rational. Therefore, any off-path belief is essentially unrestricted. Only by maintaining some common knowledge on payoffs, a player's information can interact with his beliefs about other players' rationality throughout the game. Our approach allows for these situations of interest, which are ruled out by the richness condition. Second, if we, as researchers analyzing a dynamic interaction, are confident to assume a PI structure where the richness condition is not satisfied, existing results in the literature cannot be used to help us refine our prediction. In this case, the present paper provides useful tools for us to decide which predictions are robust against misspecification of higher order beliefs within the PI structure we are confident with.

A solution concept is *robust* if it is an upper hemicontinuous correspondence defined on the universal type space (Fudenberg et al., 1988; Dekel and Fudenberg, 1990; Weinstein and Yildiz, 2007). In other words, robustness requires that the prediction for a type does not reject any strategy that is not rejected by a sufficiently "close" type.[2] We choose *extensive form rationalizability (EFR)* (Pearce, 1984; Battigalli, 1997) as the main solution concept of our analysis because it is robust in this sense, as shown by recent work of Piermont and Zuazo-Garin (2021). Moreover, EFR captures the epistemic assumption of sequential rationality and common strong belief thereof, so it imposes restrictions on players' off-path beliefs and thus is suitable for genuine dynamic games.

In the literature of dynamic games with incomplete information, EFR is either defined for a type space on which common knowledge is assumed, or for the universal type space of a PI structure. These notions cannot be directly used for our purpose. We believe the most realistic situation is where we, as researchers, impose a common knowledge assumption on a PI structure, but model players' initial beliefs using a small type space. Therefore, a player's conjecture about opponents' behaviors is restricted by the type space initially, but once an off-path history is reached, this player maintains his (strong) belief about opponents' rationality under the PI structure (but not the type space).[3] Motivated by this observation, we formulate an interim version of EFR

[2]We endow the universal type space with the product topology, so two types are considered close if their first n orders of beliefs are almost the same, with n arbitrarily large.

[3]Therefore, our approach sticks to the interpretation that a type space is only a convenient tool to

(Definition 3.3) defined for a type space. This definition relies on an auxiliary "belief-free" version of EFR, which is computed based on the PI structure without referring to any type spaces.[4] We show in Lemma 3.1 that this interim EFR is consistent with the one defined for the universal type space (Piermont and Zuazo-Garin, 2021).

We then ask a natural question: Given a finite type space in the PI structure, can we refine EFR without upsetting robustness? We focus on finite type spaces because they are ubiquitously used in applied game theory. Formally, a prediction for a finite type space specifies a set of strategies for each type, and we say a prediction is a *robust refinement* of EFR if it is consistent with an upper hemicontinuous sub-correspondence of EFR on the universal type space (Definition 3.4). To answer this question, we follow Chen et al. (2021) and employ a curb collection approach.[5] In particular, we define and compute the *upper (lower) EFR collection* of each payoff type, which consists of all sets of strategies that contain (are contained by, respectively) the set of EFR strategies for some type, with that payoff type, in the universal type space. Moreover, we define and compute a *local upper EFR collection* of each type in the given finite type space. This local collection traces out all minimal sets of EFR strategies around that type when we envision it in the universal type space. Our three main characterization results are summarized as follows:

(1) (Proposition 3.4) For a given finite type space, a prediction is a robust refinement of EFR if and only if the predicted set of strategies for every type intersects with all elements in the local upper EFR collection of that type.

(2) (Proposition 3.5) A Structure Theorem of EFR holds if and only if, in the universal type space, every rationalizable strategy of a type is uniquely rationaliable for some type with the same payoff type. In this case, EFR is the strongest robust prediction.

(3) (Proposition 3.6) Generic uniqueness of EFR holds if and only if, in the universal type space, every set of rationalizable strategies of a type contains at least one strategy that is uniquely rationalizable for some type with the same payoff type.

model beliefs, but it should not be associated with any additional common knowledge assumptions.

[4]Because of this dependence on the PI structure, our interim EFR is *not* invariant to the PI structure (see Penta, 2012).

[5]The *curb collection* is a generalization of the *curb set* proposed by Basu and Weibull (1991). Chen et al. (2014) first use this notion to study robust selection of ICR; Chen et al. (2021) also use it to study robust refinement of ICR. Both papers study static games.

This set of results can be useful in applications, especially when we are confident with certain aspects of the environment. We demonstrate the power of these results by applying them to study two forms of higher order uncertainty. In the first application, we consider the Beer-Quiche game of Cho and Kreps (1987) and introduce uncertainty about the privacy of sender's information. The PI structure is expanded in a way to allow for the possibility that the receiver *knows* the type of the sender. We still focus on the original model of player's beliefs, which now corresponds to a type space that assumes initial common belief in the privacy of sender's information. Using our results, we show that EFR is generically unique in the universal type space, and the strongest robust refinement for the type space in question is outcome equivalent to the equilibrium that satisfies the Intuitive Criterion. In the second application, we generalize the analysis in Penta and Zuazo-Garin (2022) and introduce uncertainty about the observability of actions in stage games. As an example, we consider two coordination games in the second stage, and a player can pick one of them in the first stage. We find that, when uncertainty about observability is present, the strongest robust refinement of EFR is as if the first-mover has an advantage and chooses a coordination outcome that he prefers the most. Albeit intuitive, this sharp prediction cannot be delivered by existing solution concepts.

This paper belongs to a game-theoretic literature that studies predictions robust to higher order payoff uncertainty. Several papers investigate this question in dynamic games like we do. Apart from the work by Chen (2012) and Penta (2012), Weinstein and Yildiz (2013) extend the insight of Weinstein and Yildiz (2007) to infinite-horizon repeated games and establish an "unrefinable" Folk Theorem in that setting. Piermont and Zuazo-Garin (2021) introduce a novel heterogeneity of perceptions about payoffs and explore its implications in dynamic games. While all these papers rely on a richness condition,[6] we discard it in the present paper.

A branch of the literature also considers robust predictions without richness. Penta (2013) provides sufficient conditions under which the results of Weinstein and Yildiz (2007) still hold. Chen et al. (2021) fully characterize the Structure Theorem and generic uniqueness of ICR under arbitrary payoff uncertainty. Weinstein and Yildiz (2011) characterize the sensitivity of *Bayes Nash equilibrium* to higher order beliefs without making any richness assumption. All these papers relax the richness condition in static

[6]Piermont and Zuazo-Garin (2021) do not invoke the richness condition directly. Instead, they define a notion of *higher order richness* and show that it is generically satisfied.

games. The contribution of the present paper is that we consider dynamic games and work with a solution concept designed for genuine dynamic settings.

Most of the work mentioned above is aimed at delivering a negative result: We cannot robustly refine our predictions for models we usually use, although there is a plethora of rationalizable or equilibrium strategies. However, we feel that the message of this paper is a positive one and is in line with Penta and Zuazo-Garin (2022): If we (i) are confident with a particular space of payoff uncertainty, or (ii) want to examine the relaxation of a specific form of common knowledge, then robustness to higher order uncertainty may *help* us refine the multiplicity of rationalizability.[7]

The remainder of this paper is organized as follows. Section 3.2 introduces the game-theoretic model and our solution concept. Section 3.3 contains the main characterization results. Two applications of these results are provided in Section 3.4.

3.2 Preliminaries

3.2.1 Game-Theoretic Model

We specialize our analysis in multistage games with observable actions (Fudenberg and Tirole, 1991; Osborne and Rubinstein, 1994). A dynamic game with incomplete information Γ consists of an extensive form $\mathcal{E}$ that describes the rule of the game and a preference-information structure $\mathcal{I}$ that defines players' information about their payoffs. Given a dynamic game with incomplete information, we usually attach to it a type space (or model) $\mathcal{T}$ which represents players' (exogenous) beliefs at the beginning of the game.

Formally, an *extensive form* is defined by a tuple

$$\mathcal{E} = \{I, \mathcal{H}, \mathcal{Z}, (A_i)_{i \in I}\}.$$

The finite set of players is denoted by I, and A_i is the set of actions player i can choose. A *history* is a finite concatenation of action profiles which describes the actions chosen by players in all previous stages. The set of all possible histories is partitioned into the set of partial histories $\mathcal{H}$ (including the *empty history* ϕ) and the set of terminal

[7] In a different direction, Heifetz and Kets (2018) and Germano et al. (2020) weaken the solution concept ICR by introducing (higher order) uncertainty about limited reasoning ability or limited rationality, respectively. They show that when the assumption of common belief in rationality is perturbed, robust proper refinements become possible.

histories $\mathcal{Z}$. Write $A_i(h)$ for the actions available to player i at history h, and we say player i is active at h when $A_i(h)$ is nonempty. Let $\mathcal{H}_i$ denote the set of histories at which player i is active. A strategy of player i specifies an action in $A_i(h)$ for all $h \in \mathcal{H}_i$. We identify two strategies that only differ at precluded histories, and let S_i denote the set of player i's *reduced form pure strategies* (henceforth *strategies*, for brevity). The set of player i's opponents' strategies is denoted by $S_{-i} = \times_{j \neq i} S_j$. Moreover, we write $S_i(h)$ for the set of player i's strategies that do not preclude history h, and $\mathcal{H}_i(s_i)$ for the set of player i's partial histories not precluded by $s_i \in S_i$. For each strategy profile $s \in \times_{i \in I} S_i$, let $z(s) \in \mathcal{Z}$ denote the terminal history induced by s.

A *preference-information structure* (PI structure) is a tuple

$$\mathcal{I} = \left\{\Theta_0, (\Theta_i, u_i)_{i \in I}, \bar{\Theta}\right\},$$

where Θ_0 is a finite set of the states of nature, and Θ_i is a finite set of player i's *payoff types*. Each payoff type $\theta_i \in \Theta_i$ represents a piece of player i's hard information about payoffs and is known by player i before the game starts.[8] In our model, we allow the analyst to impose common knowledge assumptions on players' information via an *information restriction* $\bar{\Theta} \subseteq \times_{i \in I} \Theta_i$. That is, it is commonly known by the players that the profile of payoff types $(\theta_i)_{i \in I}$ lies in $\bar{\Theta}$. This can happen when certain combinations of players' information are collectively exclusive. Note that we can relax such common knowledge assumption by simply letting $\bar{\Theta} = \times_{i \in I} \Theta_i$, which is common in the literature. We denote by $u_i : \mathcal{Z} \times \Theta_0 \times \bar{\Theta} \to \mathbb{R}$ the utility function of player i. For each $i \in I$ and $\theta_i \in \Theta_i$, we let $\Theta_{-i} = \times_{j \neq i} \Theta_j$ and $\bar{\Theta}_{-i}(\theta_i) \subseteq \Theta_{-i}$ be the *section of* $\bar{\Theta}$ *at* θ_i:

$$\bar{\Theta}_{-i}(\theta_i) = \left\{\theta_{-i} \in \Theta_{-i} : (\theta_i, \theta_{-i}) \in \bar{\Theta}\right\}.$$

We now describe players' belief hierarchies based on $\Theta_0 \times \bar{\Theta}$ (Mertens and Zamir, 1985; Brandenburger and Dekel, 1993). The construction is standard as in the literature except for the information restriction $\bar{\Theta}$. For each $i \in I$ and $\theta_i \in \Theta_i$, let $Z_i^1(\theta_i) =$

[8] Therefore, we sometimes also refer to θ_i as player i's *information*.

$\Delta\left(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i)\right)$ be the set of player i's first order beliefs.[9] For $n \geq 1$, let

$$Z_{-i}^n(\theta_i) = \left\{ (\theta_{-i}, \zeta_{-i}^n) \in \Theta_{-i} \times \bigtimes_{j \neq i} Z_j^n(\theta_j) : \theta_{-i} \in \bar{\Theta}_{-i}(\theta_i) \right\},$$

and define iteratively

$$Z_i^{n+1}(\theta_i) = \left\{ \begin{array}{c} \zeta_i^{n+1} = \left(\tau_i^1, \ldots, \tau_i^{n+1}\right) \in Z_i^n(\theta_i) \times \Delta\left(\Theta_0 \times Z_{-i}^n(\theta_i)\right) : \\ \text{marg}_{\Theta_0 \times Z_{-i}^{n-1}(\theta_i)} \tau_i^{n+1} = \tau_i^n \end{array} \right\},$$

where $\Delta\left(\Theta_0 \times Z_{-i}^n(\theta_i)\right)$ is the set of $(n+1)$-th order belief of player i who has information θ_i. Note that the assumptions of coherency and common belief in coherency are embedded in this iterative definition. The set of player i's collectively coherent belief hierarchies with information θ_i is

$$H_i(\theta_i) = \left\{ \zeta_i = \left(\tau_i^1, \tau_i^2, \ldots\right) \in \bigtimes_{n \geq 1} \Delta\left(\Theta_0 \times Z_{-i}^n(\theta_i)\right) : \left(\tau_i^1, \ldots, \tau_i^n\right) \in Z_i^n(\theta_i) \text{ for all } n \geq 1 \right\}.$$

When analyzing a game with incomplete information, we usually summarize players' belief hierarchies in a concise representation defined as follows.

Definition 3.1 (Type Spaces)**.** A *type space* is a tuple $\mathcal{T} = \{(T_i, \vartheta_i, \kappa_i)_{i \in I}\}$, where T_i is a compact metrizable space that contains types of player i, $\vartheta_i : T_i \to \Theta_i$ is a continuous function that specifies a payoff type for each $t_i \in T_i$, and $\kappa_i : T_i \to \Delta(\Theta_0 \times T_{-i})$, such that $\kappa_i(t_i)\left[\vartheta_{-i}(t_{-i}) \in \bar{\Theta}_{-i}(\vartheta_i(t_i))\right] = 1$, is a continuous function that describes type t_i's belief about the states of nature and his opponents' types.

A type space is called *finite* if T_i is finite for every i. Each type t_i induces a belief hierarchy of player i in $H_i(\vartheta_i(t_i))$ as usual.[10] Generalizing the analysis from Mertens and

[9] In this paper, for any metrizable space X, we write $\Delta(X)$ for the space of probability measures defined on the Borel σ-algebra of X. We endow $\Delta(X)$ with the weak* topology, a product space with the product topology, and a finite space with the discrete topology.

[10] To be specific, the first order belief of type t_i is defined by

$$\tau_i^1(t_i)[E] = \kappa_i(t_i)\left[\{(\theta_0, t_{-i}) : (\theta_0, \vartheta_{-i}(t_{-i})) \in E\}\right]$$

for every measurable $E \subseteq \Theta_0 \times \bar{\Theta}_{-i}(\theta_i)$. Moreover, for every measurable $E \subseteq \Theta_0 \times Z_{-i}^1(\vartheta_i(t_i))$,

$$\tau_i^2(t_i)[E] = \kappa_i(t_i)\left[\{(\theta_0, t_{-i}) : \left(\theta_0, \vartheta_{-i}(t_{-i}), \tau_{-i}^1(t_{-i})\right) \in E\}\right]$$

Zamir (1985) and Brandenburger and Dekel (1993), it can be shown that when $H_i(\theta_i)$ is endowed with the product topology, there exists a belief-preserving homeomorphism

$$\beta_i(\theta_i) : H_i(\theta_i) \to \Delta(\Theta_0 \times H_{-i}(\theta_i)),$$

where

$$H_{-i}(\theta_i) = \left\{ (\theta_{-i}, \zeta_{-i}) \in \Theta_{-i} \times \bigtimes_{j \neq i} H_j(\theta_j) : \theta_j \in \bar{\Theta}_{-i}(\theta_i) \right\}.$$

We now define a tuple $\mathcal{T}^* = \{(T_i^*, \vartheta_i^*, \kappa_i^*)_{i \in I}\}$, where $T_i^* = \{(\theta_i, \zeta_i) : \theta_i \in \Theta_i \text{ and } \zeta_i \in H_i(\theta_i)\}$, and for each $t_i = (\theta_i, \zeta_i) \in T_i^*$, (i) $\vartheta_i^*(t_i) = \theta_i$ and (ii) $\kappa_i^*(t_i) = \beta_i(\theta_i)(\zeta_i)$. It is easy to check that this tuple $\mathcal{T}^*$ satisfies Definition 3.1, and is therefore referred to as the *universal type space*.[11] For a given type space $\mathcal{T} = \{(T_i, \vartheta_i, \kappa_i)_{i \in I}\}$, and $t_i \in T_i$, we write $\varphi_i^* : t_i \mapsto (\vartheta_i(t_i), \tau_i^1(t_i), \tau_i^2(t_i), ...) \in T_i^*$ for the mapping that maps each type into the universal type space. Since $\beta_i(\theta_i)$ is belief-preserving, the mapping φ_i^* satisfies

$$\kappa_i^*(\varphi_i^*(t_i))[E] = \kappa_i(t_i)\left[(\theta_0, t_{-i}) : (\theta_0, \varphi_{-i}^*(t_{-i})) \in E\right]$$

for all measurable $E \subseteq \Theta_0 \times T_{-i}^*$. A type $t_i \in T_i^*$ is called a *finite type* if it can be induced by a type in a finite type space.

3.2.2 Solution Concept

The solution concept we shall employ in this dynamic environment is the *extensive form rationalizability* (EFR) defined on a type space.[12] It captures a notion of forward induction: observed behaviors shape a player's conjecture about payoffs and opponents' future play. Before formally describing this solution concept in our model, we need some definitions in advance.

Fix a player $i \in I$. For any compact metrizable space X and every history $h \in \mathcal{H}$, let $[h] \subseteq X \times S_{-i}$ denote the event that history h is not precluded by player i's opponents'

defines the second order belief of type t_i. We can therefore recursively compute the entire belief hierarchy $\left(\tau_i^1(t_i), \tau_i^2(t_i), \ldots\right) \in H_i(\vartheta_i(t_i))$ induced by type t_i.

[11] With a slight abuse of terminology, we sometimes also refer to $T^* = \times_{i \in I} T_i^*$ as the universal type space.

[12] The notion of extensive form rationalizability was first proposed by Pearce (1984) and Battigalli (1997). Battigalli and Siniscalchi (2002) study an ex ante extension under incomplete information and provide an epistemic characterization. Piermont and Zuazo-Garin (2021) examine an interim version of EFR in the universal type space.

strategies, i.e.

$$[h] = X \times S_{-i}(h).$$

Definition 3.2 (Conditional Probability Systems)**.** A collection $\mu = (\mu(h))_{h\in\mathcal{H}}$ of probability distributions $\mu(h) \in \Delta(X \times S_{-i})$ is called a *conditional probability system* (CPS) over $X \times S_{-i}$ if the following two conditions are satisfied:

(i) For each history $h \in \mathcal{H}$, $\mu(h)[h] = 1$;

(ii) For every measurable $E \subseteq [h] \subseteq [h']$, we have $\mu(h)[E] \cdot \mu(h')[h] = \mu(h')[E]$.

Let $\Delta^{\mathcal{H}}(X \times S_{-i})$ denote the set of CPS over $X \times S_{-i}$.

