Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Interactive Epistemology and Solution Concepts for Games with Asymmetric Information

Pierpaolo Battigalli, Alfredo Di Tillio, Edoardo Grillo and Antonio Penta

Working Paper n. 340

First version: July 31, 2008; This version: December 9, 2009

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Interactive Epistemology and Solution Concepts for Games with Asymmetric Information

Pierpaolo Battigalli* Bocconi

Alfredo Di Tillio Bocconi

Edoardo Grillo Princeton

Antonio Penta UPenn

First version: July 31, 2008.
This version: December 9, 2009

Abstract

We use an interactive epistemology framework to provide a systematic analysis of some solution concepts for games with asymmetric information. We characterize solution concepts using *expressible* epistemic assumptions, represented as events in the universal type space generated by primitive uncertainty about the payoff relevant state, payoff irrelevant information, and actions. In most of the paper we adopt an *interim* perspective, which is appropriate to analyze genuine incomplete information. We relate Δ-rationalizability (Battigalli and Siniscalchi, 2003) to interim correlated rationalizability (Dekel, Fudenberg, and Morris, 2007) and to rationalizability in the interim strategic form. We also consider the *ex ante* perspective, which is appropriate to analyze asymmetric information about an initial chance move. We prove the equivalence between interim correlated rationalizability and an *ex ante* notion of correlated rationalizability.

KEYWORDS: asymmetric information, type spaces, Bayesian games, rationalizability.
J.E.L. CLASSIFICATION NUMBERS: C72, D82.

1 Introduction

In the last few years, ideas related to rationalizability have been increasingly applied to the analysis of games with asymmetric information.[1] Yet there seems to be no "canonical" definition of

*Corresponding author. Address: IGIER and Ettore Bocconi Dept. of Economics, Bocconi University, Via Röntgen 1, 20136 Milano, Italy. E-mail: pierpaolo.battigalli@unibocconi.it.

[1]See Battigalli (2003, section 5), and Battigalli and Siniscalchi (2003, section 6) for references to applications of rationalizability to models of reputation, auctions and signaling. Bergemann and Morris (2005) apply a notion of iterated dominance to robust implementation. Carlsson and van Damme (1993) show that global games can be solved by iterated dominance (see also Morris and Shin (2007) for a recent evaluation of this result and its applications). Since even *strict* rationalizability lacks lower hemi-continuity with respect to belief hierarchies, this work spurred a literature on the robustness of rationalizable behavior to small perturbations of beliefs — see Dekel, Fudenberg, and Morris (2006), Weinstein and Yildiz (2007), Ely and Pęski (2008), Chen, Di Tillio, Faingold, and Xiong (2009), and for dynamic games, Penta (2009).

rationalizability for this class of games. Some authors apply rationalizability to the strategic form of Bayesian games, but different strategic forms (*ex ante* and *interim*) yield different results. Furthermore, it has been noticed by Ely and Pęski (2006) and Dekel, Fudenberg, and Morris (2007) that adding redundant types — types with the same payoff-relevant private information and the same hierarchy of beliefs — may enlarge the set of rationalizable outcomes. Dekel, Fudenberg, and Morris (2007) introduce a notion of rationalizability for Bayesian games that is weaker than rationalizability on the interim strategic form and is invariant to the addition of redundant types. Other authors put forward and apply notions of rationalizability that do not rely on the full specification of a Bayesian game and hence of a type space — see Battigalli (2003), Battigalli and Siniscalchi (2003, 2007), and Bergemann and Morris (2005, 2007).

What are the assumptions underlying these solution concepts? Why do they differ? How are they related? The existing literature provides partial and disconnected answers. In this paper we use interactive epistemology to provide a systematic analysis of the above mentioned notions of rationalizability for games with asymmetric information, interpreted either as games with genuine incomplete information or games with imperfect information about an initial chance move. In the remainder of this introduction we illustrate the issues concerning the various definitions of rationalizability and give an overview of our results. The rest of the paper is then structured as follows: section 2 introduces the basic framework; section 3 provides epistemic characterizations of solution concepts via expressible assumptions about rationality and beliefs; section 4 relates the ex ante and interim approaches to rationalizability; finally, section 5 offers some concluding remarks, including a discussion of the most related literature; the various appendices contain the proofs not given in the main text and some technical constructions and results.

1.1 Rationalizability for Bayesian games

To simplify the analysis we focus on two-person, simultaneous-move games, thus removing any issues of belief revision and correlation among opponents in the eyes of a player. Here we recall the notions of Bayesian game and belief hierarchies, discuss two issues concerning the received notions of rationalizability for Bayesian games, and briefly explain our approach to solution concepts.

Bayesian games and belief hierarchies

In a game of incomplete information, payoffs are affected by a parameter $\theta \in \Theta$ that is not common knowledge, though the set Θ and how θ affects payoffs *are* common knowledge. With no essential loss of generality, we assume that $\Theta = \Theta_0 \times \Theta_1 \times \Theta_2$ and each player $i = 1, 2$ knows the component θ_i of $\theta = (\theta_0, \theta_1, \theta_2)$. The standard methodology to analyze such situations is to model the players' beliefs about θ and about each other's beliefs by means of a *type space* à la Harsanyi (1967-68), that is, a structure $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i=1,2} \rangle$ that specifies, for each player i, a set of types T_i and mappings $\boldsymbol{\theta}_i : T_i \to \Theta_i$ and $\boldsymbol{\pi}_i : T_i \to \Delta(\Theta_0 \times T_{-i})$. (Throughout the paper we use boldface symbols to denote functions that can be interpreted as random variables.) These mappings deliver, for each Harsanyi type t_i of each player i, a Θ-*hierarchy* of beliefs, that is, a first-order belief $\boldsymbol{\pi}_i^1(t_i) \in \Delta(\Theta_0 \times \Theta_{-i})$, a second-order belief $\boldsymbol{\pi}_i^2(t_i) \in \Delta(\Theta_0 \times \Theta_{-i} \times \Delta(\Theta_0 \times \Theta_i))$, and so on (see section 2.2).