We define a *conjecture* of player i with payoff type θ_i as a CPS over $\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i}$. Given a conjecture π_i, we write $r_i(\pi_i \mid \theta_i)$ for the set of *sequentially best responses* of player i with payoff type θ_i, i.e. $s_i \in r_i(\pi_i \mid \theta_i)$ if and only if for all $h \in \mathcal{H}(s_i)$,

$$s_i \in \arg\max_{s_i' \in S_i(h)} \sum_{\theta_0, \theta_{-i}, s_{-i}} u_i(z(s_i', s_{-i}), \theta_0, \theta_i, \theta_{-i}) \pi_i(h)[\theta_0, \theta_{-i}, s_{-i}].$$

When defining a type space, a type $t_i \in T_i$ may assign zero probability to some payoff type profile $\theta_{-i} \in \bar{\Theta}_{-i}(\vartheta_i(t_i))$ of his opponents, i.e. $\kappa_i(t_i)[\Theta_0 \times \vartheta_{-i}^{-1}(\theta_{-i})] = 0$. When reaching an unexpected history, type t_i may discard his initial belief while maintaining common knowledge of the PI structure $\mathcal{I}$. Therefore, we need to keep track of the "rational behaviors" of i's opponents that are consistent with the PI structure. To achieve this, we first define an auxiliary *belief-free* version of EFR that encodes players' behaviors of all payoff types.

For every player $i \in I$ and payoff type $\theta_i \in \Theta_i$, let $\mathrm{C}_i^0(\theta_i) = \Delta^{\mathcal{H}}(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i})$ and $\mathrm{BF}_i^0(\theta_i) = S_i$. For $n \geq 1$, write $\mathrm{BF}_{-i}^{n-1}(\theta_{-i}) = \times_{j\neq i} \mathrm{BF}_j^{n-1}(\theta_j)$, and define

$$\mathrm{C}_i^n(\theta_i) = \left\{ \pi_i \in \mathrm{C}_i^{n-1}(\theta_i) : \begin{array}{l} \forall h \in \mathcal{H}, \left(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i}(h)\right) \cap \mathrm{graph}\left(\mathrm{BF}_{-i}^{n-1}\right) \neq \varnothing \\ \text{implies } \pi_i(h)\left[\Theta_0 \times \mathrm{graph}\left(\mathrm{BF}_{-i}^{n-1}\right)\right] = 1 \end{array} \right\}$$

and

$$\mathrm{BF}_i^n(\theta_i) = \{s_i \in S_i : \exists \pi_i \in \mathrm{C}_i^n(\theta_i) \text{ s.t. } s_i \in r_i(\pi_i \mid \theta_i)\}.$$

The sequence $\{\mathrm{BF}_i^n(\theta_i)\}_{n\geq 0}$ is decreasing and converges in finite steps. Notice that this procedure does not involve any type space nor its implied beliefs. It records, recursively, the set of strategies that can be played in each round of EFR for *some* type in the

universal type space with payoff type θ_i.[13] We keep the resulting sets $\{\mathrm{C}_i^n(\theta_i)\}_{n\geq 0}$ as increasingly stronger conjecture restrictions to compute EFR strategies for arbitrary type spaces smaller than the universal one.

Definition 3.3 (Extensive Form Rationalizability)**.** Fix a type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$. For every $i \in I$ and $t_i \in T_i$, let $\mathrm{EFR}_i^0(t_i) = S_i$. For $n \geq 1$, let $\mathrm{EFR}_{-i}^{n-1}(t_{-i}) = \times_{j\neq i} \mathrm{EFR}_j^{n-1}(t_j)$, and define

$$\Psi_i^n(t_i) = \left\{ \pi_i \in \mathrm{C}_i^n(\vartheta_i(t_i)) : \begin{array}{l} \exists \mu_i \in \Delta(\Theta_0 \times T_{-i} \times S_{-i}) \text{ s.t.} \\ \text{(i) } \mathrm{marg}_{\Theta_0\times T_{-i}} \mu_i = \kappa_i(t_i); \\ \text{(ii) } \mu_i \left[\left\{ (\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \mathrm{EFR}_{-i}^{n-1}(t_{-i}) \right\} \right] = 1; \\ \text{(iii) } \pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}(t_{-i}) = \theta_{-i}\}] \end{array} \right\}$$

and

$$\mathrm{EFR}_i^n(t_i) = \{ s_i \in S_i : \exists \pi_i \in \Psi_i^n(t_i) \text{ s.t. } s_i \in r_i(\pi_i \mid \vartheta_i(t_i)) \}.$$

Finally, let $\mathrm{EFR}_i(t_i) = \bigcap_{n\geq 0} \mathrm{EFR}_i^n(t_i)$.

We now describe a version of EFR defined on the universal type space $\mathcal{T}^*$. Since $\mathcal{T}^*$ contains all possible belief hierarchies consistent with the PI structure, we no longer need to use the restrictions encoded by $\{\mathrm{C}_i^n(\theta_i)\}_{n\geq 0}$ and can instead incorporate forward induction reasoning in a more direct way. The following definition is from Piermont and Zuazo-Garin (2021). For every $i \in I$ and $t_i \in T_i^*$, we write $\bar{T}_{-i}^*(t_i) = \{t_{-i} \in T_{-i}^* : \vartheta_{-i}^*(t_{-i}) \in \bar{\Theta}_{-i}(\vartheta_i^*(t_i))\}$, i.e. $\bar{T}_{-i}^*(t_i)$ is the set of opponent types that t_i considers possible due to the information restriction $\bar{\Theta}$. Now let $\Psi_i^{*,0}(t_i) = \Delta^{\mathcal{H}}(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i})$ and $\mathrm{EFR}_i^{*,0}(t_i) = S_i$. For $n \geq 1$, write $\mathrm{EFR}_{-i}^{*,n-1}(t_{-i}) = \times_{j\neq i} \mathrm{EFR}_j^{*,n-1}(t_j)$, and define

$$\Psi_i^{*,n}(t_i) = \left\{ \begin{array}{l} \pi_i \in \Psi_i^{*,n-1}(t_i) : \\ \quad \exists \tilde{\mu}_i \in \Delta^{\mathcal{H}}(\Theta_0 \times \bar{T}_{-i}^*(t_i) \times S_{-i}) \text{ s.t.} \\ \quad \text{(i) } \mathrm{marg}_{\Theta_0\times T_{-i}^*} \tilde{\mu}_i(\phi) = \kappa_i^*(t_i); \\ \quad \text{(ii) } \forall h \in \mathcal{H}, \left(\Theta_0 \times \bar{T}_{-i}^*(t_i) \times S_{-i}(h)\right) \cap \mathrm{graph}\left(\mathrm{EFR}_{-i}^{*,n-1}\right) \neq \varnothing \\ \quad \text{implies } \tilde{\mu}_i(h)\left[\Theta_0 \times \mathrm{graph}\left(\mathrm{EFR}_{-i}^{*,n-1}\right)\right] = 1; \\ \quad \text{(iii) } \forall h \in \mathcal{H}, \pi_i(h) = \mathrm{marg}_{\Theta_0\times\bar{\Theta}_{-i}(\vartheta_i(t_i))\times S_{-i}} \tilde{\mu}_i(h) \end{array} \right\}$$

and

$$\mathrm{EFR}_i^{*,n}(t_i) = \{ s_i \in S_i : \exists \pi_i \in \Psi_i^{*,n}(t_i) \text{ s.t. } s_i \in r_i(\pi_i \mid \vartheta_i^*(t_i)) \}.$$

[13] Ziegler (2022) makes a related observation for static games.

Finally, let $\text{EFR}_i^*(t_i) = \bigcap_{n\geq 0} \text{EFR}_i^{*,n}(t_i)$.

It is natural to ask whether Definition (3.3) of EFR applied to the universal type space coincides with EFR^* defined above. The following lemma gives a positive answer.

Lemma 3.1. *Fix a type space* $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$. *For every* $t_i \in T_i$, $\text{EFR}_i(t_i) = \text{EFR}_i^*(\varphi_i^*(t_i))$.

Proof. See Appendix 3.5.1. □

Corollary 3.1. *We use superscripts to indicate the type space* EFR *is defined on.*

(i) For any $t_i \in T_i^*$, $\text{EFR}_i^{\mathcal{T}^*}(t_i) = \text{EFR}_i^*(t_i)$;

(ii) For two type spaces $\mathcal{T}$ *and* $\mathcal{T}'$, $\text{EFR}_i^{\mathcal{T}}(t_i) = \text{EFR}_i^{\mathcal{T}'}(t_i')$ *if* $\varphi_i^*(t_i) = \varphi_i^*(t_i')$.

Due to this corollary, we sometimes omit the star in EFR^* and simply denote it by EFR without any confusion. The following property of EFR^* is important for our analysis.

Lemma 3.2 (Piermont and Zuazo-Garin (2021))**.** *For every* $n \geq 0$, $\text{EFR}_i^{*,n}(\cdot)$ *is upper hemicontinuous on* T_i^*. *Therefore,* $\text{EFR}_i^*(\cdot)$ *is upper hemicontinuous on* T_i^*; *that is, for each* $t_i \in T_i^*$ *and any sequence* $\{t_{i,m}\}_{m\in\mathbb{N}} \subseteq T_i^*$ *such that* $t_{i,m} \to t_i$, *if* $s_i \in \text{EFR}_i^*(t_{i,m})$ *for all* m *then* $s_i \in \text{EFR}_i^*(t_i)$.

The convergence $t_{i,m} \to t_i$ (in the product topology) means that $\vartheta_i^*(t_{i,m}) = \vartheta_i^*(t_i)$ for large enough m, and $\tau_i^n(t_{i,m}) \to \tau_i^n(t_i)$ (in the weak* topology) as $m \to \infty$ for every n. We next define a notion of unique selection that plays a crucial role.

Definition 3.4 (Unique Selections)**.** Given a type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$, we say that strategy $s_i \in S_i$ can be *uniquely selected* for type $t_i \in T_i$ if there exists a sequence $\{t_{i,m}\}_{m\in\mathbb{N}} \subseteq T_i^*$ such that $t_{i,m} \to t_i$, and $\{s_i\} = \text{EFR}_i(t_{i,m})$ for all m.

Note that we envision a type $t_i \in T_i$ as an element of T_i^* and identify it with $\varphi^*(t_i)$, i.e. the conjunction of its payoff type and belief hierarchy. An interpretation of this definition is that, if a strategy s_i can be uniquely selected for a type t_i, then s_i is the *only* EFR strategy for a type arbitrarily "close" to t_i. Therefore, a researcher cannot reject the strategy s_i being played by type t_i when she cannot precisely observe the infinite sequence of belief hierarchies.

The analyst can make a *prediction* $P = \times_{i\in I} P_i(\cdot)$ given a finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$. Each $P_i : T_i \to \mathcal{S}_i$ is a prediction for player i, where $\mathcal{S}_i$ is the collection of all nonempty subset of S_i. Following Chen et al. (2021), we now propose a

notion of robust refinement which requires that the prediction $P_i(t_i)$ of t_i coincides with an *upper hemicontinuous* refinement of EFR on T_i^*.[14]

Definition 3.5 (Robust Refinements)**.** Given a finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$, we say P is a *robust refinement* of EFR if there exists a prediction P^* on the universal type space such that for every $i \in I$

(i) $P_i^*(t_i^*) \subseteq \text{EFR}_i(t_i^*)$ for all $t_i^* \in T_i^*$;

(ii) $P_i^*(\cdot)$ is upper hemicontinuous on T_i^*;

(iii) $P_i(t_i) = P_i^*(\varphi_i^*(t_i))$ for all $t_i \in T_i$.

In words, a prediction P is a robust refinement of EFR if (i) there exists P^*, an upper hemicontinuous sub-correspondence of EFR^* on the universal type space, and (ii) the prediction P for the type space in question is consistent with P^*. Notice that by Lemmas 3.1 and 3.2, the prediction EFR is by default a robust refinement of itself. In the next section, we provide a tight condition for a prediction P to be robust refinement of EFR. We further use this condition to characterize the Structure Theorem and generic uniqueness of EFR.

3.3 Characterizations

For our characterization results, we employ a collection-based approach introduced by Chen et al. (2014, 2021). That is, we study collections of subsets of strategies and sequentially best responses against conjectures restricted by those collections. Although our characterization results bear a formal resemblance of those from Chen et al. (2021), we depart from their analysis in two ways. First, instead of studying static games, we focus on dynamic environments and use a solution concept that imposes restrictions on conjectures even at off-path histories. Second, we assume players can receive information from their payoff types and also allow for an information restriction as common knowledge among players. Therefore, the collection we study depends not only on the player, but also on the payoff type of this player.

[14]The idea of robust prediction captured by upper hemicontinuity originates from Fudenberg et al. (1988) and Dekel and Fudenberg (1990). In our setting, it requires the prediction $P_i(t_i)$ include at least some EFR strategies of the "true" type when the analyst treats t_i as a modeling tool and her observation of the belief hierarchy can be slightly imperfect. This interpretation relies on the adoption of the product topology on T_i^* and the meaning of two types being "close" in this topology. For a detailed discussion, see Weinstein and Yildiz (2007).

3.3.1 The Upper and Lower EFR Collections

Recall that $\mathcal{S}_i$ denotes the collection of all nonempty subsets of S_i. Our first goal is to compute the following two families of collections: The family of *upper EFR collections* $\left\{\mathcal{R}_i^{\uparrow}(\theta_i)\right\}_{\theta_i\in\Theta_i}$ of player i, where

$$\mathcal{R}_i^{\uparrow}(\theta_i) = \{R_i \in \mathcal{S}_i : \exists t_i \in T_i^* \text{ s.t. } \vartheta_i^*(t_i) = \theta_i \text{ and } R_i \supseteq \mathrm{EFR}_i(t_i)\},$$

and the family of *lower EFR collections* $\left\{\mathcal{R}_i^{\downarrow}(\theta_i)\right\}_{\theta_i\in\Theta_i}$ of player i, where

$$\mathcal{R}_i^{\downarrow}(\theta_i) = \{R_i \in \mathcal{S}_i : \exists t_i \in T_i^* \text{ s.t. } \vartheta_i^*(t_i) = \theta_i \text{ and } R_i \subseteq \mathrm{EFR}_i(t_i)\}.$$

Definition 3.6. Fix a payoff type $\theta_i \in \Theta_i$. For a given $\nu_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i})$, we say a distribution $\lambda_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i})$ is *consistent with* ν_i if there exists a function $f_i : \Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i} \to \Delta(S_{-i})$ such that

(i) $f_i(\theta_0, \theta_{-i}, R_{-i})[R_{-i}] = 1$ for every $R_{-i} \in \mathcal{S}_{-i}$, and

(ii) $\lambda_i[\theta_0, \theta_{-i}, s_{-i}] = \sum_{R_{-i}\in\mathcal{S}_{-i}} \nu_i[\theta_0, \theta_{-i}, R_{-i}] f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}]$.

Fixing $\nu_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i})$, for each $n \geq 1$, we define

$$\Pi_i^n(\nu_i \mid \theta_i) = \{\pi_i \in \mathrm{C}_i^n(\theta_i) : \pi_i(\phi) \text{ is consistent with } \nu_i\}.$$

To compute the upper EFR collection, let $\mathcal{R}_i^{\uparrow,0}(\theta_i) = \{S_i\}$ for each $\theta_i \in \Theta_i$ and $i \in I$. For $n \geq 1$, write $\mathcal{R}_{-i}^{\uparrow,n-1}(\theta_{-i}) = \times_{j\neq i} \mathcal{R}_j^{\uparrow,n-1}(\theta_j)$, and define recursively

$$\mathcal{R}_i^{\uparrow,n}(\theta_i) = \left\{ R_i \in \mathcal{S}_i : \begin{array}{l} \exists \nu_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i}) \text{ s.t.} \\ \text{(i) } \nu_i\Big[\Big\{(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow,n-1}(\theta_{-i})\Big\}\Big] = 1; \\ \text{(ii) } R_i \supseteq \bigcup_{\pi_i\in\Pi_i^n(\nu_i\mid\theta_i)} r_i(\pi_i \mid \theta_i) \end{array} \right\}. \tag{3.1}$$

Notice that the sequence $\mathcal{R}_i^{\uparrow,n}(\theta_i)$ is increasing in n, and converges in finitely many rounds for each θ_i of player i. To compute more efficiently, we observe that the only relevant strategy sets in computing $\mathcal{R}_i^{\uparrow,n}(\theta_i)$ are the *minimal* ones in $\mathcal{R}_{-i}^{\uparrow,n-1}(\theta_{-i})$; therefore, we only need to focus on the probability distributions ν_i that concentrate on the minimal sets in $\mathcal{R}_{-i}^{\uparrow,n-1}(\theta_{-i})$.

Similarly, to compute the lower EFR collection, we let $\mathcal{R}_i^{\downarrow,0}(\theta_i) = \mathcal{S}_i$ for each $\theta_i \in \Theta_i$ and $i \in I$. For $n \geq 1$, write $\mathcal{R}_{-i}^{\downarrow,n-1}(\theta_{-i}) = \times_{j \neq i} \mathcal{R}_j^{\downarrow,n-1}(\theta_j)$, and define recursively

$$\mathcal{R}_i^{\downarrow,n}(\theta_i) = \left\{ R_i \in \mathcal{S}_i : \begin{array}{c} \exists \nu_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i}) \text{ s.t.} \\ \text{(i) } \nu_i\Big[\Big\{(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\downarrow,n-1}(\theta_{-i})\Big\}\Big] = 1; \\ \text{(ii) } R_i \subseteq \bigcup_{\pi_i \in \Pi_i^n(\nu_i|\theta_i)} r_i(\pi_i \mid \theta_i) \end{array} \right\}.$$

This time, the sequence $\mathcal{R}_i^{\downarrow,n}(\theta_i)$ is decreasing in n, and converges in finitely many rounds for each θ_i of player i. Again, the only relevant sets in this procedure are the *maximal* ones.

Proposition 3.1. *For every $i \in I$ and every $\theta_i \in \Theta_i$, we have $\mathcal{R}_i^{\uparrow}(\theta_i) = \bigcup_{n \geq 0} \mathcal{R}_i^{\uparrow,n}(\theta_i)$, and $\mathcal{R}_i^{\downarrow}(\theta_i) = \bigcap_{n \geq 0} \mathcal{R}_i^{\downarrow,n}(\theta_i)$.*

Proof. See Appendix 3.5.2. □

Note that our characterizations of the upper and lower EFR collections do not rely on any type space. They are belief-free notions and can be computed directly from a given PI structure.

3.3.2 Unique Selections and Robust Predictions

We now focus our attention on an arbitrary *finite* type space $\mathcal{T} = \{(T_i, \vartheta_i, \kappa_i)_{i \in I}\}$. We slightly modify Definition 3.6 to adapt to the analysis of a given type space.