The type space and the payoff functions parametrized by θ form a *Bayesian game*. One can find prior beliefs $\Pi_i \in \Delta(\Theta_0 \times T_1 \times T_2)$ for each player i such that the beliefs of each type t_i are given by $\boldsymbol{\pi}_i(t_i)[\cdot] = \Pi_i[\cdot|t_i] \in \Delta(\Theta_0 \times T_{-i})$. This induces an extensive form game where an initial chance moves selects (θ_0, t_1, t_2), each player i assigns (subjective) probabilities to chance moves according to the prior Π_i and then learns his type t_i before choosing his action. A strategy profile specifies an action for each type of each player. If we take the strategic form of this game, we obtain the *ex ante strategic form* of the original Bayesian game, which is well defined even if $\Pi_1 \neq \Pi_2$ — the expected payoff for i is computed using Π_i, and in effect the particular choice of Π_i is immaterial. (Assuming T_1 and T_2 are finite, any strictly convex combination of the measures $\{\boldsymbol{\pi}_i(t_i)\}_{t_i \in T_i}$ works.) If instead we treat different types t_i as different players who compute expected payoffs using the interim beliefs $\Pi_i[\cdot|t_i]$, we obtain the *interim strategic form* of the Bayesian game (Osborne and Rubinstein, 1994, pp. 24–26). A strategy profile is an equilibrium of the ex ante strategic form if and only if it is a Nash equilibrium of the interim strategic form. Therefore both equilibrium concepts can be taken as definitions of equilibrium for the Bayesian game.

Two issues concerning rationalizability for Bayesian games

In contrast to the complete information case, there is no textbook definition of rationalizability for static games with incomplete information, represented as Bayesian games. While it seems natural to transform such games into strategic form games and apply standard rationalizability, there is more than one way to do this: one can consider the *ex ante* or the *interim* strategic form. Moreover, unlike Bayesian Nash equilibrium, rationalizability in the ex ante strategic form is a *refinement* of rationalizability in the interim strategic form.[2] The following example shows why.

Example 1. Assume that $\Theta = \{\theta_0\} \times \{\theta_1', \theta_1''\} \times \{\theta_2\}$, so that Ann (player 1) knows θ while Bob (player 2) does not, and yet only Bob's payoff depends on θ, as shown in the payoff tables below.

	l	r
u	6, 6	0, 4
m	4, 0	4, 4
d	0, 0	6, 4

θ_1'

	l	r
u	6, 0	0, 4
m	4, 0	4, 4
d	0, 6	6, 4

θ_1''

Assuming that Bob believes $\Pr[\theta_1'] = \Pr[\theta_1''] = 1/2$ and that there is common (probability one)

[2] This holds under weak conditions on players' (subjective) priors: either (a) priors have a common support, or (b) each player assigns positive prior probability to each one of his types. The difference between ex ante and interim rationalizability is related to the difference between two notions of extensive form rationalizability: the more restrictive one assumes that a player has an initial conjecture about the opponent's strategy, which may be revised only after receiving some information about the opponent's behavior; the less restrictive, adopted by Pearce (1984), drops the initial conjecture and allows a player to have different conjectures at different information sets even if they only reflect information about chance moves. When we consider the extensive form of a static Bayesian game, the first solution concept yields ex ante rationalizability and the second one yields interim rationalizability. To the best of our knowledge, Battigalli (1988, pp. 719–720, Footnote 1) is the first published work pointing out the difference.

belief of this, we obtain a Bayesian game where each Harsanyi type is uniquely determined by the corresponding private information. Pick any conjecture $\mu \in \Delta(\{l, r\})$ about Bob's action. Action u by Ann is a best reply only if $\mu[r] \leq 1/3$, while d is a best reply only if $\mu[r] \geq 2/3$. Thus the two strategies that specify u for one type of Ann and d for the other, cannot be ex ante best replies to any conjecture μ. If Bob assigns zero probability to these strategies, the expected payoff of l is at most 3, and hence l is not ex ante rationalizable. On the other hand, interim rationalizability regards θ_1' and θ_1'' as different "replicas" of Ann: θ_1' may believe $\mu[r] \geq 2/3$ while θ_1'' may believe $\mu[r] \leq 1/3$ (or vice versa). Thus, in the second iteration of the interim rationalizability procedure Bob may assign probability close to 1 to θ_1' choosing u and θ_1'' choosing d, and hence choose l as a best response. This implies that every action is interim rationalizable.

The well known difference between ex ante and interim rationalizability, as illustrated in this example, has been accepted as a natural consequence of the fact that interim rationalizability allows different types of the same player to hold different conjectures. However, we maintain that the difference between the two notions should be disturbing.

Question 1 (ex ante *vs* interim). *Rationalizability should capture the behavioral consequences of the assumption that players are expected payoff maximizers and have common belief in this fact. Moreover, ex ante expected payoff maximization is equivalent to interim expected payoff maximization.*[3] *Then, how can we explain the fact that ex ante and interim rationalizability give different results?*

Another well known fact concerns the rationalizable actions of redundant types. A type space is *redundant* if there are two types t_i, t_i' with the same private information and Θ-hierarchy:

$$(\boldsymbol{\theta}_i(t_i), \boldsymbol{\pi}_i^1(t_i), \boldsymbol{\pi}_i^2(t_i), \ldots) = (\boldsymbol{\theta}_i(t_i'), \boldsymbol{\pi}_i^1(t_i'), \boldsymbol{\pi}_i^2(t_i'), \ldots).$$

A change in the type space that has the only effect of adding redundancy may nevertheless expand the equilibrium actions. This is best understood for the simple case of games with *complete* information, i.e. when Θ is a singleton. Even in this case we can specify type spaces with multiple (and hence necessarily redundant) types for each player, and obtain Bayesian Nash equilibria that are *subjective correlated equilibria*, but not Nash equilibria of the original complete information game. However, adding redundant types to a complete information game does not change the set of rationalizable actions. More generally, the set of interim rationalizable actions is invariant to the addition of redundant types whenever interim payoff uncertainty only concerns the payoff information of the opponent, i.e. when Θ_0 is a singleton (see Corollary 2 in Dekel, Fudenberg, and Morris (2007) and our Remark 3 in section 3.4). Thus one may wonder why, when instead there is nontrivial residual payoff uncertainty (i.e. when Θ_0 has more than one element), adding redundant types can expand the set of rationalizable actions, and types with the same private information and Θ-hierarchy can have different interim rationalizable actions. This is illustrated by the following example, borrowed from Dekel, Fudenberg, and Morris (2007).[4]

Example 2. Ann and Bob play a betting game where the outcome depends on the state of nature,

[3] Interim maximization implies ex ante maximization, and under the same mild assumptions as in footnote 2, also the converse is true.

[4] See also Ely and Pęski (2006). Liu (2009) and Sadzik (2007) analyze related issues of invariance of solution concepts to redundancies.

about which they have no private information: $\Theta = \{\theta_0', \theta_0''\} \times \{\theta_1\} \times \{\theta_2\}$. Ann wins if both bet and θ_0' occurs, while Bob wins if both bet and θ_0'' occurs. Placing a bet costs 4. The loser gives 12 to the winner. Payoffs are summarized by the tables below.

	B	N
B	$8, -16$	$-4, 0$
N	$0, -4$	$0, 0$

θ_0'

	B	N
B	$-16, 8$	$-4, 0$
N	$0, -4$	$0, 0$

θ_0''

Assume that it is common belief that each player attaches equal probabilities to θ_0' and θ_0''. The simplest Bayesian game representing this situation has only one type for each player. The ex ante and interim strategic forms coincide and betting is dominated, hence not rationalizable:

	B	N
B	$-4, -4$	$-4, 0$
N	$0, -4$	$0, 0$

Now take the type space with two types for each player generated by the following common prior:

	t_2'	t_2''
t_1'	$1/4$	0
t_1''	0	$1/4$

θ_0'

	t_2'	t_2''
t_1'	0	$1/4$
t_1''	$1/4$	0

θ_0''

It is clear that, as before, it is common belief that θ_0' and θ_0'' are considered equally likely, therefore we are just adding redundant types. But in the Bayesian game induced by this type space, betting is rationalizable. Since ex ante rationalizability implies interim rationalizability, to see this it suffices to show that there are ex ante rationalizable strategies where at least one type bets. Let XY denote the strategy where t_i' chooses X and t_i'' chooses Y. The ex ante strategic form is as follows:

	BB	BN	NB	NN
BB	$-4, -4$	$-4, -2$	$-4, -2$	$-4, 0$
BN	$-2, -4$	$\mathbf{1}, -5$	$-5, \mathbf{1}$	$-2, 0$
NB	$-2, -4$	$-5, \mathbf{1}$	$\mathbf{1}, -5$	$-2, 0$
NN	$0, -4$	$0, -2$	$0, -2$	$\mathbf{0}, \mathbf{0}$

Note that BB is dominated, but the set of strategy profiles $\{BN, NB, NN\} \times \{BN, NB, NN\}$ has the best response property (Pearce, 1984): as the highlighted payoffs indicate, each strategy in the subset of player i is a best response to some strategy in (and hence to some belief on) the set of player $-i$. Therefore BN and NB are rationalizable, which implies that betting is rationalizable. ◊

Question 2 (non-invariance). *Adding redundant types can expand the rationalizable set of the strategic form. Does interim rationalizability capture more than just common belief of expected payoff maximization in a situation of incomplete information? How are the additional hidden assumptions related to the presence of redundant types?*

Addressing the questions: expressible assumptions about rationality and beliefs

The somewhat puzzling facts illustrated above should make us suspicious about solution concepts mechanically obtained by applying a known solution algorithm (rationalizability) to the strategic forms of Bayesian games. The problem with these notions is that they are not completely transparent because, unlike rationalizability in games of complete information, they have not been characterized using *expressible assumptions* about rationality and beliefs.

To see what we mean, let us first consider games with complete information, the special case where Θ is a singleton. Tan and Werlang (1988) show that an action is rationalizable if and only if it is consistent with rationality, i.e. expected payoff maximization, and common belief of rationality. (Brandenburger and Dekel (1987) prove a related result.) These assumptions can be expressed in a language that starts from primitive terms (actions), terms derived from the primitives, like beliefs about actions, and terms derived from the primitives and other derived terms, like joint beliefs about the actions and beliefs of others. As explained in Heifetz and Samet (1998), such assumptions can be represented as (and indeed identified with) measurable subsets of a *canonical* state space, where each state specifies the players' actions and hierarchies of beliefs about actions — beliefs about others' actions, beliefs about others' actions and beliefs, and so on. Every state satisfying a natural coherency property is represented in this state space, hence the set of states satisfying an assumption like "each player maximizes his expected payoff" represents exactly that assumption and nothing more. Of course, we may want to consider other assumptions beside rationality and common belief in rationality. For example, in games with more than two players, we can assume that each player regards the actions of his opponents as stochastically independent random variables and that there is common belief of this fact too. Indeed, while not necessarily compelling in every application, this assumption is also expressible.[5]

We can understand rationalizability in games of incomplete information applying the same methodology. Is it possible to characterize, say, interim rationalizability by means of expressible assumptions about rationality and beliefs? What solution concept do we obtain if we assume (only) rationality and common belief of rationality? Here, too, answering these questions requires that we specify the *primitive terms* of our language, which now must include not only actions, but also the payoff state θ.[6] But players may have further private information that can be thought to be correlated with θ. Economic examples abound: geological information and satellite photographs of a tract of land on sale are thought to be correlated with the value of the recoverable resources, expert reports on an object are thought to be correlated with the value of this object, personality traits and propensities may be thought to be correlated with ability, etc. The applied theorist who models a particular situation typically specifies these payoff irrelevant, but strategically relevant aspects. Thus, in our abstract framework, we let ξ_i denote a realization of all the payoff-irrelevant

[5] Rationality, stochastic independence of beliefs about others, and common belief of both together characterize independent rationalizability, a refinement of correlated rationalizability (which in turn is equivalent to iterated dominance). Independent rationalizability was introduced by Bernheim (1984) and Pearce (1984). Rationalizability with correlated beliefs became the default solution concept later on (see e.g. Osborne and Rubinstein, 1994, Ch. 4).