Definition 3.7. Fix a type $t_i \in T_i$. For a given $\nu_i \in \Delta(\Theta_0 \times T_{-i} \times \mathcal{S}_{-i})$ such that $\text{marg}_{\Theta_0 \times T_{-i}} \nu_i = \kappa_i(t_i)$, we say a distribution $\lambda_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times S_{-i})$ is *consistent with* ν_i if there exists a function $f_i : \Theta_0 \times T_{-i} \times \mathcal{S}_{-i} \to \Delta(S_{-i})$ such that

(i) $f_i(\theta_0, t_{-i}, R_{-i})[R_{-i}] = 1$,

(ii) $\lambda_i[\theta_0, \theta_{-i}, s_{-i}] = \sum_{t_{-i}: \vartheta_{-i}(t_{-i}) = \theta_{-i}} \sum_{R_{-i} \in \mathcal{S}_{-i}} \nu_i[\theta_0, t_{-i}, R_{-i}] f_i(\theta_0, t_{-i}, R_{-i})[s_{-i}]$.

Moreover, denote by $\Lambda_i^{\nu_i}(t_i)$ the set of all conjectures of type t_i that are consistent with ν_i.

Our next step is to characterize, for each (finite) type $t_i \in T_i$, the *local upper EFR collection* $\mathcal{R}_i^{\text{loc}}(t_i)$, which contains all sets of strategies that "curb" the rationalizable

behaviors of types in the neighborhood of t_i. Formally, for each type $t_i \in T_i$, let

$$\mathcal{R}_i^{\text{loc}}(t_i) = \{R_i \in \mathcal{S}_i : \exists \{t_{i,m}\}_{m\in\mathbb{N}} \subseteq T_i^* \text{ s.t. } t_{i,m} \to t_i \text{ and } R_i \supseteq \text{EFR}_i(t_{i,m})\ \forall m\}.$$

For given $\nu_i \in \Delta(\Theta_0 \times T_{-i} \times \mathcal{S}_{-i})$, $\tilde{\nu}_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i})$, and $\varepsilon \in (0,1]$, we define

$$\Lambda_i^n(\nu_i, \tilde{\nu}_i, \varepsilon \mid t_i) = \left\{ \pi_i \in \text{C}_i^n(\vartheta_i(t_i)) : \begin{array}{l} \exists \lambda_i, \tilde{\lambda}_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times S_{-i}) \text{ s.t.} \\ \text{(i) } \lambda_i \text{ is consistent with } \nu_i; \\ \text{(ii) } \tilde{\lambda}_i \text{ is consistent with } \tilde{\nu}_i; \\ \text{(iii) } \pi_i(\phi) = (1-\varepsilon)\lambda_i + \varepsilon\tilde{\lambda}_i \end{array} \right\}.$$

We now describe an iterative procedure to compute $\mathcal{R}_i^{\text{loc}}(t_i)$. For each $i \in I$ and $t_i \in T_i$, let $\mathcal{R}_i^{\text{loc},0}(t_i) = \mathcal{R}_i^{\uparrow}(\vartheta_i(t_i))$. For $n \geq 1$, let $\mathcal{R}_{-i}^{\text{loc},n-1}(t_{-i}) = \times_{j\neq i} \mathcal{R}_j^{\text{loc},n-1}(t_j)$, and define

$$\mathcal{R}_i^{\text{loc},n}(t_i) = \left\{ \begin{array}{l} R_i \in \mathcal{S}_i : \\ \quad \forall \varepsilon \in (0,1], \\ \quad \exists (\nu_i, \tilde{\nu}_i) \in \Delta(\Theta_0 \times T_{-i} \times \mathcal{S}_{-i}) \times \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i}) \text{ s.t.} \\ \quad \text{(i) } \text{marg}_{\Theta_0 \times T_{-i}} \nu_i = \kappa_i(t_i); \\ \quad \text{(ii) } \nu_i\Big[\Big\{(\theta_0, t_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\text{loc},n-1}(t_{-i})\Big\}\Big] = 1; \\ \quad \text{(iii) } \tilde{\nu}_i\Big[\Big\{(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow}(\theta_{-i})\Big\}\Big] = 1; \\ \quad \text{(iv) } R_i \supseteq \bigcup_{\pi_i \in \Lambda_i^n(\nu_i, \tilde{\nu}_i, \varepsilon \mid t_i)} r_i(\pi_i \mid \vartheta_i(t_i)). \end{array} \right\} \tag{3.2}$$

Each $\left\{\mathcal{R}_i^{\text{loc},n}(t_i)\right\}_{n\geq 0}$ is a decreasing sequence, and reaches its limit in finitely many rounds. Intuitively, at each round, ν_i captures a small perturbation of type t_i's belief hierarchy up to order n, while $\tilde{\nu}_i$ captures an arbitrary perturbation which is governed by an arbitrarily small probability ε.

We now prove that the limit of this profile characterizes the local upper EFR collections.

Proposition 3.2. *Fix a finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$. For each player $i \in I$ and type $t_i \in T_i$, we have $\mathcal{R}_i^{\text{loc}}(t_i) = \bigcap_{n\geq 0} \mathcal{R}_i^{\text{loc},n}(t_i)$.*

Proof. See Appendix 3.5.3. □

Using Proposition 3.2, we can further characterize unique selections and robust predictions for any finite type space.

Proposition 3.3 (Unique Selections)**.** *A strategy s_i can be uniquely selected for a finite type t_i if and only if $\{s_i\} \in \mathcal{R}_i^{\text{loc}}(t_i)$.*

Proof. By Definition 3.4 and the definition of $\mathcal{R}_i^{\text{loc}}(t_i)$. □

Proposition 3.4 (Robust Refinements)**.** *Given a finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$. A prediction P is a robust refinement if and only if for every $i \in I$ and $t_i \in T_i$, we have $P_i(t_i) \cap R_i \neq \varnothing$ for all $R_i \in \mathcal{R}_i^{\text{loc}}(t_i)$.*

Proof. See Appendix 3.5.4. □

3.3.3 The Structure Theorem and Generic Uniqueness

Our next task is to provide exact conditions on PI structures under which the structure theorem and generic uniqueness of EFR hold, respectively. To this end, we first collect all singletons in $\mathcal{R}_i^{\uparrow}(\theta_i)$ for each $\theta_i \in \Theta_i$, and let

$$R_i^{\text{u}}(\theta_i) = \left\{ s_i \in S_i : \{s_i\} \in \mathcal{R}_i^{\uparrow}(\theta_i) \right\}.$$

Each $R_i^{\text{u}}(\theta_i)$ is the set of strategies that are uniquely rationalizable for some $t_i \in T_i^*$ with payoff type $\vartheta^*(t_i) = \theta_i$.

Proposition 3.5 (The Structure Theorem)**.** *For a given PI structure, the following statements are equivalent:*

(1) For every finite type $t_i \in T_i^*$, any strategy $s_i \in \text{EFR}_i(t_i)$ can be uniquely selected for t_i;

(2) For every $i \in I$, $\theta_i \in \Theta_i$, and $R_i \in \mathcal{R}_i^{\downarrow}(\theta_i)$, we have $R_i \subseteq R_i^{\text{u}}(\theta_i)$.

Proof. See Appendix 3.5.5. □

Once we assume a PI structure, this characterization provides us a tool to determine whether a Structure Theorem of EFR holds by simply computing the families of upper EFR collections $\left\{\mathcal{R}_i^{\uparrow}(\theta_i)\right\}_{\theta_i\in\Theta_i}$ and lower EFR collections $\left\{\mathcal{R}_i^{\downarrow}(\theta_i)\right\}_{\theta_i\in\Theta_i}$ for every player i. Our condition is both sufficient and necessary: The Structure Theorem holds if and only if every strategy that is rationalizable for a type is uniquely rationalizable for

some type with the same payoff type. Note that when the richness condition of Penta (2012) holds, $R_i^{\mathrm{u}}(\theta_i) = S_i$ for all θ_i so condition (2) is satisfied; Moreover, EFR coincides with interim sequential rationalizability (ISR, Penta, 2012) due to unrestricted off-path conjectures. Therefore, Proposition 3.5 implies the Structure Theorem of ISR under richness.

The Structure Theorem of EFR generalizes the one of ICR in static games. Here, we emphasize a caveat in the interpretation of this generalization. In static games, a richer space of payoff uncertainty allows for more perturbations of higher order beliefs, but it does not affect the solution concept ICR. Therefore, when a researcher uses a finite type space to make predictions, she can make a statement as follows: "if richness is satisfied, the ICR strategies I computed for the type space I assume delivers the strongest robust prediction." However, the same statement *cannot* be made for EFR (and also ISR) in dynamic games. This is because EFR is sensitive to the common knowledge assumption carried by the PI structure, so a richer PI structure may change the EFR strategies that the researcher has computed.

Proposition 3.6 (Generic Uniqueness)**.** *For a given PI structure, the following statements are equivalent:*

(1) For every $i \in I$, the set $\mathcal{U}_i = \{t_i \in T_i^* : |\mathrm{EFR}_i(t_i)| = 1\}$ is open and dense in T_i^*;

(2) For every $i \in I$, $\theta_i \in \Theta_i$, and $R_i \in \mathcal{R}_i^{\uparrow}(\theta_i)$, we have $R_i \cap R_i^{\mathrm{u}}(\theta_i) \neq \varnothing$.

Proof. See Appendix 3.5.6. □

This result shows that the EFR correspondence on T^* is generically a singleton if and only if every set of rationalizable strategies of a type contains at least one strategy that is uniquely rationalizable for some type with the same payoff type. This is a strictly weaker condition than the one that characterizes the Structure Theorem. Importantly, the gap between these two conditions delineates a region where EFR admits a strongest robust proper refinement (see Section 3.4).

We now use an example to illustrate how to apply our characterization results.

Example 3.1. Consider the two-player two-stage game presented in Figure 3.1. Player 1 moves first and chooses to opt out (O) or enter the second stage (I). When player 1 enters, two players simultaneously choose either left or right, and the game ends. The only potential payoff uncertainty is player 1's payoff when the second stage is played,

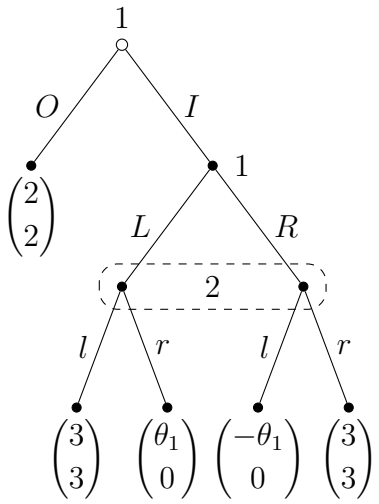


Figure 3.1: Example 3.1 — A two-stage game.

and it is common knowledge that player 1 knows his payoff. Therefore, both Θ_0 and Θ_2 are degenerate. For Θ_1, we consider three cases.

Case 1. First, suppose the analyst models this strategic interaction as a complete information game with $\Theta_1 = \{0\}$ and applies the solution concept EFR. In this case, the universal type space only contains one type for each player: $\vartheta_1^*\left(t_1^{\text{CB}}\right) = 0$, and for $i = 1, 2$,

$$T_i^* = \left\{t_i^{\text{CB}}\right\} \text{ and } \kappa_i\left(t_i^{\text{CB}}\right)\left[t_{-i}^{\text{CB}}\right] = 1.$$

It is easy to see that all strategies are extensive form rationalizable: For player 1, IL is the best response to ℓ, IR is the best response to r, and O is the best response to a conjecture that attaches equal probabilities to ℓ and r; For player 2, ℓ is the best response to IL, and r to IR. Therefore, $\text{EFR}_1\left(t_1^{\text{CB}}\right) = \{O, IL, IR\}$ and $\text{EFR}_2\left(t_2^{\text{CB}}\right) = \{\ell, r\}$. Moreover, since the universal type space is degenerate, any refinement of EFR is robust by definition.

Case 2. Now suppose the analyst still holds the assumption that there is common initial belief in $\theta_1 = 0$, but acknowledges the possibility that $\theta_1 = 4$. In other words, the analyst still employs the type space that only contains t_1^{CB} and t_2^{CB}, but the universal type space becomes larger due to the uncertainty $\Theta_1 = \{0, 4\}$. Using the argument in Case 1, we have $\text{EFR}_1\left(t_1^{\text{CB}}\right) = \{O, IL, IR\}$ and $\text{EFR}_2\left(t_2^{\text{CB}}\right) = \{\ell, r\}$.[15] The upper EFR

[15]Note that such invariance of EFR strategies to PI structures is not a general property. A type (i.e. payoff type and belief hierarchy) may have different EFR strategies when envisioned in two different

collections can be computed according to definition (3.1), and the *minimal elements* in the iterative procedure are shown in the following table

n	0	1	2	3	$\cdots$
$\mathcal{R}_1^{\uparrow,n}(\theta_1 = 0)$	S_1	S_1	S_1	$\{IL\}$	$\{IL\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1 = 4)$	S_1	$\{IL\}$	$\{IL\}$	$\{IL\}$	$\{IL\}$
$\mathcal{R}_2^{\uparrow,n}$	S_2	S_2	$\{\ell\}$	$\{\ell\}$	$\{\ell\}$

There are two key steps in the computation above. In the second round, we can pick a distribution $\nu_2 \in \Delta(\Theta_1 \times \mathcal{S}_1)$ for player 2 such that $\nu_2[\theta_1 = 4, \{IL\}] = 1$; therefore, $\{\ell\}$ becomes an element of $\mathcal{R}_2^{\uparrow,2}$. Then in the third round, we can pick a distribution $\nu_1 \in \Delta(\mathcal{S}_2)$ for player 1 such that $\nu_1[\{\ell\}] = 1$, which makes $\{IL\}$ an element of $\mathcal{R}_1^{\uparrow,3}(\theta_1 = 0)$. The process converges after four rounds.

Now by Proposition 3.6, the prediction of EFR is generically unique on the universal type space. In addition, we know that only IL is uniquely rationalizable for player 1, and ℓ for player 2. Therefore, $P_1\left(t_1^{\text{CB}}\right) = \{IL\}$ and $P_2\left(t_2^{\text{CB}}\right) = \{\ell\}$ is the strongest robust refinement of EFR for the types that the analyst has in mind.

Case 3. The analyst maintains the assumption of common initial belief in $\theta_1 = 0$, but also assumes common knowledge that $\Theta_1 = \{0, 4, -4\}$. As before, we have $\text{EFR}_1\left(t_1^{\text{CB}}\right) = \{O, IL, IR\}$ and $\text{EFR}_2\left(t_2^{\text{CB}}\right) = \{\ell, r\}$. Again, we summarize the iterative procedure of computing upper EFR collections by the minimal elements in each round

n	0	1	2	3	$\cdots$
$\mathcal{R}_1^{\uparrow,n}(\theta_1 = 0)$	S_1	S_1	S_1	$\{O\}, \{IL\}, \{IR\}$	$\{O\}, \{IL\}, \{IR\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1 = 4)$	S_1	$\{IL\}$	$\{IL\}$	$\{IL\}$	$\{IL\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1 = -4)$	S_1	$\{IR\}$	$\{IR\}$	$\{IR\}$	$\{IR\}$
$\mathcal{R}_2^{\uparrow,n}$	S_2	S_2	$\{\ell\}, \{r\}$	$\{\ell\}, \{r\}$	$\{\ell\}, \{r\}$

Note that $\{O\}$ is an element of $\mathcal{R}_1^{\uparrow,n}(\theta_1 = 0)$ from the third round, because it can be supported by a distribution $\nu_1 \in \Delta(\mathcal{S}_2)$ such that $\nu_1[\{\ell\}] = \nu_1[\{r\}] = \frac{1}{2}$.

On the other hand, the lower EFR collections can be seen from the facts that $\text{EFR}_1\left(t_1^{\text{CB}}\right) = S_1$, $\text{EFR}_2\left(t_2^{\text{CB}}\right) = S_2$, and player 1 has a dominant strategy when $\theta_1 = 4$ or -4. Specifically,

$$\mathcal{R}_1^{\downarrow}(\theta_1 = 0) = \mathcal{S}_1,\ \mathcal{R}_1^{\downarrow}(\theta_1 = 4) = \{IL\},\ \mathcal{R}_1^{\downarrow}(\theta_1 = -4) = \{IR\},\ \text{and } \mathcal{R}_2^{\downarrow} = \mathcal{S}_2$$

PI structures. See the discussion on *information invariance* in Penta (2012).

By Proposition 3.5, we conclude that the Structure Theorem holds for EFR; in other words, every rationalizable strategy can be uniquely selected for t_1^{CB} and t_2^{CB}. Therefore, any proper refinement of EFR is not a robust one.[16] ♢

3.4 Applications

In this section, we demonstrate how to use our characterization results in two applications. Our approach is to start with a model that we normally use in applied game theory, such as a finite type space or complete information. The situation is that we, as researchers, are confident with some aspects of our model, but are uncertain about others. We show that by relaxing common knowledge assumptions in a way that is suitable for our concern, robustness to higher order uncertainty can help us refine our predictions.

3.4.1 Privacy of Information

When modeling strategic interactions with incomplete information, we sometimes assume that a party receives a hard piece of information, while the other party is uninformed. In other words, the information is privately learned and exclusive to the owner. For example, in an auction with private values, the seller does not possess any information about bidders' valuations (except the prior distribution); in a signaling game, the "type" of a sender is private information and unavailable to the receiver. In this subsection, we consider predictions that are robust when we maintain the payoff implication of information, but perturb common knowledge about the *privacy* of such information.

We use the well-known Beer-Quiche game (Cho and Kreps, 1987) as an example to illustrate how to relax the common knowledge assumption on the privacy of information. Two players, a sender (player 1) and a receiver (player 2), move sequentially in a two-stage game. There are two types of the sender, which are payoff-relevant for both players. The sender can be wimpish or surly, with probability 0.1 or 0.9, respectively. In the original version of this game, the sender knows his type but the receiver does not, and this is common knowledge. We now extend the structure of information to include the possibility that sender's information is not private, i.e. his type is also known by the receiver. Let $\Theta_1 = \{\theta_1^w, \theta_1^s\}$, where superscripts indicate the type of the sender.

[16] The PI structure in Case 3 also provides an example where the richness condition of Penta (2012) is not satisfied, but the Structure Theorem does hold by our characterization result.

Moreover, let $\Theta_2 = \{\theta_2^p, \theta_2^w, \theta_2^s\}$, where θ_2^p implies the receiver knows nothing about the sender,[17] and θ_2^w or θ_2^s means the receiver learns the sender's type. There is an information restriction $\bar{\Theta}$ as follows.

	θ_1^w	θ_1^s
θ_2^p	1 is wimpish and 2 does not know	1 is surly and 2 does not know
θ_2^w	1 is wimpish and 2 knows	$\times$
θ_2^s	$\times$	1 is surly and 2 knows

Notice that some information pairs are mutually exclusive: For example, it cannot be the case that the sender is surly but the receiver knows he is wimpish.