[6] The payoff state θ parametrizes the mapping from actions to payoffs, and it is informally assumed that this parametrization is common knowledge between players.

(though potentially strategically relevant) aspects known by player i. The pair (θ_i, ξ_i) describes i's *private information*.[7] Note that the payoff-irrelevant information ξ_i is strategically relevant for two (related) reasons: (a) Ann's action may depend on ξ_{Ann}, (b) Bob may believe that ξ_{Ann} is correlated with θ_0, thus inducing a potential correlation between θ_0 and a_{Ann}. Furthermore, explicitly taking into account the players' (payoff-relevant and payoff-irrelevant) information allows us to express restrictions on beliefs — information-based conditional independence, see below — that otherwise would not be expressible. (See section 5 for further discussion.)

Thus the basic elements of the language are given by a structure $\mathcal{E}$, the economic *environment* specified by the modeler, that lists players $i \in I$, actions $a_i \in A_i$, residual payoff uncertainty $\theta_0 \in \Theta_0$, payoff-relevant and payoff irrelevant information $x_i = (\theta_i, \xi_i) \in \Theta_i \times \Xi_i = X_i$, and payoff functions $\boldsymbol{g}_i : \Theta \times A \to \mathbb{R}$. In this framework a first-order belief of player i concerns $(\theta_0, x_{-i}, a_{-i})$, a second-order belief concerns (the same as before and) the first-order belief of $-i$, and so on. Thus, following Heifetz and Samet (1998), we define an *expressible assumption* as a measurable subset of the canonical space — in section 3.1 we briefly review Heifetz and Samet's definitions and explain why this is meaningful — and we state assumptions concerning (i) first-order beliefs and the relationship between players' actions, information and first-order beliefs, (ii) second-order beliefs, (iii) third-order beliefs, etc. Then we show that the behavioral consequences of these assumptions can be derived by appropriate iterative deletion procedures, and we relate these procedures to old and new notions of "rationalizability". An example of assumption about first-order beliefs is that they satisfy *information-based conditional independence*: the beliefs of i are such that, conditional on x_{-i}, there is no residual correlation between θ_0 and a_{-i}. The standard assumption about the relationship between actions, information and first-order beliefs is that players are *rational*, i.e. maximize their expected payoff, given their information and first-order beliefs. Second-order beliefs are then assumed to assign probability one to the previously stated assumptions, and so on.

1.2 Preview of results

Our exploration of solution concepts begins with Δ-rationalizability, an "umbrella notion" defined on the environment $\mathcal{E}$ and parametrized by information-dependent restrictions Δ on players' beliefs about the primitives (Battigalli, 2003 and Battigalli and Siniscalchi, 2003, 2007). Formally, for each player $i = 1, 2$ and each $x_i \in X_i$ we postulate a set $\Delta_{x_i} \subseteq \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ of possible beliefs about the exogenous state and the opponent's action, and we define an iterative deletion procedure that takes such restrictions on beliefs into account. Next we consider interim *correlated* rationalizability and interim *independent* rationalizability in the Bayesian game induced by some type space T (T-ICR and T-IIR, Dekel, Fudenberg, and Morris, 2007). IIR is equivalent to rationalizability in the interim strategic form and requires that the players' beliefs satisfy a conditional independence property (whereas ICR does not): conditional on the opponent's type t_{-i}, there is no residual correlation between θ_0 and a_{-i}. Finally, we consider *ex ante* notions of rationalizability and relate them to

[7] The environment $\mathcal{E}$ is similar to what Battigalli and Siniscalchi (2007) call "game with payoff uncertainty" and Bergemann and Morris (2007) call "belief-free incomplete information game". But, unlike these papers, we make explicit the difference between payoff-relevant and payoff-irrelevant private information.

corresponding interim solution concepts. A partial list of our results follows. (The first is already known and we report it for completeness only.)

Result 1 (Lemma 1, cf. Battigalli and Siniscalchi, 2007, Propositions 1,2) *Δ-rationalizability is characterized by the following assumptions: (a) players are rational, (b) their first-order beliefs satisfy the restrictions Δ, and (c) there is common belief of (a) and (b).*

Say that a type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i \in I} \rangle$ has *information types* if, for each i, the set of types T_i is (isomorphic to) X_i. In this case we can obtain a set Δ of (information-dependent) restrictions on beliefs about the primitives that exactly identifies T: for each i and x_i, the set Δ_{x_i} is the set of measures $\mu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ such that (M) $\mathrm{marg}_{\Theta_0 \times X_{-i}} \mu_i = \boldsymbol{\pi}_i(x_i)$ (that is, $\mathrm{marg}_{\Theta_0 \times X_{-i}} \mu_i$ is the belief of information-type x_i in T).

Result 2 (Proposition 1) *If a type space T has information types and Δ is the set of restrictions derived from T, then Δ-rationalizability coincides with T-ICR.*

A corollary of Results 1 and 2 is that if T has information types, then T-ICR can be characterized by expressible assumptions about rationality and interactive beliefs. Indeed, it turns out that such a characterization is possible for every type space, and that the ICR actions of a type depend only on its expressible features, which we call the *explicit type*. (Indeed, they depend only on the private information about θ and on the Θ-hierarchy that the explicit type induces — see section 2.2.)

Result 3 (Theorem 1, cf. Dekel, Fudenberg, and Morris, 2007, Proposition 2) *ICR is characterized by rationality and common belief of rationality in the following sense: for each type space T and each type t_i in T, the set of T-ICR actions of t_i is the set of actions consistent with rationality, common belief of rationality and player i having explicit type $(\boldsymbol{\theta}_i(t_i), \boldsymbol{\pi}_i^1(t_i), \boldsymbol{\pi}_i^2(t_i), \ldots)$.*

Fix a type space T with information types. Say that a set Δ of restrictions on first-order beliefs is *CI-derived* from T if, for each player i and each type x_i, Δ_{x_i} is the set of measures $\mu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ such that (M) $\mathrm{marg}_{\Theta_0 \times X_{-i}} \mu_i = \boldsymbol{\pi}_i(x_i)$ and (CI) $\mu_i[x_{-i}] > 0$ implies $\mu_i[\theta_0, a_{-i} | x_{-i}] = \mu_i[\theta_0 | x_{-i}] \mu_i[a_{-i} | x_{-i}]$, that is, i believes that θ_0 and a_{-i} are independent conditional on x_{-i}. The following result shows that T-IIR is equivalent to rationalizability in the interim strategic form of the Bayesian game induced by T (see Appendix B). Furthermore, if T has information types, then computing the rationalizable strategies in the interim strategic form of the Bayesian game induced by T amounts to imposing the conditional independence restriction (CI) on top of the restrictions (M) implied by the type space.

Result 4 (Remark 2 and Proposition 3) *Fix a type space T. The set of T-IIR actions of every type t_i in T is the set of actions that are rationalizable for player/type t_i in the interim strategic form of the Bayesian game induced by T. If T has information types and Δ is the set of restrictions CI-derived from T, then Δ-rationalizability coincides with T-IIR.*

(Note the parallel between Result 2 and the second statement in Result 4: Requiring that Δ be derived from T delivers ICR, whereas requiring that Δ be CI-derived from T gives IIR.) As a corollary of Results 1 and 4 we obtain a characterization of IIR via expressible assumptions on rationality and interactive beliefs for the special case of a type space with information types:

Result 5 (Corollary 3, cf. Dekel, Fudenberg, and Morris, 2007, Proposition 3) *If a type space T has* *information types, then for every player i and every (information) type x_i in T the set of T-IIR actions of x_i is the set of actions consistent with rationality, conditional independence, common belief of rationality and conditional independence, and player i having explicit type $(\boldsymbol{\theta}_i(x_i), \boldsymbol{\pi}_i^1(x_i), \boldsymbol{\pi}_i^2(x_i), \ldots)$.*

We were not able to provide a general characterization of IIR via expressible assumptions. The difficulty lies in the fact that Harsanyi types are self-referentially defined: a type is a belief about the payoff state and the type of the opponent. Whereas each type t_i encodes the payoff-relevant information $\boldsymbol{\theta}_i(t_i)$ and the belief hierarchy $(\boldsymbol{\pi}_i^1(t_i), \boldsymbol{\pi}_i^2(t_i), \ldots)$, which are expressible, we have seen in Example 2 that specifying these (and only these) features of a type does not allow to determine the set of interim rationalizable actions. Insofar as interim rationalizability depends on non-expressible features of types, we cannot give it a characterization via expressible assumptions (see also our discussion of Ely and Pęski (2006) in section 5). However, we do obtain a characterization as in Result 3 as a consequence of the following:

Result 6 [Remark 3, Corollary 2] *If in $\mathcal{E}$ there is distributed knowledge of the payoff state, i.e. if Θ_0 is a singleton, then T-ICR and T-IIR coincide for every type space T. Therefore the set of T-IIR actions of each type t_i is the set of actions consistent with rationality, common belief of rationality and player i having explicit type $(\boldsymbol{\theta}_i(t_i), \boldsymbol{\pi}_i^1(t_i), \boldsymbol{\pi}_i^2(t_i), \ldots)$.*

Finally we turn to the relationship between the ex ante and interim perspectives. First we note that ex ante rationalizability makes sense only if types correspond to information that players can learn, i.e. if we have information types. Given the economic environment $\mathcal{E}$ and a type space T with information types, we derive the 3-player strategic form where an indifferent, fictitious player (nature) chooses the profile (θ_0, x_1, x_2) and each $i = 1, 2$ chooses a strategy $\boldsymbol{f}_i : X_i \to A_i$. Then we define a correlated rationalizability solution procedure subject to the restriction that each player assigns positive probability to each of his own information types and has beliefs consistent with T. Letting $\boldsymbol{F}_{-i}$ denote the set of mappings from X_{-i} to A_{-i}, and given a belief $\mu_i \in \Delta(\Theta_0 \times X_1 \times X_2 \times \boldsymbol{F}_{-i})$ of player i, the conditional belief $\mu_i[\cdot|x_i]$ is then well defined and satisfies $\mathsf{marg}_{\Theta_0 \times X_{-i}} \mu_i[\cdot|x_i] = \boldsymbol{\pi}_i(x_i)$ for every x_i. We call this solution T-*ex ante correlated rationalizability*, or T-ACR.

Result 7 [Theorem 2] *For every type space T with information types, T-ACR is equivalent to T-ICR.*[8]

The intuition of this result is that ex ante expected payoff maximization is equivalent to interim expected payoff maximization, and since the ex ante beliefs of i about nature and $-i$ may be correlated, the interim belief of i about the strategy of $-i$ may depend on the information type x_i. Of course, one can define a notion of *ex ante independent* rationalizability by imposing that i's ex ante beliefs about the fictitious player (nature) and the real opponent satisfy ex ante independence: $\mu_i = \mu_i^0 \times \mu_i^{-i}$ with $\mu_i^0 \in \Delta(\Theta_0 \times X_1 \times X_2)$ and $\mu_i^{-i} \in \Delta(\boldsymbol{F}_{-i})$. It is easily verified that this is rationalizability on the ex ante strategic form.