We can now envision the original situation as a type space $\mathcal{T}^{\text{CB}}$ defined as

$$\begin{aligned} T_1 = \{t_1^w, t_1^s\},\ T_2 = \{t_2^p\},\\ \vartheta_1(t_1^w) = \theta_1^w,\ \vartheta_1(t_1^s) = \theta_1^s,\ \vartheta_2(t_2^p) = \theta_2^p,\\ \kappa_1(t_1^w)[t_2^p] = 1,\ \kappa_1(t_1^s)[t_2^p] = 1, \text{ and } \kappa_2(t_2^p)[t_1^w] = 1 - \kappa_2(t_2^p)[t_1^s] = 0.1. \end{aligned}$$

By employing this type space, the analyst assumes common initial belief in the privacy of sender's information. However, higher order uncertainty about information privacy is present when types in $\mathcal{T}^{\text{CB}}$ are perturbed in the universal type space.

Actions and payoffs are depicted in Figure 3.2. (Note that this figure is an ex ante description of the two-stage game while our solution concept is interim.) The dotted rectangle in the center corresponds to the original Beer-Quiche game, and the additional branches are cases where the sender's type is no longer private information.

EFR has weak predictive power for types in $\mathcal{T}^{\text{CB}}$. First consider the receiver t_2^p. If his initial conjecture is $\pi_2(\phi)[\theta_1^w, Q] = 1 - \pi_2(\phi)[\theta_1^s, B] = 0.1$, then NF (i.e. Not fight if Beer and Fight if Quiche) is the best response; if his initial conjecture is $\pi_2(\phi)[\theta_1^w, B] = 1 - \pi_2(\phi)[\theta_1^s, Q] = 0.1$, then FN is the best response; finally, if $\pi_2(\phi)[\theta_1^w, Q] = 0.1$ and $\pi_2(\phi)[\theta_1^s, Q] = \pi_2(\phi)[\theta_1^s, B] = 0.45$, then NN is the best response. In particular, FF is *never* a best response and thus can be removed in the first round. For either type of the sender, B is the best response if his initial conjecture is $\pi_1(\phi)[\theta_2^p, NF] = 1$, and Q is the best response if $\pi_1(\phi)[\theta_2^p, FN] = 1$. The iterative procedure converges in the second round, and we have

$$\text{EFR}_1(t_1^w) = \text{EFR}_1(t_1^s) = \{B, Q\}, \text{ and } \text{EFR}_2(t_2^p) = \{NF, FN, NN\}.$$

[17] Superscript p stands for *p*rivacy.

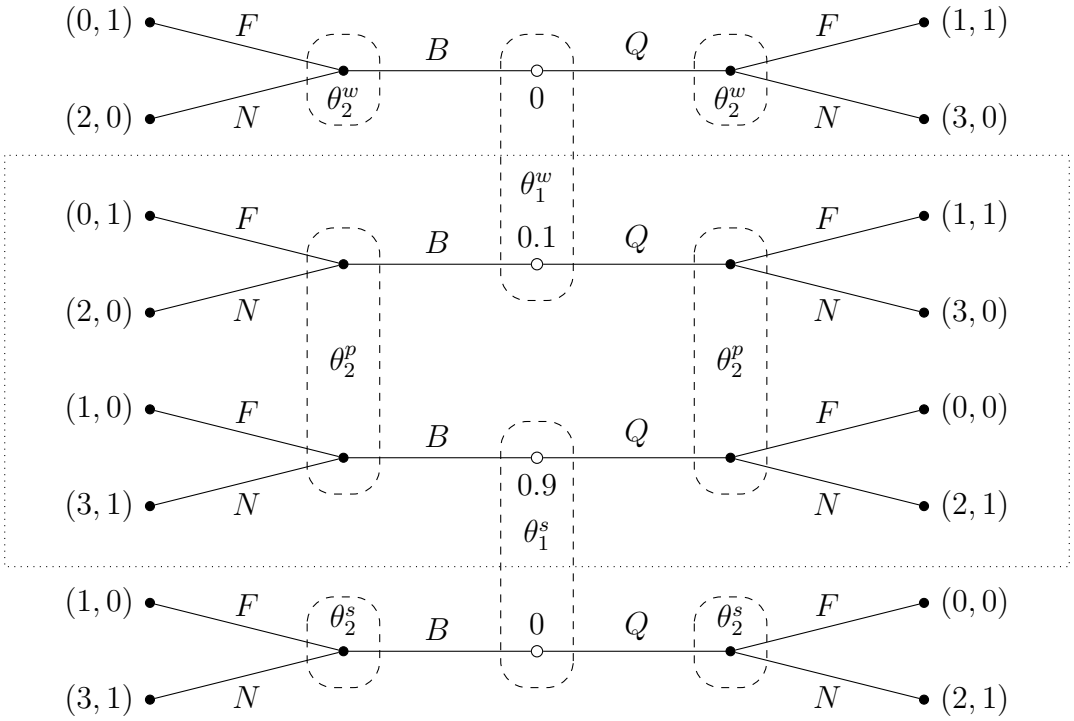


Figure 3.2: Beer-Quiche game with uncertainty about information privacy.

We next show that EFR has a robust refinement. To do so, we need to compute the upper EFR collections as in the following table. As before, only minimal elements are listed in each round.

n	0	1	2	3	4	$\cdots$
$\mathcal{R}_1^{\uparrow,n}(\theta_1^w)$	S_1	S_1	$\{Q\}$	$\{Q\}$	$\{B\},\{Q\}$	$\{B\},\{Q\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1^s)$	S_1	S_1	$\{B\}$	$\{B\}$	$\{B\}$	$\{B\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^p)$	S_2	S_1	S_1	$\{NF\}$	$\{NF\}$	$\{NF\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^w)$	S_2	$\{FF\}$	$\{FF\}$	$\{FF\}$	$\{FF\}$	$\{FF\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^s)$	S_2	$\{NN\}$	$\{NN\}$	$\{NN\}$	$\{NN\}$	$\{NN\}$

By Proposition 3.6, we know that EFR is generically unique on the universal type space, so it admits a proper robust refinement. Because $R_1^{\mathrm{u}}(\theta_1^s) = \{B\}$ and $R_2^{\mathrm{u}}(\theta_1^p) = \{NF\}$, the strongest robust refinement P satisfies

$$P_1(t_1^s) = \{B\}, \text{ and } P_2(t_2^p) = \{NF\}.$$

Now note that the minimal elements in $\mathcal{R}_2^{\mathrm{loc}}(t_2^p)$ and $\mathcal{R}_2^{\uparrow}(\theta_2^w)$ are $\{NF\}$ and $\{FF\}$,

respectively. Applying the definition of local upper EFR collection (3.2), we can see that the only minimal element in $\mathcal{R}_1^{\text{loc}}(t_1^w)$ is $\{B\}$,[18] which implies

$$P_1(t_1^w) = \{B\}.$$

Therefore, the strongest robust refinement of EFR predicts that both types of the sender play B in the first stage, and the receiver plays NF as a response. Notably, this prediction coincides with the only sequential equilibrium outcome that satisfies the Intuitive Criterion (Cho and Kreps, 1987). However, our argument is concerned with relaxing common knowledge of information in a particular way and higher order uncertainty due to such relaxation, but does not rely on the common knowledge of a fixed equilibrium outcome (see Cho and Kreps, 1987; Battigalli and Siniscalchi, 2002, 2003).

3.4.2 Observability of Actions

In games with imperfect information, the presence of an information set indicates that the player to whom this information set belongs cannot distinguish between the decision nodes in it. Implicitly, this is common knowledge among the players. It is then natural and interesting to investigate the consequences of perturbing such common knowledge assumption. Penta and Zuazo-Garin (2022) first formulate this question in static games. They transform the uncertainty over structure of games into payoff uncertainty and make predictions based on rationality and common belief in rationality. Because our framework generalizes theirs, and EFR captures rationality and common belief in rationality in static games, we can reproduce the results in Penta and Zuazo-Garin (2022) using Proposition 3.6.[19] In this subsection, we use an example to illustrate how to generalize their analysis to dynamic environments.

Consider the two-stage game depicted in Figure 3.3. In the second stage, two players play a coordination game where they choose H or L. (We follow the convention that player 1 is the row player.) There are two versions of the coordination game, and the players disagree on which stage game is better in terms of the *best* payoff they can obtain. Player 1, in the first stage, picks one of them. It can be checked that EFR has

[18] This comes from the fact that type t_1^w attaches probability 1 to type t_2^p whose payoff type is θ_2^p. If type t_1^w were to attach a high enough probability to another type of player 2 whose payoff type is θ_2^w, the strongest robust prediction for t_1^w would have been $\{Q\}$.

[19] An earlier version of this paper contains the proof.

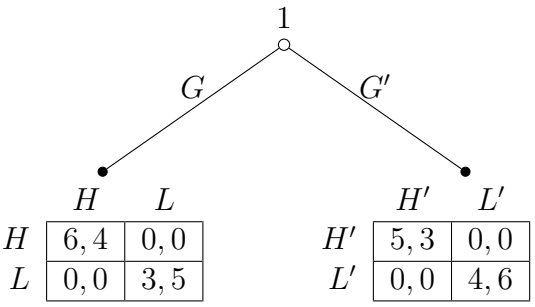


Figure 3.3: Selecting a coordination game by a "first-mover."

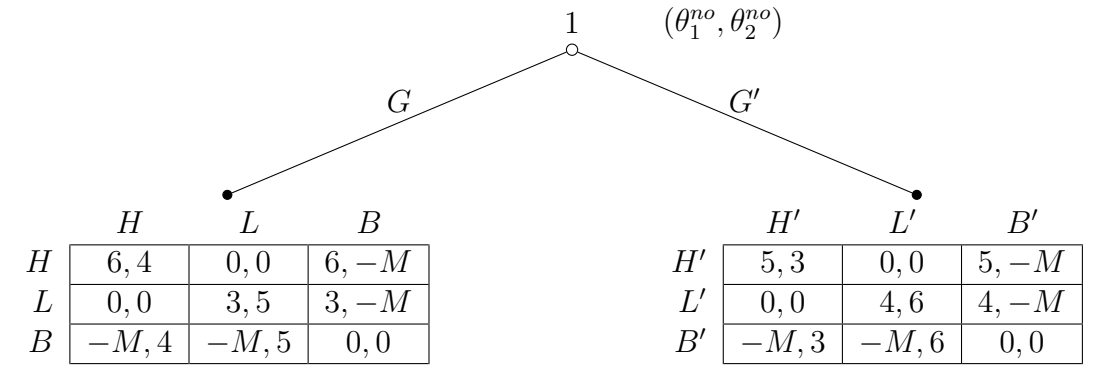
no predictive power in this game. For player 1, GH is the best response to HH', GL is the best response to $L(\frac{1}{2}H' + \frac{1}{2}L')$, $G'H'$ is to LH', and $G'L'$ is to LL'. Because all strategies of player 1 survive, we cannot remove any strategy for player 2: For example, HH' is the best response to $\frac{1}{2}GH + \frac{1}{2}G'H'$.

We now embed this game into a larger one which captures possibilities of observing opponent's action in the stage game. Suppose there are three cases: either no one can observe the opponent's action, or player 1 or 2 (knows he) has the ability to observe and respond contingently in the second stage.[20] Each player has two payoff types $\Theta_i = \{\theta_i^{no}, \theta_i^o\}$, where θ_i^{no} and θ_i^o imply the opponent's action is unobservable and observable, respectively. Since it cannot be the case that both players can observe, there is an information restriction as follows:

	θ_2^{no}	θ_2^o
θ_1^{no}	no one can observe	2 can observe
θ_1^o	1 can observe	$\times$

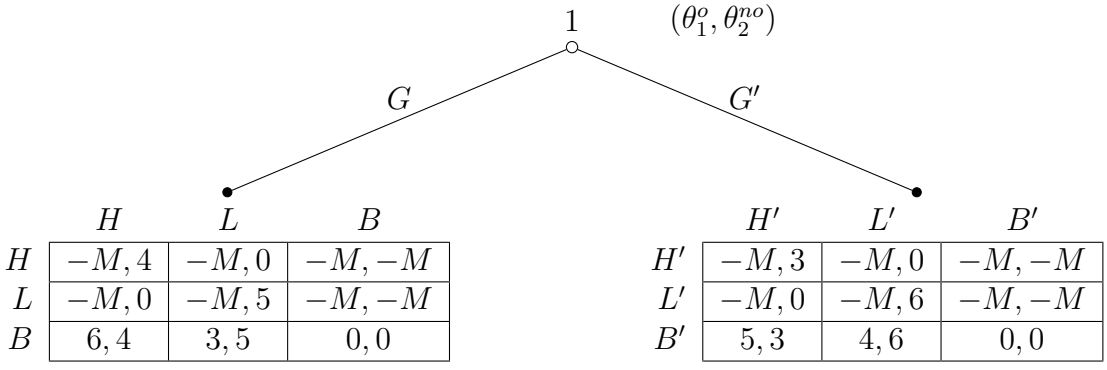
In order to fix the extensive form so that it does not depend on players' payoff types, we slightly modify the stage game and add an action B which represents the "backward induction action." Moreover, we modify the utility functions accordingly so that the action B is uniquely optimal (due to sequential rationality) whenever a player can observe opponent's action and respond contingently, and B is unavailable whenever a player cannot. Figure 3.4 describes players' payoffs in three different cases, where M is a sufficiently large number.

[20]For simplicity, we assume that if a player can observe the opponent's action, he is able to do so in both versions of the stage game, but the analysis can be generalized with no conceptual difficulty.



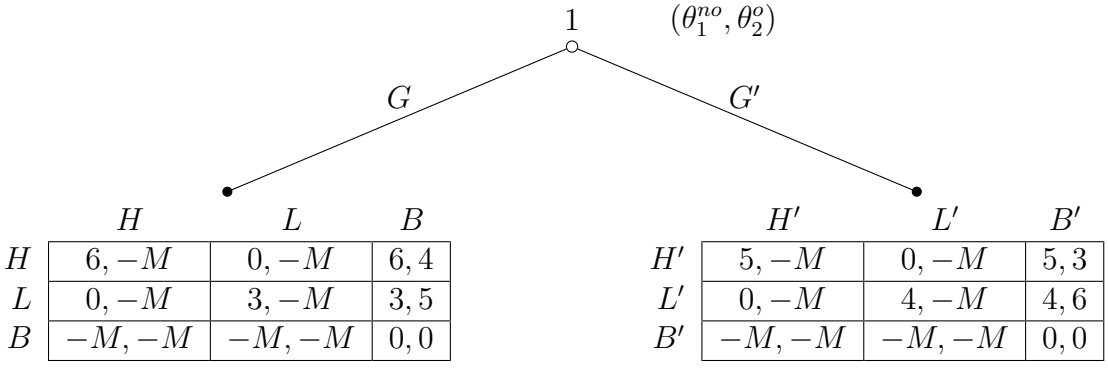
	H	L	B
H	$6,4$	$0,0$	$6,-M$
L	$0,0$	$3,5$	$3,-M$
B	$-M,4$	$-M,5$	$0,0$

	H'	L'	B'
H'	$5,3$	$0,0$	$5,-M$
L'	$0,0$	$4,6$	$4,-M$
B'	$-M,3$	$-M,6$	$0,0$



	H	L	B
H	$-M,4$	$-M,0$	$-M,-M$
L	$-M,0$	$-M,5$	$-M,-M$
B	$6,4$	$3,5$	$0,0$

	H'	L'	B'
H'	$-M,3$	$-M,0$	$-M,-M$
L'	$-M,0$	$-M,6$	$-M,-M$
B'	$5,3$	$4,6$	$0,0$



	H	L	B
H	$6,-M$	$0,-M$	$6,4$
L	$0,-M$	$3,-M$	$3,5$
B	$-M,-M$	$-M,-M$	$0,0$

	H'	L'	B'
H'	$5,-M$	$0,-M$	$5,3$
L'	$0,-M$	$4,-M$	$4,6$
B'	$-M,-M$	$-M,-M$	$0,0$

Figure 3.4: Selecting a coordination game with possibly observable actions.

The original game in Figure 3.3 can now be envisioned as a type space $\mathcal{T}^{\text{CB}}$ (in the enlarged game) that assumes common initial belief in the second stage being simultaneous. That is,

$$T_i = \{t_i^{no}\},\ \vartheta_i(t_i^{no}) = \theta_i^{no},\ \text{and}\ \kappa_i(t_i^{no})[t_{-i}^{no}] = 1,\ \text{for}\ i = 1, 2$$

We now use our characterization results to study robust predictions for this type space. We first compute the upper EFR collections in this game, and display the minimal elements in each round in the table below.

n	0	1	2	3
$\mathcal{R}_1^{\uparrow,n}(\theta_1^{no})$	S_1	$\{GH, GL, G'H', G'L'\}$	$\{GH\}$	$\{GH\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1^{o})$	S_1	$\{GB, G'B'\}$	$\{GB, G'B'\}$	$\{GB, G'B'\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^{no})$	S_2	$\{HH', HL', LH', LL'\}$	$\{HL', LH', LL'\}$	$\{HH', HL'\}, \{HL', LH', LL'\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^{o})$	S_2	$\{BB'\}$	$\{BB'\}$	$\{BB'\}$

n	4	5
$\mathcal{R}_1^{\uparrow,n}(\theta_1^{no})$	$\{GH\}$	$\{GH\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1^{o})$	$\{GB\}$	$\{GB\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^{no})$	$\{HH', HL'\}, \{HL', LH', LL'\}$	$\{HH', HL'\}, \{LH', LL'\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^{o})$	$\{BB'\}$	$\{BB'\}$

n	6	7	$\cdots$
$\mathcal{R}_1^{\uparrow,n}(\theta_1^{no})$	$\{GH\}$	$\{GH\}$	$\{GH\}$
$\mathcal{R}_1^{\uparrow,n}(\theta_1^{o})$	$\{GB\}, \{G'B'\}$	$\{GB\}, \{G'B'\}$	$\{GB\}, \{G'B'\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^{no})$	$\{HH', HL'\}, \{LH', LL'\}$	$\{HL'\}, \{LL'\}$	$\{HL'\}, \{LL'\}$
$\mathcal{R}_2^{\uparrow,n}(\theta_2^{o})$	$\{BB'\}$	$\{BB'\}$	$\{BB'\}$

By Proposition 3.6, the prediction delivered by EFR is generically unique in the universal type space. The strongest robust refinement P for the type space $\mathcal{T}^{\text{CB}}$ is such that[21]

$$P_1(t_1^{no}) = \{GH\},\ \text{and}\ P_2(t_2^{no}) = \{HL'\}.$$

Perhaps surprisingly, not only can two players achieve perfect coordination in the second stage, but it is *as if* player 1 has a "first-mover advantage" so that he can select the coordination outcome he prefers the most. Such a strong prediction is not implied by forward induction reasoning, because GL is not strictly dominated as it is a best response to $L(\frac{1}{2}H' + \frac{1}{2}L')$.

[21]For type t_1^{no}, this comes from the fact that $R_1^{\text{u}}(\theta_1^{no}) = \{GH\}$. For type t_2^{no}, we note that the only minimal element in $\mathcal{R}_1^{\text{loc}}(t_1^{no})$ is $\{GH\}$ and $\kappa_2(t_2^{no})[t_1^{no}] = 1$; applying definition (3.2) of local upper EFR collections yields the unique selection.

Remark. The generic uniqueness of EFR in this example depends on the fact that player 1's payoff (3) from coordinating on (L, L) in version G is lower than both 5 and 4, i.e. player 1's payoffs when two players coordinate in version G'. If the former payoff is higher than 4, the procedure computing upper EFR collections converges after 5 rounds, and there is robust multiplicity in the universal type space. However, such multiplicity only comes from player 2's action in the coordination game G': In a strongest robust refinement P for the type space $\mathcal{T}^{\text{CB}}$, we have $P_1(t_1^{no}) = \{GH\}$, but $P_2(t_2^{no}) = \{HH'\}$ or $P_2(t_2^{no}) = \{HL'\}$ depending on how we choose to refine. The strongest robust *outcome* is always such that player 1 chooses version G and obtains the highest payoff by coordinating on (H, H).

3.5 Appendix

3.5.1 Proof of Lemma 3.1

We prove this lemma in two steps.

Claim 3.1. *For all $n \geq 0$, $s_i \in \text{BF}_i^n(\theta_i)$ if and only if there exists $t_i \in T_i^*$ such that $\vartheta_i^*(t_i) = \theta_i$ and $s_i \in \text{EFR}_i^{*,n}(t_i)$.*

Proof. By definition, $\text{BF}_i^0(\theta_i) = \text{EFR}_i^{*,0}(t_i) = S_i$ for all θ_i and t_i. Now suppose the statement hold for all $k = 0, 1, \ldots, n-1$, and we prove them for n.

Take $s_i \in \text{BF}_i^n(\theta_i)$. By definition, there exists $\pi_i \in \text{C}_i^n(\theta_i) \in \Delta^{\mathcal{H}}(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i})$ such that

(i) For all $h \in \mathcal{H}$ and $k = 0, 1, \ldots, n-1$,

$$\Big(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i}(h)\Big) \cap \text{graph}\left(\text{BF}_{-i}^{k-1}\right) \neq \varnothing \;\Rightarrow\; \pi_i(h)\left[\Theta_0 \times \text{graph}\left(\text{BF}_{-i}^{k-1}\right)\right] = 1;$$

(ii) $s_i \in r_i(\pi_i \mid \theta_i)$.

By induction hypothesis, for every $k = 0, 1, \ldots, n-1$ and $s_{-i} \in \text{BF}_{-i}^{k-1}(\theta_{-i})$, there is a type profile $y_{-i}^{(\theta_{-i}, s_{-i}, k)} \in T_{-i}^*$ such that $\vartheta_{-i}^*\Big(y_{-i}^{(\theta_{-i}, s_{-i}, k)}\Big) = \theta_{-i}$ and $s_{-i} \in \text{EFR}_{-i}^{*,k}\Big(y_{-i}^{(\theta_{-i}, s_{-i}, k)}\Big)$. Now define a type $t_i \in T_i^*$ as

$$\kappa_i^*(t_i)[\theta_0, y_{-i}] = \pi_i(\phi)\left[(\theta_0, \theta_{-i}, s_{-i}) : y_{-i}^{(\theta_{-i}, s_{-i}, n-1)} = y_{-i}\right].$$

For this type t_i, define $\tilde{\mu}_i \in \Delta^{\mathcal{H}}(\Theta_0 \times \bar{T}^*_{-i}(t_i) \times S_{-i})$ as follows: For all $k = 0, 1, \ldots, n-1$ and $h \in \mathcal{H}$ such that $\left(\Theta_0 \times \bar{T}^*_{-i}(t_i) \times S_{-i}(h)\right) \cap \text{graph}\left(\text{EFR}^{*,k}_{-i}\right) \neq \varnothing$, let

$$\tilde{\mu}_i(h)[\theta_0, y_{-i}, s_{-i}] = \pi_i(h)\left[(\theta_0, \theta_{-i}, s_{-i}) : y^{(\theta_{-i}, s_{-i}, k)}_{-i} = y_{-i}\right];$$

for all other histories h, arbitrarily fix a $\hat{y}^{\theta_{-i}}_{-i} \in T^*_{-i}$ such that $\vartheta^*_{-i}\left(\hat{y}^{\theta_{-i}}_{-i}\right) = \theta_{-i}$ for each θ_{-i}, and let

$$\tilde{\mu}_i(h)[\theta_0, \hat{y}^{\theta_{-i}}_{-i}, s_{-i}] = \pi_i(h)[\theta_0, \theta_{-i}, s_{-i}].$$

By construction, $\tilde{\mu}_i$ is a well-defined CPS, and satisfies

(i') $\text{marg}_{\Theta_0 \times T^*_{-i}} \tilde{\mu}_i(\phi) = \kappa^*_i(t_i)$;

(ii') For all $h \in \mathcal{H}$ and $k = 0, 1, \ldots, n-1$,

$$\left(\Theta_0 \times \bar{T}^*_{-i}(t_i) \times S_{-i}(h)\right) \cap \text{graph}\left(\text{EFR}^{*,k}_{-i}\right) \neq \varnothing \;\Rightarrow\; \tilde{\mu}_i(h)\left[\Theta_0 \times \text{graph}\left(\text{EFR}^{*,k}_{-i}\right)\right] = 1;$$

(iii') For all $h \in \mathcal{H}$, $\pi_i(h) = \text{marg}_{\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times S_{-i}} \tilde{\mu}_i(h)$.

These facts, recursively, imply that $\pi_i \in \Psi^k_i(t_i)$ for all $k = 0, 1, \ldots, n$. Because $s_i \in r_i(\pi_i \mid \theta_i)$, we conclude that $s_i \in \text{EFR}^{*,n}_i(t_i)$.

For the converse, take $t_i \in T^*_i$ such that $\vartheta^*_i(t_i) = \theta_i$ and $s_i \in \text{EFR}^{*,n}_i(t_i)$. Let $\pi_i \in \Psi^{*,n}_i(t_i)$ be the conjecture such that $s_i \in r_i(\pi_i \mid \theta_i)$; Moreover, there exists $\tilde{\mu}_i \in \Delta^{\mathcal{H}}(\Theta_0 \times \bar{T}^*_{-i}(t_i) \times S_{-i})$ such that conditions (ii') and (iii') above are satisfied. By induction hypothesis and (iii'), we have

$$\begin{aligned} &\left(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times S_{-i}(h)\right) \cap \text{graph}\left(\text{BF}^{k-1}_{-i}\right) \neq \varnothing \\ &\quad \Rightarrow \quad \left(\Theta_0 \times \bar{T}^*_{-i}(t_i) \times S_{-i}(h)\right) \cap \text{graph}\left(\text{EFR}^{*,k}_{-i}\right) \neq \varnothing, \end{aligned}$$

and

$$\tilde{\mu}_i(h)\left[\Theta_0 \times \text{graph}\left(\text{EFR}^{*,k}_{-i}\right)\right] = 1 \;\Rightarrow\; \pi_i(h)\left[\Theta_0 \times \text{graph}\left(\text{BF}^{k-1}_{-i}\right)\right] = 1,$$

for all $h \in \mathcal{H}$ and $k = 0, 1, \ldots, n-1$. Combining these with (ii') implies $\pi_i \in \text{C}^n_i(\theta_i)$. But $s_i \in r_i(\pi_i \mid \theta_i)$, so we conclude that $s_i \in \text{BF}^n_i(\theta_i)$. □

Claim 3.2. *Fix a type space* $\{(T_i, \vartheta_i, \kappa_i)_{i \in I}\}$. *For every* $t_i \in T_i$ *and* $n \geq 0$, $\Psi^n_i(t_i) = \Psi^{*,n}_i(\varphi^*_i(t_i))$. *This in turn implies* $\text{EFR}^n_i(t_i) = \text{EFR}^{*,n}_i(\varphi^*_i(t_i))$.

Proof. The proof is by induction. To show $\Psi_i^n(t_i) = \Psi_i^{*,n}(\varphi_i^*(t_i))$, we can adapt the proof from Dekel et al. (2007, Lemma 1) (with suitable modifications) for conjectures at the empty history ϕ. At all off-path histories, we can then apply Claim 3.1 to show equivalence. We omit the details. □

3.5.2 Proof of Proposition 3.1

We first prove a lemma.

Lemma 3.3. *For any $n \geq 0$, we have*

$$\mathcal{R}_i^{\uparrow,n}(\theta_i) = \{R_i \in \mathcal{S}_i : \exists t_i \in T_i^* \text{ s.t. } \vartheta_i^*(t_i) = \theta_i \text{ and } R_i \supseteq \mathrm{EFR}_i^n(t_i)\},$$

and

$$\mathcal{R}_i^{\downarrow,n}(\theta_i) = \{R_i \in \mathcal{S}_i : \exists t_i \in T_i^* \text{ s.t. } \vartheta_i^*(t_i) = \theta_i \text{ and } R_i \subseteq \mathrm{EFR}_i^n(t_i)\},$$

Proof. We focus on proving the first equality. The proof for the second one is analogous, and hence we omit. We prove by induction. Since $\mathrm{EFR}_i^0(t_i) = S_i$ for any $t_i \in T_i^*$ by definition, the case for $n = 0$ is obvious. Suppose the equality holds for $n - 1$.

For the "$\subseteq$" direction, take any $R_i \in \mathcal{R}_i^{\uparrow,n}(\theta_i)$. Therefore, there exists $\nu_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i})$ such that $\nu_i\Big[(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow,n-1}(\theta_{-i})\Big] = 1$ and $R_i \supseteq \bigcup_{\pi_i \in \Pi_i^n(\nu_i|\theta_i)} r_i(\pi_i \mid \theta_i)$. By the induction hypothesis, for every $R_{-i} \in \mathcal{R}_{-i}^{\uparrow,n-1}(\theta_{-i})$, there exists $t_{-i}^{(\theta_{-i},R_{-i})} \in T_{-i}^*$ such that $\vartheta_{-i}^*\Big(t_{-i}^{(\theta_{-i},R_{-i})}\Big) = \theta_{-i}$ and $R_{-i} \supseteq \mathrm{EFR}_{-i}^{n-1}\Big(t_{-i}^{(\theta_{-i},R_{-i})}\Big)$. Define a type $t_i \in T_i^*$ by letting $\vartheta_i^*(t_i) = \theta_i$ and

$$\kappa_i^*(t_i)[\theta_0, t_{-i}] = \nu_i\left(\left\{(\theta_0, \theta_{-i}, R_{-i}) : t_{-i}^{(\theta_{-i},R_{-i})} = t_{-i}\right\}\right).$$

This type is well-defined because ν_i only puts positive probability on $\theta_{-i} \in \bar{\Theta}_{-i}(\theta_i)$. We next show that $R_i \supseteq \mathrm{EFR}_i^n(t_i)$. Take any $s_i \in \mathrm{EFR}_i^n(t_i)$. By definition 3.3, this implies that $s_i \in r_i(\pi_i \mid \vartheta_i^*(t_i))$ for some $\pi_i \in \Psi_i^n(t_i)$. Let $\mu_i \in \Delta(\Theta_0 \times T_{-i}^* \times S_{-i})$ be such that

(i) $\mathrm{marg}_{\Theta_0 \times T_{-i}} \mu_i = \kappa_i^*(t_i)$;

(ii) $\mu_i\left[\left\{(\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \mathrm{EFR}_{-i}^{n-1}(t_{-i})\right\}\right] = 1$;

(iii) $\pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}^*(t_{-i}) = \theta_{-i}\}]$.

We now show that $\pi_i \in \Pi_i^n(\nu_i \mid \theta_i)$. Define f_i as the conditional probability of μ_i on each $\left(\theta_0, t_{-i}^{(\theta_{-i}, R_{-i})}\right)$, i.e.

$$f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}] = \mu_i\left[s_{-i} \mid \theta_0, t_{-i}^{(\theta_{-i}, R_{-i})}\right].$$

Condition (i) in Definition 3.6 is satisfied because $\mu_i\left[s_{-i} \in \mathrm{EFR}_{-i}^{n-1}(t_{-i})\right] = 1$ and $R_{-i} \supseteq \mathrm{EFR}_{-i}^{n-1}\left(t_{-i}^{(\theta_{-i}, R_{-i})}\right)$ for every $R_{-i} \in \mathcal{R}_{-i}^{\uparrow, n-1}(\theta_{-i})$ by the induction hypothesis. The initial conjecture $\pi_i(\phi)$ is consistent with ν_i (Definition 3.6) because

$$\begin{aligned}
\pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] &= \sum_{t_{-i}: \vartheta_{-i}^*(t_{-i}) = \theta_{-i}} \kappa_i^*(t_i)[\theta_0, t_{-i}]\mu_i[s_{-i} \mid \theta_0, t_{-i}] \\
&= \sum_{R_{-i} \in \mathcal{R}_{-i}^{\uparrow, n-1}(\theta_{-i})} \kappa_i^*(t_i)\left[\theta_0, t_{-i}^{(\theta_{-i}, R_{-i})}\right]\mu_i\left[s_{-i} \mid \theta_0, t_{-i}^{(\theta_{-i}, R_{-i})}\right] \\
&= \sum_{R_{-i} \in \mathcal{S}_{-i}} \nu_i[\theta_0, \theta_{-i}, R_{-i}] f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}].
\end{aligned}$$

Thus, we have $\pi_i \in \Pi_i^n(\nu_i \mid \theta_i)$, which implies $s_i \in r_i(\pi_i \mid \theta_i) \subseteq R_i$. We can conclude that $R_i \supseteq \mathrm{EFR}_i^n(t_i)$.

For the converse direction "$\supseteq$", take $R_i \in \mathcal{S}_i$ such that there is a type $t_i \in T_i^*$ for which $\vartheta_i^*(t_i) = \theta_i$ and $R_i \supseteq \mathrm{EFR}_i^n(t_i)$. Define $\nu_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i})$ as

$$\nu_i[\theta_0, \theta_{-i}, R_{-i}] = \kappa_i^*(t_i)\left[(\theta_0, t_{-i}) : \vartheta_{-i}^*(t_{-i}) = \theta_{-i} \text{ and } \mathrm{EFR}_{-i}^{n-1}(t_{-i}) = R_{-i}\right].$$

By the induction hypothesis, we have $\nu_i\left[(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow, n-1}(\theta_{-i})\right] = 1$. It is left to show that for any $\pi_i \in \Pi_i^n(\nu_i \mid \theta_i)$, we have $R_i \supseteq r_i(\pi_i \mid \theta_i)$. Suppose $\pi_i \in \Pi_i^n(\nu_i \mid \theta_i)$, so there exists a function $f_i : \Theta_0 \times \bar{\Theta}_{-i}(\theta_i) \times \mathcal{S}_{-i} \to \Delta(S_{-i})$ such that

(i) $f_i(\theta_0, \theta_{-i}, R_{-i})[R_{-i}] = 1$,

(ii) $\pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \sum_{R_{-i} \in \mathcal{S}_{-i}} \nu_i[\theta_0, \theta_{-i}, R_{-i}] f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}]$.

Define a distribution $\mu_i \in \Delta(\Theta_0 \times T_{-i}^* \times S_{-i})$ as follows: For every measurable $E_{-i} \subseteq T_{-i}^*$, let

$$\begin{aligned}
&\mu_i\left[\{\theta_0\} \times E_{-i} \times \{s_{-i}\}\right] \\
&= \sum_{R_{-i} \in \mathcal{S}_{-i}} \kappa_i^*(t_i)\left[\{(\theta_0, t_{-i}) : t_{-i} \in E_{-i} \text{ and } \mathrm{EFR}_{-i}^{n-1}(t_{-i}) = R_{-i}\}\right] f_i(\theta_0, \vartheta_{-i}^*(t_{-i}), R_{-i})[s_{-i}].
\end{aligned}$$

It follows that $\text{marg}_{\Theta_0 \times T^*_{-i}} \mu_i = \kappa^*_i(t_i)$ and $\mu_i \left[s_{-i} \in \text{EFR}^{n-1}_{-i}(t_{-i}) \right] = 1$. Moreover, we have

$$\begin{aligned}
&\mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta^*_{-i}(t_{-i}) = \theta_{-i}\}] \\
&= \sum_{R_{-i} \in \mathcal{S}_{-i}} \kappa^*_i(t_i) \left[(\theta_0, t_{-i}) : \vartheta^*_{-i}(t_{-i}) = \theta_{-i} \text{ and } \text{EFR}^{n-1}_{-i}(t_{-i}) = R_{-i} \right] f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}] \\
&= \sum_{R_{-i} \in \mathcal{S}_{-i}} \nu_i(\theta_0, \theta_{-i}, R_{-i}) f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}] \\
&= \pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}],
\end{aligned}$$

which means $\pi_i \in \Psi^n_i(t_i)$. By Definition 3.3, we know $\text{EFR}^n_i(t_i) \supseteq r_i(\pi_i \mid \theta_i)$. Since $R_i \supseteq \text{EFR}^n_i(t_i)$, we conclude $R_i \supseteq r_i(\pi_i \mid \theta_i)$ for arbitrary $\pi_i \in \Pi^n_i(\nu_i \mid \theta_i)$. Hence, $R_i \in \mathcal{R}^{\uparrow,n}_i(\theta_i)$. □

We now turn to the proof of Proposition 3.1.

Proof of Proposition 3.1. (i) $\mathcal{R}^{\uparrow}_i(\theta_i) = \bigcup_{n \geq 0} \mathcal{R}^{\uparrow,n}_i(\theta_i)$: For "$\subseteq$", let $R_i \in \mathcal{R}^{\uparrow}_i(\theta_i)$. Then there exist $t_i \in T^*_i$ and $m \in \mathbb{N}$ such that $\vartheta^*_i(t_i) = \theta_i$ and $R_i \supseteq \text{EFR}_i(t_i) = \text{EFR}^m_i(t_i)$. By Lemma 3.3, we have $R_i \in \mathcal{R}^{\uparrow,m}_i(\theta_i)$. For "$\supseteq$", take $R_i \in \mathcal{R}^{\uparrow,n}_i(\theta_i)$ for some n. By Lemma 3.3, there exists $t_i \in T^*_i$ such that $\vartheta^*_i(t_i) = \theta_i$ and $R_i \supseteq \text{EFR}^n_i(t_i) \supseteq \text{EFR}_i(t_i)$. Hence, $R_i \in \mathcal{R}^{\uparrow}_i(\theta_i)$.