An answer to Question 1. Ex ante and interim rationalizability can be compared for Bayesian games with information types. We show that they both rely on rationality, independence and common certainty of rationality and independence. *Ex ante rationalizability is stronger than interim rationalizability because ex ante independence is stronger than interim independence:* Ex ante independence

[8]The mathematical result actually holds for every type space, but it is meaningful for spaces with information types.

means that $\mu_i = \mu_i^0 \times \mu_i^{-i}$, therefore $\mu_i[\cdot|x_i] = (\text{marg}_{\Theta_0\times X_{-i}}\,\mu_i^0[\cdot|x_i]) \times \mu_i^{-i}$ and each information type of i has the same conjecture, μ_i^{-i}, about the strategy of the opponent. Interim independence instead means that, for each x_i, $\mu_i[\cdot|x_i] = (\text{marg}_{\Theta_0\times X_{-i}}\,\mu_i[\cdot|x_i]) \times (\text{marg}_{F_{-i}}\,\mu_i[\cdot|x_i])$, which implies conditional independence: $\mu_i[\theta_0, a_{-i}|x_i, x_{-i}] = \mu_i[\theta_0|x_i, x_{-i}] \times \mu_i[a_{-i}|x_i, x_{-i}]$ for each a_{-i} and x_{-i} such that $\mu_i[x_{-i}|x_i] > 0$; thus, different information types may hold different conjectures about the strategy of the opponent. *Removing the independence restriction removes any difference between (the appropriate versions of) ex ante and interim rationalizability.*

An answer to Question 2. Adding redundant types is certainly meaningful when types correspond to actual information (we do not exclude that it may be meaningful also in other circumstances). Payoff-irrelevant information may be thought to be correlated with θ_0. As Example 2 shows, this is possible even if payoff-irrelevant information does not affect hierarchies of beliefs about θ. Since actions may depend on this information, it is possible that i's beliefs satisfy conditional independence when considering all the information of the opponent, and yet when they are conditioned on (θ_{-i} and) the hierarchy of beliefs about θ of the opponent they exhibit correlation between θ_0 and a_{-i}. Therefore, *considering (payoff-irrelevant) information that does not affect the hierarchy of beliefs about the payoff state (redundant information types) decreases the bite of the conditional independence assumption and hence expands the set of interim rationalizable actions.* On the other hand, interim correlated rationalizability is invariant to the addition of redundant types because it allows conditional correlation and therefore adding redundant types has no effect.

2 Preliminaries

In this section we introduce the basic elements of our analysis (section 2.1) and define belief hierarchies, type spaces and information types (section 2.2).

2.1 The economic environment

The basic ingredients of our model are collected in an *economic environment*, that is, a structure

$$\mathcal{E} = \left\langle \Theta_0, (\Theta_i, \Xi_i, A_i, \boldsymbol{g}_i)_{i\in I} \right\rangle$$

where:

- $I = \{1, 2\}$ is the set of *players*, and for each $i \in I$ we let $-i$ denote the other player.
- A_i is the finite set of feasible *actions* of player i, and we define $A = A_1 \times A_2$.
- Θ_i and Ξ_i are finite sets representing, respectively, the payoff-relevant and payoff-irrelevant *private information* of player i; we define $X_i = \Theta_i \times \Xi_i$ and $X = X_1 \times X_2$ and refer to an element $x_i \in X_i$ as an *information type* of player i.
- Θ_0 is a finite set representing payoff-relevant uncertainty that persists even after pooling the players' private; we let $\Theta = \Theta_0 \times \Theta_1 \times \Theta_2$ and we refer to an element $\theta \in \Theta$ as a *payoff state*,

to an element $(\theta_0, x) \in \Theta_0 \times X$ as an *exogenous external state*, and to an element $(\theta_0, x, a) \in \Theta_0 \times X \times A$ as an *external state*.

- $\boldsymbol{g}_i : \Theta \times A \to \mathbb{R}$ is the *payoff function* of player i.

The environment $\mathcal{E}$ will be kept fixed throughout the paper and informally assumed to be common knowledge between players.[9] According to the *incomplete information interpretation*, interaction starts at the *interim* stage in a given exogenous external state $(\theta_0, x_1, x_2) \in \Theta_0 \times X$. Each player i knows (only) x_i and chooses some $a_i \in A_i$. The actual payoff function $\boldsymbol{g}_i(\theta, \cdot) : A \to \mathbb{R}$ of player i is not commonly known, unless Θ is a singleton; in the latter case there is *complete information*, that is, common knowledge of the payoff state, whereas if Θ_0 is a singleton we say that there is *distributed knowledge* of the payoff state.[10] According to the *complete* but *asymmetric information interpretation*, interaction starts at an *ex ante* stage where players are symmetrically uniformed. Then some exogenous external state $(\theta_0, x_1, x_2) \in \Theta_0 \times X$ is selected at random, each i observes (only) x_i and chooses some $a_i \in A_i$.

2.2 Belief hierarchies and type spaces

For any standard Borel space Z we write $\Delta(Z)$ for the set of all probability measures on Z, endowed with the topology of weak convergence and the corresponding Borel σ-algebra.[11] The space $\Delta(Z)$ is also standard Borel and its σ-algebra is the same as the one generated by the family of sets of the form $\{\mu \in \Delta(Z) \mid \mu[E] \geq p\}$, where $p \in [0,1]$ and $E \subseteq Z$ is measurable.[12] Given another standard Borel space Z', each measurable function $g : Z \to Z'$ induces the measurable function $\hat{g} : \Delta(Z) \to \Delta(Z')$ such that $\hat{g}(\mu)[E] = \mu[g^{-1}(E)]$ for each measurable $E \subseteq Z'$. For each $\mu \in \Delta(Z)$, the measure $\hat{g}(\mu)$ is the *pushforward* of μ given by g.