(ii) $\mathcal{R}^{\downarrow}_i(\theta_i) = \bigcap_{n \geq 0} \mathcal{R}^{\downarrow,n}_i(\theta_i)$: For "$\subseteq$", let $R_i \in \mathcal{R}^{\downarrow}_i(\theta_i)$. Then there exists $t_i \in T^*_i$ such that $\vartheta^*_i(t_i) = \theta_i$ and $R_i \subseteq \text{EFR}_i(t_i) \subseteq \text{EFR}^n_i(t_i)$ for all n. By Lemma 3.3, we have $R_i \in \mathcal{R}^{\downarrow,n}_i(\theta_i)$ for all n. For "$\supseteq$", take $R_i \in \bigcap_{n \geq 0} \mathcal{R}^{\downarrow,n}_i(\theta_i)$. By Lemma 3.3, there exists a sequence $\{t_{i,n}\} \subseteq T^*_i$ such that $\vartheta^*_i(t_{i,n}) = \theta_i$ and $R_i \subseteq \text{EFR}^n_i(t_{i,n})$ for each n. Since T^*_i is a compact metric space, there exists a convergent subsequence $\{t_{i,n_k}\}$, and let t_i denote its limit. Fix any n, and therefore for all $n_k \geq n$, we have $R_i \subseteq \text{EFR}^{n_k}_i(t_{i,n_k}) \subseteq \text{EFR}^n_i(t_{i,n_k})$. Since EFR^n_i is upper hemicontinuous for each n and $t_{i,n_k} \to t_i$, we have $R_i \subseteq \text{EFR}^n_i(t_i)$. This is true for all n, and thus we conclude that $R_i \subseteq \text{EFR}_i(t_i)$, which implies $R_i \in \mathcal{R}^{\downarrow}_i(\theta_i)$. □

3.5.3 Proof of Proposition 3.2

Fixing a type $t_i \in T_i$ in the finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i \in I}\}$, we split the statement of Proposition 3.2 into two lemmas and prove them in order.

Lemma 3.4. $\mathcal{R}^{\text{loc}}_i(t_i) \supseteq \bigcap_{n \geq 0} \mathcal{R}^{\text{loc},n}_i(t_i)$.

Proof. Let $\tilde{\mathcal{R}}_i^{\mathrm{loc},0}(t_i) = \mathcal{R}_i^{\uparrow}(\vartheta_i(t_i))$, and define for each n,

$$\tilde{\mathcal{R}}_i^{\mathrm{loc},n}(t_i) = \left\{ R_i \in \mathcal{S}_i : \begin{array}{l} \exists \{t_{i,m}\}_{m\in\mathbb{N}} \subset T_i^* \text{ s.t.} \\ \text{(i) } \vartheta_i^*(t_{i,m}) = \vartheta_i(t_i)\ \forall m, \text{ and } \tau_i^n(t_{i,m}) \to \tau_i^n(t_i) \text{ as } m \to \infty \\ \text{(ii) } R_i \supseteq \mathrm{EFR}_i(t_{i,m})\ \forall m \end{array} \right\}.$$

It suffices to show that $\tilde{\mathcal{R}}_i^{\mathrm{loc},n}(t_i) \supseteq \mathcal{R}_i^{\mathrm{loc},n}(t_i)$ for all n, and then $\mathcal{R}_i^{\mathrm{loc}}(t_i) \supseteq \bigcap_{n\geq 0} \mathcal{R}_i^{\mathrm{loc},n}(t_i)$ is implied by taking a diagonal sequence.

We prove by induction. For $n = 0$, we have $\tilde{\mathcal{R}}_i^{\mathrm{loc},0}(t_i) = \mathcal{R}_i^{\uparrow}(\vartheta_i(t_i)) \supseteq \mathcal{R}_i^{\mathrm{loc},0}(t_i)$ by definition. Suppose the inclusion holds for $n-1$, i.e. $\tilde{\mathcal{R}}_i^{\mathrm{loc},n-1}(t_i) \supseteq \mathcal{R}_i^{\mathrm{loc},n-1}(t_i)$ for all $i \in I$ and all $t_i \in T_i$. Take $R_i \in \mathcal{R}_i^{\mathrm{loc},n}(t_i)$, and we need to show that $R_i \in \tilde{\mathcal{R}}_i^{\mathrm{loc},n}(t_i)$. By definition of $\mathcal{R}_i^{\mathrm{loc},n}(t_i)$, for each $m \in \mathbb{N}$, there exists $(\nu_{i,m}, \tilde{\nu}_{i,m}) \in \Delta(\Theta_0 \times T_{-i} \times \mathcal{S}_{-i}) \times \Delta\left(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i}\right)$ such that

(i) $\mathrm{marg}_{\Theta_0 \times T_{-i}} \nu_{i,m} = \kappa_i(t_i)$;

(ii) $\nu_{i,m}[\{(\theta_0, t_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\mathrm{loc},n-1}(t_{-i})\}] = 1$;

(iii) $\tilde{\nu}_{i,m}\left[(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow}(\theta_{-i})\right] = 1$;

(iv) $R_i \supseteq \bigcup_{\pi_i \in \Lambda_i^n(\nu_i, \tilde{\nu}_i, \frac{1}{m+1}|t_i)} r_i(\pi_i \mid \vartheta_i(t_i))$.

For each $R_{-i} \in \mathcal{R}_{-i}^{\uparrow}(\theta_{-i})$ where $\theta_{-i} \in \bar{\Theta}_{-i}(\vartheta_i(t_i))$, there exists $y_{-i}^{(\theta_{-i},R_{-i})} \in T_{-i}^*$ such that $\vartheta_{-i}^*\left(y_{-i}^{(\theta_{-i},R_{-i})}\right) = \theta_{-i} \in \bar{\Theta}_{-i}(\vartheta_i(t_i))$ and $R_{-i} \supseteq \mathrm{EFR}_{-i}\left(y_{-i}^{(\theta_{-i},R_{-i})}\right)$. For each $t_{-i} \in T_{-i}$ and $R_{-i} \in \mathcal{R}_{-i}^{\mathrm{loc},n-1}(t_{-i})$, by the induction hypothesis, there exists a sequence $\left\{y_{-i,m}^{(t_{-i},R_{-i})}\right\}_{m\in\mathbb{N}} \subseteq T_{-i}^*$ such that $\vartheta_{-i}^*\left(y_{-i,m}^{(t_{-i},R_{-i})}\right) = \vartheta_{-i}(t_{-i})$ for all m, $\tau_{-i}^{n-1}\left(y_{-i,m}^{(t_{-i},R_{-i})}\right) \to \tau_{-i}^{n-1}(t_{-i})$ as $m \to \infty$ and $R_{-i} \supseteq \mathrm{EFR}_{-i}\left(y_{-i,m}^{(t_{-i},R_{-i})}\right)$ for all m. (For $n = 1$, define $y_{-i,m}^{(t_{-i},R_{-i})} = y_{-i}^{(\theta_{-i},R_{-i})}$ for all m.) We now define, for each m, a type $t_{i,m} \in T_i^*$ such that $\vartheta_i^*(t_{i,m}) = \vartheta_i(t_i)$, and

$$\begin{aligned} \kappa_i^*(t_{i,m})[\theta_0, y_{-i}] = {} & \frac{m}{m+1}\nu_{i,m}\left[\left\{(\theta_0, t_{-i}, R_{-i}) : y_{-i,m}^{(t_{-i},R_{-i})} = y_{-i}\right\}\right] \\ & + \frac{1}{m+1}\tilde{\nu}_{i,m}\left[\left\{(\theta_0, \theta_{-i}, R_{-i}) : y_{-i}^{(\theta_{-i},R_{-i})} = y_{-i}\right\}\right]. \end{aligned} \tag{3.3}$$

for every $(\theta_0, y_{-i}) \in \Theta_0 \times T_{-i}^*$. Notice that each $\kappa_i^*(t_{i,m})$ has finite support. Since $\tau_{-i}^{n-1}\left(y_{-i,m}^{(t_{-i},R_{-i})}\right) \to \tau_{-i}^{n-1}(t_{-i})$ as $m \to \infty$ and $\mathrm{marg}_{\Theta_0 \times T_{-i}} \nu_{i,m} = \kappa_i(t_i)$ for each m, we have $\tau_i^n(t_{i,m}) \to \tau_i^n(t_i)$ as $m \to \infty$.

We need to show that $R_i \supseteq \mathrm{EFR}_i(t_{i,m})$ for each m. By Definition 3.3 of EFR, for each $s_i \in \mathrm{EFR}_i(t_{i,m})$, there exist a distribution $\mu_i \in \Delta(\Theta_0 \times T^*_{-i} \times S_{-i})$ and a conjecture $\hat{\pi}_i \in \mathrm{C}^n_i(\vartheta_i(t_i))$ such that

(i) $\mathrm{marg}_{\Theta_0 \times T^*_{-i}} \mu_i = \kappa^*_i(t_{i,m})$;

(ii) $\mu_i[\{(\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \mathrm{EFR}_{-i}(t_{-i})\}] = 1$;

(iii) $\hat{\pi}_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta^*_{-i}(t_{-i}) = \theta_{-i}\}]$;

(iv) $s_i \in r_i(\hat{\pi}_i \mid \vartheta^*_i(t_{i,m}))$.

For $\nu_{i,m} \in (\Theta_0 \times T_{-i} \times \mathcal{S}_{-i})$, define a function $f_i : \Theta_0 \times T_{-i} \times \mathcal{S}_{-i} \to \Delta(S_{-i})$ as follows: If (t_{-i}, R_{-i}) are such that $y^{(t_{-i},R_{-i})}_{-i,m} = y_{-i}$, then let $f_i(\theta_0, t_{-i}, R_{-i})[s_{-i}] = \mu_i[s_{-i} \mid \theta_0, y_{-i}]$. Because $R_{-i} \supseteq \mathrm{EFR}_{-i}\Big(y^{(t_{-i},R_{-i})}_{-i,m}\Big)$, we have $f_i(\theta_0, t_{-i}, R_{-i})[R_{-i}] = 1$. Then we define

$$\begin{aligned}\lambda_i[\theta_0, \theta_{-i}, s_{-i}] &= \sum_{t_{-i}:\vartheta_{-i}(t_{-i})=\theta_{-i}} \sum_{R_{-i} \in \mathcal{C}_{-i}(t_{-i})} \nu_{i,m}[\theta_0, t_{-i}, R_{-i}] f_i(\theta_0, t_{-i}, R_{-i})[s_{-i}] \\ &= \sum_{y_{-i}:\vartheta^*_{-i}(y_{-i})=\theta_{-i}} \nu_{i,m}\left[\left\{(\theta_0, t_{-i}, R_{-i}) : y^{(t_{-i},R_{-i})}_{-i,m} = y_{-i}\right\}\right] \mu_i[s_{-i} \mid \theta_0, y_{-i}].\end{aligned} \tag{3.4}$$

Similarly, for $\tilde{\nu}_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i})$, define a function $\tilde{f}_i : \Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i} \to \Delta(S_{-i})$ as follows: If (θ_{-i}, R_{-i}) is such that $y^{(\theta_{-i},R_{-i})}_{-i} = y_{-i}$, then let

$$\tilde{f}_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}] = \mu_i[s_{-i} \mid \theta_0, y_{-i}].$$

Because $R_{-i} \supseteq \mathrm{EFR}_{-i}\Big(y^{(\theta_{-i},R_{-i})}_{-i}\Big)$, we have $f_i(\theta_0, t_{-i}, R_{-i})[R_{-i}] = 1$. Then we define

$$\begin{aligned}\tilde{\lambda}_i[\theta_0, \theta_{-i}, s_{-i}] &= \sum_{R_{-i} \in \mathcal{R}_{-i}(\theta_{-i})} \tilde{\nu}_i[\theta_0, \theta_{-i}, R_{-i}] f_i(\theta_0, \theta_{-i}, R_{-i})[s_{-i}] \\ &= \sum_{y_{-i}:\vartheta^*_{-i}(y_{-i})=\theta_{-i}} \tilde{\nu}_{i,m}\left[\left\{(\theta_0, \theta_{-i}, R_{-i}) : y^{(\theta_{-i},R_{-i})}_{-i} = y_{-i}\right\}\right] \mu_i[s_{-i} \mid \theta_0, y_{-i}].\end{aligned} \tag{3.5}$$

It is left to show that $\hat{\pi}_i \in \Lambda^n_i(\nu_i, \tilde{\nu}_i, \frac{1}{m+1} \mid t_i)$. By construction, $\lambda_i, \tilde{\lambda}_i \in \Delta(\Theta_0 \times$

$\bar{\Theta}_{-i}(\vartheta_i(t_i)) \times S_{-i})$ are consistent with ν_i and $\tilde{\nu}_i$, respectively. Moreover, we have

$$\begin{aligned}
&\hat{\pi}_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] \\
=& \sum_{y_{-i}:\vartheta^*_{-i}(y_{-i})=\theta_{-i}} \mu_i[\theta_0, y_{-i}, s_{-i}] \\
=& \sum_{y_{-i}:\vartheta^*_{-i}(y_{-i})=\theta_{-i}} \kappa^*_i(t_{i,m})[\theta_0, y_{-i}]\mu_i[s_{-i} \mid \theta_0, y_{-i}] \\
=& \frac{m}{m+1} \sum_{y_{-i}:\vartheta^*_{-i}(y_{-i})=\theta_{-i}} \nu_{i,m}\left[\left\{(\theta_0, t_{-i}, R_{-i}) : y^{(t_{-i},R_{-i})}_{-i,m} = y_{-i}\right\}\right] \mu_i[s_{-i} \mid \theta_0, y_{-i}] \\
&+ \frac{1}{m+1} \sum_{y_{-i}:\vartheta^*_{-i}(y_{-i})=\theta_{-i}} \tilde{\nu}_{i,m}\left[\left\{(\theta_0, \theta_{-i}, R_{-i}) : y^{(\theta_{-i},R_{-i})}_{-i} = y_{-i}\right\}\right] \mu_i[s_{-i} \mid \theta_0, y_{-i}] \\
=& \frac{m}{m+1}\lambda_i[\theta_0, \theta_{-i}, s_{-i}] + \frac{1}{m}\tilde{\lambda}_i[\theta_0, \theta_{-i}, s_{-i}].
\end{aligned}$$

The second equality is by the fact that $\mathrm{marg}_{\Theta_0 \times T^*_{-i}} \mu_i = \kappa^*_i(t_{i,m})$; the third equality is by equation (3.3); and the last one is due to equations (3.4) and (3.5).

Therefore, we have $\hat{\pi}_i \in \Lambda^n_i(\nu_i, \tilde{\nu}_i, \frac{1}{m+1} \mid t_i)$. But then $R_i \supseteq r_i(\hat{\pi}_i \mid \vartheta_i(t_i))$, which means $s_i \in R_i$. Since $s_i \in \mathrm{EFR}_i(t_{i,m})$ is arbitrary, we have $R_i \supseteq \mathrm{EFR}_i(t_{i,m})$, completing the proof of this lemma. □

Lemma 3.5. $\mathcal{R}^{\mathrm{loc}}_i(t_i) \subseteq \bigcap_{n\geq 0} \mathcal{R}^{\mathrm{loc},n}_i(t_i)$.

Proof. We prove this by showing that the profile $\left\{\left(\mathcal{R}^{\mathrm{loc}}_i(t_i)\right)_{t_i \in T_i}\right\}_{i\in I}$ survives each round of definition (3.2) in the main text. Since every $R_i \in \mathcal{R}^{\mathrm{loc}}_i(t_i)$ satisfies $R_i \in \mathcal{R}^{\uparrow}_i(\vartheta_i(t_i))$ by definition, we have $\mathcal{R}^{\mathrm{loc}}_i(t_i) \subseteq \mathcal{R}^{\mathrm{loc},0}_i(t_i)$. Now suppose $\mathcal{R}^{\mathrm{loc}}_i(t_i) \subseteq \mathcal{R}^{\mathrm{loc},n-1}_i(t_i)$ for all $i \in I$ and $t_i \in T_i$. We want to show that, if $R_i \in \mathcal{R}^{\mathrm{loc}}_i(t_i)$, then for any $\varepsilon \in (0, 1]$, there exists a pair $(\nu_i, \tilde{\nu}_i) \in \Delta\Big(\Theta_0 \times T_{-i} \times \mathcal{S}_{-i}\Big) \times \Delta\Big(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i}\Big)$ such that the following are true:

(i) $\mathrm{marg}_{\Theta_0 \times T_{-i}} \nu_i = \kappa_i(t_i)$;

(ii) $\nu_i\Big[(\theta_0, t_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}^{\mathrm{loc},n-1}_i(t_i)\Big] = 1$;

(iii) $\tilde{\nu}_i\Big[(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}^{\uparrow}_{-i}(\theta_{-i})\Big] = 1$;

(iv) $R_i \supseteq \bigcup_{\pi_i \in \Lambda^n_i(\nu_i, \tilde{\nu}_i, \varepsilon \mid t_i)} r_i(\pi_i \mid \vartheta_i(t_i))$.

Since $R_i \in \mathcal{R}^{\mathrm{loc}}_i(t_i)$, by definition, there exists a sequence $\{t_{i,m}\}_{m\in\mathbb{N}} \subset T^*_i$ such that $t_{i,m} \to t_i$, and $R_i \supseteq \mathrm{EFR}_i(t_{i,m})$ for all m. Because Θ_i is finite, we can assume $\vartheta^*_i(t_{i,m}) =$

$\vartheta_i(t_i)$ for all m without loss of generality. For each m, define $\nu_{i,m} \in \Delta(\Theta_0 \times T^*_{-i} \times \mathcal{S}_{-i})$ by

$$\nu_{i,m}[\{\theta_0\} \times E_{-i} \times \{R_{-i}\}] = \kappa^*_i(t_{i,m})\left[(\theta_0, y_{-i}) : y_{-i} \in E_{-i} \text{ and } \mathrm{EFR}_{-i}(y_{-i}) = R_{-i}\right], \quad (3.6)$$

for every measurable $E_{-i} \subseteq T^*_{-i}$ and every $(\theta_0, R_{-i}) \in \Theta_0 \times \mathcal{S}_{-i}$. Because the space of probability measures $\Delta(\Theta_0 \times T^*_{-i} \times \mathcal{S}_{-i})$ is weak* compact metrizable, $\{\nu_{i,m}\}$ has a convergent subsequence $\{\nu_{i,m_k}\}$, and let ν_i denote its limit. We first verify that ν_i satisfies conditions (i) and (ii) above. For (i), since $\mathrm{marg}_{\Theta_0 \times T^*_{-i}} \nu_{i,m_k} = \kappa^*_i(t_{i,m_k})$ by definition, $t_{i,m_k} \to t_i$ and $\nu_{i,m_k} \to \nu_i$ as $k \to \infty$, and κ^*_i is continuous, we have $\mathrm{marg}_{\Theta_0 \times T_{-i}} \nu_i = \kappa_i(t_i)$. For (ii), let $\ell \in \mathbb{N}$ and define

$$F_\ell = \mathrm{cl}\left\{(\theta_0, y_{-i}, R_{-i}) : \exists y'_{-i} \in T^*_{-i} \text{ s.t. } d_{-i}(y_{-i}, y'_{-i}) \le \frac{1}{\ell} \text{ and } \mathrm{EFR}_{-i}(y'_{-i}) = R_{-i}\right\},$$
$$F_\infty = (\Theta_0 \times T_{-i} \times \mathcal{S}_{-i}) \cap \bigcap_{\ell \ge 1} F_\ell,$$

where d_{-i} is the metric on T^*_{-i}. Observe that

$$F_\ell \supseteq \left\{(\theta_0, y_{-i}, R_{-i}) : y_{-i} \in T^*_{-i} \text{ and } \mathrm{EFR}_{-i}(y_{-i}) = R_{-i}\right\}, \ \forall \ell \ge 0$$
$$F_\infty \subseteq \left\{(\theta_0, y_{-i}, R_{-i}) : y_{-i} \in T_{-i} \text{ and } R_{-i} \in \mathcal{R}^{\mathrm{loc}}_{-i}(t_{-i})\right\}.$$

By definition, $\nu_{i,m}[\{(\theta_0, y_{-i}, R_{-i}) : \mathrm{EFR}_{-i}(y_{-i}) = R_{-i}\}] = 1$, so $\nu_{i,m}[F_\ell] = 1$ for all ℓ. Since F_ℓ is closed and $\nu_{i,m_k} \to \nu_i$ as $k \to \infty$, we have $\nu_i[F_\ell] \ge \limsup \nu_{i,m_k}[F_\ell] = 1$. Because $\mathrm{marg}_{\Theta_0 \times T_{-i}} \nu_i = \kappa_i(t_i)$, we also have $\nu_i[\Theta_0 \times T_{-i} \times \mathcal{S}_{-i}] = 1$. Therefore, we conclude that $\nu_i[F_\infty] = 1$, which implies $\nu_i\left[\left\{(\theta_0, t_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}^{\mathrm{loc}}_{-i}(t_{-i})\right\}\right] = 1$. Condition (ii) above is then implied by the induction hypothesis.

We next construct $\tilde{\nu}_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i})$ as follows:

$$\tilde{\nu}_i[\theta_0, \theta_{-i}, R_{-i}] = \frac{1}{\varepsilon}\Bigg(\nu_{i,m}[\{\theta_0\} \times (\vartheta^*_{-i})^{-1}(\theta_{-i}) \times \{R_{-i}\}]$$
$$- (1-\varepsilon) \sum_{t_{-i} : \vartheta_{-i}(t_{-i}) = \theta_{-i}} \nu_i[\theta_0, t_{-i}, R_{-i}]\Bigg). \quad (3.7)$$

Since $\nu_{i,m_k} \to \nu_i$, we can choose a sufficiently large m so that $\tilde{\nu}_i[\theta_0, \theta_{-i}, R_i] \ge 0$ for every $(\theta_0, \theta_{-i}, R_i)$. When $\tilde{\nu}_i[\theta_0, \theta_{-i}, R_i] > 0$, we must have $\nu_{i,m}\left[\{\theta_0\} \times (\vartheta^*_{-i})^{-1}(\theta_{-i}) \times \{R_{-i}\}\right] >$

0. By definition of $\nu_{i,m}$, there exists a type $y_{-i}^{(\theta_{-i},R_{-i})}$ such that $\vartheta_{-i}^*(y_{-i}) = \theta_{-i}$ and $R_{-i} = \mathrm{EFR}_{-i}\left(y_{-i}^{(\theta_{-i},R_{-i})}\right)$, so $R_{-i} \in \mathcal{R}_{-i}^{\uparrow}(\theta_{-i})$. Hence, condition (iii) holds.

Finally, we need to show that for any conjecture $\pi_i \in \Lambda_i^n(\nu_i, \tilde{\nu}_i, \varepsilon \mid t_i)$, we have $R_i \supseteq r_i(\pi_i \mid \vartheta_i(t_i))$. Let $\lambda_i, \tilde{\lambda}_i \in \Delta(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times S_{-i})$ be the distributions consistent with ν_i and $\tilde{\nu}_i$, respectively, such that

$$\pi_i(\phi) = (1-\varepsilon)\lambda_i + \varepsilon\tilde{\lambda}_i.$$

We now define a distribution $\mu_i \in \Delta\left(\Theta_0 \times T_{-i}^* \times S_{-i}\right)$ as follows: For every $\theta_{-i} \in \bar{\Theta}_{-i}(\vartheta_i(t_i))$ and measurable $E_{-i} \subseteq (\vartheta_{-i}^*)^{-1}(\theta_{-i})$, let

$$\mu_i[\{\theta_0\} \times E_{-i} \times \{s_{-i}\}] = K_i(\theta_0, E_{-i}) \cdot \left((1-\varepsilon)\lambda_i + \varepsilon\tilde{\lambda}_i\right)[\theta_0, \theta_{-i}, s_{-i}]. \tag{3.8}$$

where $K_i(\theta_0, E_{-i})$ is a multiplier defined by

$$K_i(\theta_0, E_{-i}) = \frac{\left(\mathrm{marg}_{\Theta_0 \times T_{-i}^*}\nu_{i,m}\right)[\{\theta_0\} \times E_{-i}]}{\left(\mathrm{marg}_{\Theta_0 \times T_{-i}^*}\nu_{i,m}\right)[\{\theta_0\} \times (\vartheta_{-i}^*)^{-1}(\theta_{-i})]}. \tag{3.9}$$

It can be checked that the probability distribution μ_i is well-defined. Moreover, for every measurable $E_{-i} \subseteq (\vartheta_{-i}^*)^{-1}(\theta_{-i})$ and $\theta_0 \in \Theta_0$, we have

$$\begin{aligned}
&\left(\mathrm{marg}_{\Theta_0 \times T_{-i}^*}\mu_i\right)[\{\theta_0\} \times E_{-i}] \\
&= K_i(\theta_0, E_{-i}) \cdot \left(\mathrm{marg}_{\Theta_0 \times T_{-i}^*}\nu_{i,m}\right)\left[\{\theta_0\} \times (\vartheta_{-i}^*)^{-1}(\theta_{-i})\right] \\
&= \left(\mathrm{marg}_{\Theta_0 \times T_{-i}^*}\nu_{i,m}\right)[\{\theta_0\} \times E_{-i}] \\
&= \kappa_i^*(t_{i,m})\left[\{\theta_0\} \times E_{-i}\right],
\end{aligned}$$

where the first equality is by equations (3.8) and (3.7); the second equality is by equation (3.9); and the third equality is by equation (3.6). The above implies $\mathrm{marg}_{\Theta_0 \times T_{-i}^*}\mu_i = \kappa_i^*(t_{i,m})$. Also, by the definitions of ν_i and $\tilde{\nu}_i$, and the fact that λ_i and $\tilde{\lambda}_i$ are consistent with ν_i and $\tilde{\nu}_i$ respectively, we have $\mu_i[\{(\theta_0, y_{-i}, s_{-i}) : s_{-i} \in \mathrm{EFR}_{-i}(y_{-i})\}] = 1$.

Our last step is to show that $\pi_i \in \Psi_i^n(t_{i,m})$, but this is because $\pi_i \in \mathrm{C}_i^n(\vartheta_i(t_{i,m}))$ and

$$\begin{aligned}
\mu_i\left[\{(\theta_0, y_{-i}, s_{-i}) : \vartheta_{-i}^*(y_{-i}) = \theta_{-i}\}\right] &= (1-\varepsilon)\lambda_i[\theta_0, \theta_{-i}, s_{-i}] + \varepsilon\tilde{\lambda}_i[\theta_0, \theta_{-i}, s_{-i}] \\
&= \pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}],
\end{aligned}$$

where the first equality is by combining equations (3.8) and (3.9). Therefore, by Definition 3.3 of EFR, we have $r_i(\pi_i \mid \vartheta_i(t_i)) \subseteq \text{EFR}_i(t_{i,m})$ (note that $\vartheta_i^*(t_{i,m}) = \vartheta_i(t_i)$). Since $R_i \supseteq \text{EFR}_i(t_{i,m})$, we can conclude $R_i \supseteq r_i(\pi_i \mid \vartheta_i(t_i))$, which wraps up the proof. □

3.5.4 Proof of Proposition 3.4

Since each $\text{EFR}_i(\cdot)$ is an upper hemicontinuous correspondence on T_i^*, the proof of Proposition 1 from Chen et al. (2021) can be directly used to show the following lemma.

Lemma 3.6. *Given a finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$. A prediction P is a robust refinement if and only if for every $i \in I$ and $t_i \in T_i$, there exists an open neighborhood $E_{t_i} \subseteq T_i^*$ of t_i such that $P_i(t_i) \cap \text{EFR}_i(t_i^*) \neq \varnothing$ for every $t_i^* \in E_{t_i}$.*

We now use this lemma to prove Proposition 3.4. For "⇒", suppose P is robust. Fix any $i \in I$ and $t_i \in T_i$. By definition, for any $R_i \in \mathcal{R}_i^{\text{loc}}(t_i)$, there exists a sequence $\{t_{i,m}\}_{m\in\mathbb{N}} \subseteq T_i^*$ such that $t_{i,m} \to t_i$ and $R_i \supseteq \text{EFR}_i(t_{i,m})$ for all m. Since P is a robust refinement, by Lemma 3.6, $P_i(t_i) \cap \text{EFR}_i(t_{i,m}) \neq \varnothing$ for sufficiently large m, which means $P_i(t_i) \cap R_i \neq \varnothing$.

For the converse "⇐", suppose P is not a robust refinement. Then by Lemma 3.6, there exists a type $t_i \in T_i$ and a sequence $\{t_{i,m_k}\}_{k\in\mathbb{N}} \subseteq T_i^*$ such that $t_{i,m_k} \to t_i$ and $P_i(t_i) \cap \text{EFR}_i(t_{i,m_k}) = \varnothing$ for all k. This means $\bigcup_{k\geq 0}(P_i(t_i) \cap \text{EFR}_i(t_{i,m_k})) = P_i(t_i) \cap \left(\bigcup_{k\geq 0} \text{EFR}_i(t_{i,m_k})\right) = \varnothing$. But $\bigcup_{k\geq 0} \text{EFR}_i(t_{i,m_k}) \in \mathcal{R}_i^{\text{loc}}(t_i)$ by definition.

3.5.5 Proof of Proposition 3.5

3.5.5.1 (1) ⇒ (2)

We split this direction into two steps.

First, we construct a finite type space such that the collection of EFR sets contains all maximal sets in $\mathcal{R}_i^{\downarrow}(\theta_i)$. To achieve this, for each $i \in I$, partition the space T_i^* into $\mathbf{T}_i^{\text{EFR}} = \left\{\tilde{T}_i^{(\theta_i,R_i)}\right\}$ by payoff types and the EFR correspondence, i.e. $t_i, t_i' \in \tilde{T}_i^{(\theta_i,R_i)}$ if and only if $\vartheta_i^*(t_i) = \vartheta_i^*(t_i') = \theta_i$ and $\text{EFR}_i(t_i) = \text{EFR}_i(t_i') = R_i$. Note that the partition $\mathbf{T}_i^{\text{EFR}}$ is finite and measurable (in T_i^*) by continuity of ϑ_i^*, upper hemicontinuity of EFR, and finiteness of Θ_i and S_i. Now for each $\tilde{T}_i^{(\theta_i,R_i)} \in \mathbf{T}_i^{\text{EFR}}$, fix a type $\tilde{t}_i^{(\theta_i,R_i)} \in \tilde{T}_i^{(\theta_i,R_i)}$, and define a one-to-one and onto mapping $g_i : \tilde{t}_i^{(\theta_i,R_i)} \mapsto \tilde{T}_i^{(\theta_i,R_i)}$. Define a finite type

space $\left\{\left(\mathbf{T}_i^{\mathrm{EFR}}, \tilde{\vartheta}_i, \tilde{\kappa}_i\right)_{i\in I}\right\}$ such that $\tilde{\vartheta}_i\left(\tilde{T}_i^{(\theta_i,R_i)}\right) = \theta_i$, and

$$\tilde{\kappa}_i\left(\tilde{T}_i^{(\theta_i,R_i)}\right)\left[\theta_0, \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}\right] = \kappa_i^*\left(\tilde{t}_i^{(\theta_i,R_i)}\right)\left[(\theta_0, t_{-i}) : t_{-i} \in \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}\right]. \tag{3.10}$$

Note that $\tilde{T}_i^{(\theta_i,R_i)}$ denotes both a type in $\mathbf{T}_i^{\mathrm{EFR}}$ (on the left-hand-side) and a subset of T_i^* (on the right-hand-side). We claim that $\mathrm{EFR}_i\left(\tilde{T}_i^{(\theta_i,R_i)}\right) \supseteq \mathrm{EFR}_i\left(\tilde{t}_i^{(\theta_i,R_i)}\right)$ for every $\tilde{T}_i^{(\theta_i,R_i)}$. To prove this, we will show that $\mathrm{EFR}_i^n\left(\tilde{T}_i^{(\theta_i,R_i)}\right) \supseteq \mathrm{EFR}_i\left(\tilde{t}_i^{(\theta_i,R_i)}\right)$ for all $n \geq 0$. This is clearly true for $n = 0$. Suppose it holds for $n-1$. Take $s_i \in \mathrm{EFR}_i\left(\tilde{t}_i^{(\theta_i,R_i)}\right)$. Then there exists a distribution $\mu_i \in \Delta\left(\Theta_0 \times T_{-i}^* \times S_{-i}\right)$ and $\pi_i \in \mathrm{C}_i^n(\vartheta_i^*(t_i))$ such that[22]

(i) $\mathrm{marg}_{\Theta_0\times T_{-i}^*}\mu_i = \kappa_i^*\left(\tilde{t}_i^{(\theta_i,R_i)}\right)$;

(ii) $\mu_i[\{(\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \mathrm{EFR}_{-i}(t_{-i})\}] = 1$;

(iii) $\pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}^*(t_{-i}) = \theta_{-i}\}]$;

(iv) $s_i \in r_i(\pi_i \mid \vartheta_i^*(t_i))$.

Now define a new distribution $\tilde{\mu}_i \in \Delta(\Theta_0 \times \mathbf{T}_{-i}^{\mathrm{EFR}} \times S_{-i})$ for type $\tilde{T}_i^{(\theta_i,R_i)} \in \mathbf{T}_i^{\mathrm{EFR}}$ by

$$\tilde{\mu}_i\left[\theta_0, \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}, s_{-i}\right] = \mu_i\left[\left\{(\theta_0, t_{-i}, s_{-i}) : t_{-i} \in \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}\right\}\right].$$

By equation (3.10) and condition (i) above, we have $\mathrm{marg}_{\Theta_0\times\mathbf{T}_{-i}^{\mathrm{EFR}}}\tilde{\mu}_i = \tilde{\kappa}_i\left(\tilde{T}_i^{(\theta_i,R_i)}\right)$. Since for every $t_{-i} \in \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}$, we have $\mathrm{EFR}_{-i}(t_{-i}) = \mathrm{EFR}_{-i}\left(\tilde{t}_{-i}^{(\theta_{-i},R_{-i})}\right) = R_{-i}$, condition (ii) above and the induction hypothesis implies that

$$\tilde{\mu}_i\left[\left(\theta_0, \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}, s_{-i}\right) : s_{-i} \in \mathrm{EFR}_{-i}^{n-1}\left(\tilde{T}_{-i}^{(\theta_{-i},R_{-i})}\right)\right] = 1.$$

Moreover, by construction,

$$\begin{aligned}\tilde{\mu}_i\left[\left\{\left(\theta_0, \tilde{T}_{-i}^{(\theta_{-i},R_{-i})}, s_{-i}\right) : \tilde{\vartheta}_{-i}\left(\tilde{T}_{-i}^{(\theta_{-i},R_{-i})}\right) = \theta_{-i}\right\}\right] &= \mu_i\left[\left\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}^*(t_{-i}) = \theta_{-i}\right\}\right] \\ &= \pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}].\end{aligned}$$

Therefore, the fact that $s_i \in r_i(\pi_i \mid \vartheta_i^*(t_i))$ implies $s_i \in \mathrm{EFR}_i^n\left(\tilde{T}_i^{(\theta_i,R_i)}\right)$, and thus we have proved our claim.

[22]This is because the procedure EFR converges in finitely many steps and the fact that $\{\mathrm{C}_i^n(\theta_i)\}_{n\geq 0}$ is a decreasing sequence.

To summarize the first step, for any $R_i \in \mathcal{R}_i^{\downarrow}(\theta_i)$, there is a *finite* type t_i such that $\vartheta_i^*(t_i) = \theta_i$ and $R_i \subseteq \text{EFR}(t_i)$. The second step is simple. Since any strategy $s_i \in \text{EFR}(t_i)$ can be uniquely selected for t_i, we have $\text{EFR}(t_i) \subseteq R_i^{\text{u}}(\theta_i)$. Hence, $R_i \subseteq R_i^{\text{u}}(\theta_i)$.

3.5.5.2 (2) ⇒ (1)

Fixing a finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$, we need to show that for each $t_i \in T_i$ and any strategy $s_i \in \text{EFR}_i(t_i)$, the singleton set $\{s_i\} \in \mathcal{R}_i^{\text{loc}}(t_i)$. By Proposition (3.2), we can prove this by showing $\{s_i\} \in \mathcal{R}_i^{\text{loc},n}(t_i)$ for all $n \geq 0$. For $n = 0$, note that $\text{EFR}_i(t_i) \in \mathcal{R}_i^{\downarrow}(\vartheta_i(t_i))$, and hence $\text{EFR}_i(t_i) \subseteq R_i^{\text{u}}(\vartheta_i(t_i))$ by assumption. This means each $s_i \in \text{EFR}_i(t_i)$ forms a singleton set in $\mathcal{R}_i^{\uparrow}(\vartheta_i(t_i))$. But $\mathcal{R}_i^{\text{loc},0}(t_i) = \mathcal{R}_i^{\uparrow}(\vartheta_i(t_i))$ by definition, so $\{s_i\} \in \mathcal{R}_i^{\text{loc},0}(t_i)$ is established.

Suppose for each $t_i \in T_i$, $s_i \in \text{EFR}_i(t_i)$ implies $\{s_i\} \in \mathcal{R}_i^{\text{loc},n-1}(t_i)$. We now show that the same statement holds for n. Fix a type $t_i \in T_i$. If $s_i \in \text{EFR}_i(t_i)$, there exists a distribution $\mu_i \in \Delta\Big(\Theta_0 \times T_{-i} \times S_{-i}\Big)$ and $\pi_i \in \text{C}_i^n(\vartheta_i(t_i))$ such that

(i) $\text{marg}_{\Theta_0\times T_{-i}}\mu_i = \kappa_i(t_i)$;

(ii) $\mu_i[\{(\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \text{EFR}_{-i}(t_{-i})\}] = 1$;

(iii) $\pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}(t_{-i}) = \theta_{-i}\}]$;

(iv) $s_i \in r_i(\pi_i \mid \vartheta_i(t_i))$.