Our analysis concerns the players' interactive beliefs over a basic uncertainty space of the form $Y = \Theta_0 \times Y_1 \times Y_2$ where either $Y_i = \Theta_i$ for each $i \in I$, or $Y_i = X_i = \Theta_i \times \Xi_i$ for each $i \in I$, or $Y_i = X_i \times A_i = \Theta_i \times \Xi_i \times A_i$ for each $i \in I$. Note that in the three cases considered we have $Y = \Theta$, $Y = \Theta_0 \times X$, and $Y = \Theta_0 \times X \times A$, respectively. Given a set Y as above, for all $i \in I$ define $H^1_{Y,i} = \Delta(\Theta_0 \times Y_{-i})$ and recursively

$$H^{k+1}_{Y,i} = \left\{ (\delta^1_i, \ldots, \delta^{k+1}_i) \in H^k_{Y,i} \times \Delta(\Theta_0 \times Y_{-i} \times H^k_{Y,-i}) \;\middle|\; \mathrm{marg}_{\Theta_0 \times Y_{-i} \times H^{k-1}_{Y,i}} \delta^{k+1}_i = \delta^k_i \right\} \qquad \forall k \geq 1.$$

[9] The analysis can be easily extended to the case of compact Polish $\Theta_0 \times X$ and continuous $\boldsymbol{g}_i$. Finiteness of the action sets can be relaxed at the cost of some additional complications. We assume two players for simplicity. This allows us to focus our attention on issues of correlation that do not arise in games with complete information. Throughout the paper we consistently use bold symbols to denote functions that may be interpreted as random variables. An example of this sort is the payoff function $\boldsymbol{g}_i$, since $\boldsymbol{g}_i(\cdot, a)$ can be interpreted as the random payoff induced by action profile a, which is a function of the payoff state and hence of the state of the world.

[10] See Fagin, Halpern, Moses, and Vardi (1995, pp. 23–24).

[11] A measurable space is a standard Borel space if it is isomorphic to a separable and completely metrizable (i.e. Polish) topological space, endowed with the Borel σ-algebra (see e.g. Kechris, 1995, Definition 12.5).

[12] This fact, whose proof can be found e.g. in (Kechris, 1995, Theorems 17.23 and 17.24), is what motivates and renders meaningful our definition of expressible assumptions in section 3.1.

By the *coherency* conditions on marginal distributions, each element of $H^k_{Y,i}$ is determined by its last coordinate, so we can identify $H^k_{Y,i}$, the space of *Y-based k-order hierarchies* of player i, with $\Delta(\Theta_0 \times Y_{-i} \times H^{k-1}_{Y,-i})$, the space of *$Y$-based k-order beliefs* of player i. Accordingly, define the space of *Y-based belief hierarchies* of player i as

$$H^*_{Y,i} = \Big\{(\delta^k_i)_{k\geq 1} \in \bigtimes_{k\geq 1} \Delta(\Theta_0 \times Y_{-i} \times H^{k-1}_{Y,-i}) \;\Big|\; (\delta^1_i, \ldots, \delta^k_i) \in H^k_{Y,i} \quad \forall k \geq 1\Big\}.$$

From Mertens and Zamir (1985) we know that $H^*_{Y,i}$ is compact metrizable (hence standard Borel) in the product topology. Moreover, there exists a homeomorphism

$$\boldsymbol{\phi}_{Y,i} : H^*_{Y,i} \to \Delta(\Theta_0 \times Y_{-i} \times H^*_{Y,-i})$$

that is *belief-preserving*: $\mathsf{marg}_{\Theta_0\times Y_{-i}\times H^{k-1}_{Y,-i}}\, \boldsymbol{\phi}_{Y,i}(h^*_i) = \delta^k_i$ for all $h^*_i = (\delta^\ell_i)_{\ell\geq 1} \in H^*_{Y,i}$ and $k \geq 1$.

Appending a Y-based belief hierarchy $h^*_i \in H^*_{Y,i}$ to some primitive information $y_i \in Y_i$ we obtain a *Y-based explicit type* $t^*_i = (y_i, h^*_i)$ of player i. Thus, the space of Y-based explicit types of i is

$$T^*_{Y,i} = Y_i \times H^*_{Y,i}.$$

We can describe explicit Y-based types with a *Y-based type space* à la Harsanyi (1967-68), a structure

$$\mathsf{T} = \langle Y, (T_i, \boldsymbol{\pi}_i, \boldsymbol{y}_i)_{i\in I}\rangle$$

where each T_i is a standard Borel space and the functions $\boldsymbol{\pi}_i : T_i \to \Delta(\Theta_0 \times T_{-i})$ and $\boldsymbol{y}_i : T_i \to Y_i$ are measurable. Indeed, each type $t_i \in T_i$ induces a Y-based explicit type

$$\boldsymbol{\tau}^{\mathsf{T}}_i(t_i) = (\boldsymbol{y}_i(t_i), \boldsymbol{\pi}^{\mathsf{T},1}_i(t_i), \boldsymbol{\pi}^{\mathsf{T},2}_i(t_i), \ldots)$$

in a natural way: $\boldsymbol{\pi}^{\mathsf{T},1}_i(t_i)$ is the pushforward of $\boldsymbol{\pi}_i(t_i)$ given by $(\theta_0, t_{-i}) \mapsto (\theta_0, \boldsymbol{y}_{-i}(t_{-i}))$, and recursively for all $k \geq 2$, $\boldsymbol{\pi}^{\mathsf{T},k}_i(t_i)$ is the pushforward of $\boldsymbol{\pi}_i(t_i)$ given by

$$(\theta_0, t_{-i}) \mapsto (\theta_0, \boldsymbol{y}_{-i}(t_{-i}), \boldsymbol{\pi}^{\mathsf{T},1}_{-i}(t_{-i}), \ldots, \boldsymbol{\pi}^{\mathsf{T},k-1}_{-i}(t_{-i})).$$

A particular case that will play an important role in our analysis is that of a Θ-based type space with *information types*, namely a Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$ such that

$$T_i = X_i = \Theta_i \times \Xi_i \quad \text{and} \quad \boldsymbol{\theta}_i(t_i) = \mathsf{proj}_{\Theta_i}\, t_i \qquad \forall i \in I,\ \forall t_i \in T_i.$$

Thus, in a type space with information types, each player's beliefs are determined by his information, so that Harsanyi types can be interpreted as private information. Such type spaces are often used in applications, usually assuming a common prior.