Now define $\nu_i \in \Delta\left(\Theta_0 \times T_{-i} \times \mathcal{S}_{-i}\right)$ by

$$\nu_i[\theta_0, t_{-i}, \{s_{-i}\}] = \mu_i[\theta_0, t_{-i}, s_{-i}].$$

By construction, we have $\text{marg}_{\Theta_0\times T_{-i}}\nu_i = \kappa_i(t_i)$, and the *only* λ_i consistent with ν_i is the one such that $\lambda_i = \pi_i(\phi)$. Moreover, by the induction hypothesis, we also have

$$\nu_i\left[\Big\{(\theta_0, t_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\text{loc},n-1}(t_{-i})\Big\}\right] = 1.$$

Since $\{s_i\} \in R_i^{\text{u}}(\vartheta_i(t_i))$ by assumption, there exists a type $t_i^{s_i} \in T_i^*$ such that $\vartheta_i^*\Big(t_i^{s_i}\Big) = \vartheta_i(t_i)$ and $\text{EFR}_i\Big(t_i^{s_i}\Big) = \{s_i\}$. Therefore, for *any* distribution $\tilde{\mu}_i \in \Delta(\Theta_0 \times T_{-i}^* \times S_{-i})$ and $\tilde{\pi}_i \in \text{C}_i^n(\vartheta_i(t_i))$ such that

(i) $\text{marg}_{\Theta_0\times T_{-i}^*}\tilde{\mu}_i = \kappa_i^*\Big(t_i^{s_i}\Big)$;

(ii) $\tilde{\mu}_i[\{(\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \text{EFR}_{-i}(t_{-i})\}] = 1$;

(iii) $\tilde{\pi}_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \tilde{\mu}_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}(t_{-i}) = \theta_{-i}\}]$,

we have $\{s_i\} = r_i(\tilde{\pi}_i \mid \vartheta_i(t_i))$ by the definition of EFR. We now define $\tilde{\nu}_{i,\eta} \in \Delta\Big(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i}\Big)$ by

$$\tilde{\nu}_i[\theta_0, \theta_{-i}, R_{-i}] = \kappa_i^*\Big(t_i^{s_i}\Big)[\{(\theta_0, t_{-i}) : \vartheta_{-i}^*(t_{-i}) = \theta_{-i} \text{ and } \mathrm{EFR}_{-i}(t_{-i}) = R_{-i}\}].$$

By construction, for any $\tilde{\lambda}_i$ consistent with $\tilde{\nu}_i$ and $\tilde{\pi}_i \in \mathrm{C}_i^n(\vartheta_i(t_i))$ such that $\tilde{\lambda}_i = \tilde{\pi}_i(\phi)$, we have $\{s_i\} = r_i(\tilde{\pi}_i \mid \vartheta_i(t_i))$. Moreover,

$$\tilde{\nu}_i\left[\left\{(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow}(\theta_{-i})\right\}\right] = 1.$$

Finally, take any $\varepsilon \in (0, 1]$ and the distributions ν_i and $\tilde{\nu}_i$ above. For *any* $\hat{\pi}_i \in \mathrm{C}_i^n(\vartheta_i(t_i))$ such that

$$\hat{\pi}_i(\phi) = (1 - \varepsilon)\lambda_i + \varepsilon\tilde{\lambda}_i,$$

where λ_i and $\tilde{\lambda}_i$ are consistent with ν_i and $\tilde{\nu}_i$ respectively, we must have $\{s_i\} = r_i(\hat{\pi}_i \mid \vartheta_i(t_i))$. To see this, note that for any $h \in \mathcal{H}_i(s_i)$ such that $[h] \cap \mathrm{supp}\lambda_i = \varnothing$ or $[h] \cap \mathrm{supp}\tilde{\lambda}_i = \varnothing$, $\hat{\pi}(h) = \tilde{\pi}(h)$ for some $\tilde{\lambda}_i$ consistent with $\tilde{\nu}_i$ and $\tilde{\pi}_i \in \mathrm{C}_i^n(\vartheta_i(t_i))$ such that $\tilde{\lambda}_i = \tilde{\pi}_i(\phi)$. Thus, s_i is the unique best response at history h. On the other hand, for any $h \in \mathcal{H}_i(s_i)$ such that $[h] \cap \mathrm{supp}\lambda_i \cap \mathrm{supp}\tilde{\lambda}_i \neq \varnothing$, $\hat{\pi}(h)$ is a convex combination of distributions such that (with probability $1-\varepsilon$) s_i is a best response and (with probability ε) s_i is the only best response. Hence, for all $h \in \mathcal{H}_i(s_i)$,

$$\{s_i\} = \operatorname*{arg\,max}_{s_i' \in S_i(h)} \sum_{\theta_0, \theta_{-i}, s_{-i}} u_i(z(s_i', s_{-i}), \theta_0, \vartheta_i(t_i), \theta_{-i})\hat{\pi}_i(h)[\theta_0, \theta_{-i}, s_{-i}].$$

Therefore, $\{s_i\} = \bigcup_{\pi_i \in \Lambda_i^n(\nu_i, \tilde{\nu}_i, \varepsilon | t_i)} r_i(\pi_i \mid \vartheta_i(t_i))$. This means $\{s_i\} \in \mathcal{R}_i^{\mathrm{loc},n}(t_i)$, completing the proof.

3.5.6 Proof of Proposition 3.6

3.5.6.1 (1) ⇒ (2)

Fix a player $i \in I$ and $\theta_i \in \Theta_i$. If $R_i \in \mathcal{R}_i^{\uparrow}(\theta_i)$, there exists a type $t_i \in T_i^*$ such that $\vartheta_i^*(t_i) = \theta_i$ and $R_i \supseteq \mathrm{EFR}_i(t_i)$. By denseness of $\mathcal{U}_i$, there is a sequence $\{t_{i,m}\}_{m \in \mathbb{N}} \subseteq T_i^*$ such that $t_{i,m} \to t_i$ as $m \to \infty$, and $|\mathrm{EFR}_i(t_{i,m})| = 1$. Since Θ_i and S_i are finite, there exists a subsequence $\{t_{i,m_k}\}$ such that $\vartheta_i^*(t_{i,m_k}) = \theta_i$ and $\{s_i\} = \mathrm{EFR}_i(t_{i,m_k})$ for some

$s_i \in S_i$. This means $s_i \in R_i^{\mathrm{u}}(\theta_i)$. By upper hemicontinuity of EFR, we must have $s_i \in \mathrm{EFR}_i(t_i)$. Therefore, $s_i \in \mathrm{EFR}_i(t_i) \cap R_i^{\mathrm{u}}(\theta_i) \neq \varnothing$, which implies $R_i \cap R_i^{\mathrm{u}}(\theta_i) \neq \varnothing$.

3.5.6.2 (2) ⇒ (1)

Openness of $\mathcal{U}_i$ is implied by the fact that $\mathrm{EFR}_i(\cdot)$ is upper hemicontinuous and nonempty on T_i^*. Denseness of $\mathcal{U}_i$ is proved in two steps.

First, we show that for every $t_i \in T_i$ in any finite type space $\{(T_i, \vartheta_i, \kappa_i)_{i\in I}\}$, the set $\mathrm{EFR}_i(t_i)$ contains some strategies that can be uniquely selected. To achieve this, we define a refinement of EFR which collects all strategies that can be uniquely selected for a type. For each $t_i \in T_i$, let $\mathrm{EFR}_i^{\mathrm{u},0}(t_i) = R_i^{\mathrm{u}}(\vartheta_i(t_i))$, which is nonempty by assumption. For $n \geq 1$, let $\mathrm{EFR}_{-i}^{\mathrm{u},n-1}(t_{-i}) = \times_{j\neq i} \mathrm{EFR}_j^{\mathrm{u},n-1}(t_j)$, and define

$$\Psi_i^{\mathrm{u},n}(t_i) = \left\{ \begin{array}{l} \pi_i \in \mathrm{C}_i^n(\vartheta_i(t_i)) : \\ \quad \exists \mu_i \in \Delta(\Theta_0 \times T_{-i} \times S_{-i}) \text{ s.t.} \\ \quad \text{(i) } \mathrm{marg}_{\Theta_0 \times T_{-i}} \mu_i = \kappa_i(t_i); \\ \quad \text{(ii) } \mu_i \left[\left\{ (\theta_0, t_{-i}, s_{-i}) : s_{-i} \in \mathrm{EFR}_{-i}^{\mathrm{u},n-1}(t_{-i}) \right\} \right] = 1; \\ \quad \text{(iii) } \pi_i(\phi)[\theta_0, \theta_{-i}, s_{-i}] = \mu_i[\{(\theta_0, t_{-i}, s_{-i}) : \vartheta_{-i}(t_{-i}) = \theta_{-i}\}] \end{array} \right\}$$

and

$$\mathrm{EFR}_i^{\mathrm{u},n}(t_i) = \left\{ s_i \in R_i^{\mathrm{u}}(\vartheta_i(t_i)) : \exists \pi_i \in \Psi_i^{\mathrm{u},n}(t_i) \text{ s.t. } s_i \in r_i(\pi_i \mid \vartheta_i(t_i)) \right\}.$$

Finally, let $\mathrm{EFR}_i^{\mathrm{u}}(t_i) = \bigcap_{n\geq 0} \mathrm{EFR}_i^{\mathrm{u},n}(t_i)$. Observe that $\{\mathrm{EFR}_i^{\mathrm{u},n}(t_i)\}_{n\in\mathbb{N}} \subseteq R_i^{\mathrm{u}}(\vartheta_i(t_i))$ is a decreasing sequence for each t_i and converges in finitely many steps.

Claim 3.3. *For all $n \geq 0$, $\mathrm{EFR}_i^{\mathrm{u},n}(t_i) \subseteq \mathrm{EFR}_i^n(t_i)$. Hence, $\mathrm{EFR}_i^{\mathrm{u}}(t_i) \subseteq \mathrm{EFR}_i(t_i)$.*

Proof. For $n = 0$, we have $R_i^{\mathrm{u}}(\vartheta_i(t_i)) = \mathrm{EFR}_i^{\mathrm{u},0}(t_i) \subseteq \mathrm{EFR}_i^0(t_i) = S_i$. Suppose $\mathrm{EFR}_i^{\mathrm{u},n-1}(t_i) \subseteq \mathrm{EFR}_i^{n-1}(t_i)$. Then $\Psi_i^{\mathrm{u},n}(t_i) \subseteq \Psi_i^n(t_i)$ by definition, which implies $\mathrm{EFR}_i^{\mathrm{u},n}(t_i) \subseteq \mathrm{EFR}_i^n(t_i)$. □

Claim 3.4. *For all $n \geq 0$, $\mathrm{EFR}_i^{\mathrm{u},n}(t_i) \neq \varnothing$.*

Proof. We argue this by induction. Each $\mathrm{EFR}_i^{\mathrm{u},0}(t_i) = R_i^{\mathrm{u}}(\vartheta_i(t_i))$ is nonempty by assumption. Suppose $\mathrm{EFR}_i^{\mathrm{u},n-1}(t_i)$ is nonempty for every $i \in I$ and $t_i \in T_i$. Take any distribution $\mu_i \in \Delta(\Theta_0 \times T_{-i} \times S_{-i})$ such that conditions (i) and (ii) in the definition of $\Psi_i^{\mathrm{u},n}(t_i)$ are satisfied. We need to show that there exists a strategy $s_i \in \bigcup_{\pi_i \in \Psi_i^{\mathrm{u},n}(t_i)} r_i(\pi_i \mid$

$\vartheta_i(t_i)$) such that $s_i \in R_i^{\mathrm{u}}(\vartheta_i(t_i))$. Define $\nu_i \in \Delta\Big(\Theta_0 \times \bar{\Theta}_{-i}(\vartheta_i(t_i)) \times \mathcal{S}_{-i}\Big)$ by

$$\nu_i[\theta_0, \theta_{-i}, \{s_{-i}\}] = \sum_{t_{-i}:\vartheta_{-i}(t_{-i})=\theta_{-i}} \mu_i[\theta_0, t_{-i}, s_{-i}].$$

Because $\mathrm{EFR}_i^{\mathrm{u},n-1}(t_i)$ is a nonempty subset of $R_i^{\mathrm{u}}(\vartheta_i(t_i))$, this ν_j is well-defined and satisfies

$$\nu_i\left[\left\{(\theta_0, \theta_{-i}, R_{-i}) : R_{-i} \in \mathcal{R}_{-i}^{\uparrow}(\theta_{-i})\right\}\right] = 1.$$

By construction, we have $\Pi_i^n(\nu_i \mid \vartheta_i(t_i)) = \Psi_i^{\mathrm{u},n}(t_i)$ because for any $\pi_i \in \Pi_i^n(\nu_i \mid \vartheta_i(t_i))$, the initial probability distribution $\pi_i(\phi)$ is uniquely pinned down by μ_i. Because $\{\Pi_i^n(\nu_i \mid \vartheta_i(t_i))\}_{n\geq 0}$ is decreasing in n and $\{\mathcal{R}_i^{\uparrow,n}(\vartheta_i(t_i))\}$ defined by (3.1) converges to $\mathcal{R}_i^{\uparrow}(\vartheta_i(t_i))$, it must be that $\bigcup_{\pi_i \in \Psi_i^{\mathrm{u},n}(t_i)} r_i(\pi_i \mid \vartheta_i(t_i)) \in \mathcal{R}_i^{\uparrow}(\vartheta_i(t_i))$. By assumption, we have

$$\left(\bigcup_{\pi_i \in \Psi_i^{\mathrm{u},n}(t_i)} r_i(\pi_i \mid \vartheta_i(t_i))\right) \cap R_i^{\mathrm{u}}(\vartheta_i(t_i)) \neq \varnothing.$$

Therefore, $\mathrm{EFR}_i^{\mathrm{u},n}(t_i) \neq \varnothing$, which means the iterative procedure converges to a nonempty set of strategies for every $t_i \in T_i$. □

We claim that for every $s_i \in \mathrm{EFR}_i^{\mathrm{u}}(t_i)$, the singleton $\{s_i\} \in \mathcal{R}_i^{\mathrm{loc}}(t_i)$. We prove this by induction. Since any $s_i \in \mathrm{EFR}_i^{\mathrm{u},0}(t_i)$ is in the set $R_i^{\mathrm{u}}(\vartheta_i(t_i))$, we know $\{s_i\} \in \mathcal{R}_i^{\mathrm{loc},0}(t_i)$ by definition. Suppose $s_i \in \mathrm{EFR}_i^{\mathrm{u},n-1}(t_i)$ implies $\{s_i\} \in \mathcal{R}_i^{\mathrm{loc},n-1}(t_i)$ and take any $s_i \in \mathrm{EFR}_i^{\mathrm{u},n}(t_i)$. Constructions of ν_i and $\tilde{\nu}_i$ in the proof of Proposition 3.5, with obvious modifications, can be used to show that $\{s_i\} \in \mathcal{R}_i^{\mathrm{loc},n}(t_i)$ (see Appendix 3.5.5.2).

Therefore, our first step shows that for any finite type $t_i \in T_i^*$, there is some strategy $s_i \in \mathrm{EFR}_i(t_i)$ that can be uniquely selected. The second step is standard: Let $T_i^{\mathrm{f}} \subseteq T_i^*$ be the collection of all finite types of player i. Since $t_i \in \mathrm{cl}(\mathcal{U}_i)$ for all $t_i \in T_i^{\mathrm{f}}$, we have $T_i^{\mathrm{f}} \subseteq \mathrm{cl}(\mathcal{U}_i)$. Because T_i^{f} is dense in T_i^* (Mertens and Zamir, 1985), we know that $T_i^* = \mathrm{cl}(T_i^{\mathrm{f}}) \subseteq \mathrm{cl}(\mathcal{U}_i)$, which means $\mathcal{U}_i$ is also dense in T_i^*.

3.6 Bibliography

Basu, K. and J. W. Weibull (1991): "Strategy Subsets Closed under Rational Behavior," *Economic Letters*, 36(2), 141–146.

Battigalli, P. (1997): "On Rationalizability in Extensive Games," *Journal of Economic Theory*, 74(1), 40–61.

Battigalli, P. and M. Siniscalchi (2002): "Strong Belief and Forward Induction Reasoning," *Journal of Economic Theory*, 106(2), 356–391.

——— (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3(1), Article 3.

Brandenburger, A. and E. Dekel (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59(1), 189–198.

Carlsson, H. and E. van Damme (1993): "Global Games and Equilibrium Selection," *Econometrica*, 61(5), 989–1018.

Chen, Y.-C. (2012): "A Structure Theorem for Rationalizability in the Normal Form of Dynamic Games," *Games and Economic Behavior*, 75(2), 587–597.

Chen, Y.-C., S. Takahashi, and S. Xiong (2014): "The Robust Selection of Rationalizability," *Journal of Economic Theory*, 151, 448–475.

——— (2021): "Robust Refinement of Rationalizability with Arbitrary Payoff Uncertainty," *Working paper*.

Cho, I.-K. and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179–222.

Dekel, E. and D. Fudenberg (1990): "Rational Behavior with Payoff Uncertainty," *Journal of Economic Theory*, 52(2), 243–267.

Dekel, E., D. Fudenberg, and S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2(1), 15–40.

Fudenberg, D., D. M. Kreps, and D. K. Levine (1988): "On the Robustness of Equilibrium Refinements," *Journal of Economic Theory*, 44(2), 354–380.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, Cambridge, Mass. and London: The MIT Press.

Germano, F., J. Weinstein, and P. Zuazo-Garin (2020): "Uncertain Rationality, Depth of Reasoning and Robustness in Games with Incomplete Information," *Theoretical Economics*, 15(1), 89–122.

Heifetz, A. and W. Kets (2018): "Robust Multiplicity with a Grain of Naiveté," *Theoretical Economics*, 13(1), 415–465.

Mertens, J.-F. and S. Zamir (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

Morris, S. and H. S. Shin (2000): "Rethinking Multiple Equilibria in Macroeconomic Modeling," *NBER Macroeconomics Annual*, 15, 139–161.

Osborne, M. J. and A. Rubinstein (1994): *A Course in Game Theory*, Cambridge, Mass. and London: The MIT Press.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52(4), 1029–1050.

Penta, A. (2012): "Higher Order Uncertainty and Information: Static and Dynamic Games," *Econometrica*, 80(2), 631–660.

——— (2013): "On the Structure of Rationalizability for Arbitrary Spaces of Uncertainty," *Theoretical Economics*, 8(2), 405–430.

Penta, A. and P. Zuazo-Garin (2022): "Rationalizability, Observability and Common Knowledge," *Review of Economic Studies*, 89(2), 948–975.

Piermont, E. and P. Zuazo-Garin (2021): "Heterogeneously Perceived Incentives in Dynamic Environments: Rationalization, Robustness and Unique Selections," *Working paper*.

Rubinstein, A. (1989): "The Electronic Mail Game: Strategic Behavior under "Almost Common Knowledge"," *American Economic Review*, 79(3), 385–391.

Weinstein, J. and M. Yildiz (2007): "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 75(2), 365–400.

——— (2011): "Sensitivity of Equilibrium Behavior to Higher-Order Beliefs in Nice Games," *Games and Economic Behavior*, 72(1), 288–300.

——— (2013): "Robust Predictions in Infinite-Horizon Games—an Unrefinable Folk Theorem," *Review of Economic Studies*, 80(1), 365–394.

Ziegler, G. (2022): "Informational Robustness of Common Belief in Rationality," *Games and Economic Behavior*, 132, 592–597.