The type space $\mathsf{T}^*_Y = \langle Y, (T^*_{Y,i}, \boldsymbol{\pi}^*_{Y,i}, \boldsymbol{y}^*_i)_{i\in I}\rangle$ where $\boldsymbol{y}^*_i : T^*_{Y,i} \to Y_i$ is the natural projection and $\boldsymbol{\pi}^*_{Y,i} : T^*_{Y,i} \to \Delta(\Theta_0 \times T^*_{Y,-i})$ is the mapping $(y_i, h^*_i) \mapsto \boldsymbol{\phi}_{Y,i}(h^*_i)$ for each $i \in I$, is the *canonical universal* Y-based type space. Indeed, for every Y-based type space T there are unique belief-preserving mappings from $(T_i)_{i\in I}$ into $(T^*_{Y,i})_{i\in I}$, namely the mappings $(\boldsymbol{\tau}^{\mathsf{T}}_i)_{i\in I}$ above.[13] When the

[13] The mappings $(\boldsymbol{\tau}^{\mathsf{T}}_i)_{i\in I}$ constitute the *canonical belief morphism* from T to T^*_Y. We introduce belief morphisms below.

mappings $(\boldsymbol{\tau}_i^{\mathsf{T}})_{i\in I}$ are injective the type space T is called *non-redundant*. In this case, $(\boldsymbol{\tau}_i^{\mathsf{T}})_{i\in I}$ are measurable embeddings onto their images $(\boldsymbol{\tau}_i^{\mathsf{T}}(T_i))_{i\in I}$, which are measurable and can be viewed as a non-redundant type space, since we have $\boldsymbol{\pi}_i^*(\boldsymbol{\tau}_i^{\mathsf{T}}(t_i))[\Theta_0 \times \boldsymbol{\tau}_{-i}^{\mathsf{T}}(T_{-i})] = 1$ for all $i \in I$ and $t_i \in T_i$. Conversely, any $(T_i)_{i\in I}$ such that $T_i \subseteq T_{Y,i}^*$ and $\boldsymbol{\pi}_{Y,i}^*(t_i)[\Theta_0 \times T_{-i}] = 1$ for all $i \in I$ and $t_i \in T_i$ can be viewed as a non-redundant type space.

Every Y-based type space $\mathsf{T} = \langle Y, (T_i, \boldsymbol{\pi}_i, \boldsymbol{y}_i)_{i\in I}\rangle$ induces a Θ-based type space, namely

$$\langle \Theta, (T_i, \boldsymbol{\pi}_i, \mathrm{proj}_{\Theta_i} \boldsymbol{y}_i(\cdot))_{i\in I}\rangle.$$

More generally, a *belief morphism* from a Y-based type space $\mathsf{T} = \langle Y, (T_i, \boldsymbol{\pi}_i, \boldsymbol{y}_i)_{i\in I}\rangle$ to a Θ-based type space $\mathsf{T}' = \langle \Theta, (T_i', \boldsymbol{\pi}_i', \boldsymbol{\theta}_i')_{i\in I}\rangle$ is a pair $(\boldsymbol{m}_i)_{i\in I}$ where for each $i \in I$ the mapping $\boldsymbol{m}_i : T_i \to T_i'$ is measurable and the following diagram commutes:

$$\begin{array}{ccc}
Y_i & \xrightarrow{\quad \mathrm{proj}_{\Theta_i} \quad} & \Theta_i \\
\boldsymbol{y}_i \uparrow & & \uparrow \boldsymbol{\theta}_i' \\
T_i & \xrightarrow{\quad \boldsymbol{m}_i \quad} & T_i' \\
\boldsymbol{\pi}_i \downarrow & & \downarrow \boldsymbol{\pi}_i' \\
\Delta(\Theta_0 \times T_{-i}) & \xrightarrow{\quad \widehat{\mathrm{id}_{\Theta_0}, \boldsymbol{m}_{-i}} \quad} & \Delta(\Theta_0 \times T_{-i}')
\end{array}$$

Note that if $Y = \Theta$ then this reduces to the usual definition of belief morphism between Θ-based type spaces, as in Mertens and Zamir (1985). In any case, the existence of a belief morphism $(\boldsymbol{m}_i)_{i\in I}$ from T to T' implies that every Θ-hierarchy that can be computed from a type in T (if $Y \neq \Theta$, via the induced Θ-based type space) is also generated by some type in T', and if each $\boldsymbol{m}_i$ is onto, then the converse is also true.

The belief morphism from the universal $(\Theta_0 \times X \times A)$-based type space $\mathsf{T}^*_{\Theta_0\times X\times A}$ onto the universal Θ-based type space T^*_Θ, which we denote by $(\boldsymbol{m}_i^*)_{i\in I}$, will be especially relevant for our purposes. The mappings $(\boldsymbol{m}_i^*)_{i\in I}$ are defined as follows: let $\boldsymbol{m}_i^1 : H^1_{\Theta_0\times X\times A,i} \to H^1_{\Theta,i}$ designate the pushforward mapping given by the projection $(\theta_0, \theta_{-i}, \xi_{-i}, a_{-i}) \mapsto (\theta_0, \theta_{-i})$, and recursively, let $\boldsymbol{m}_i^k : H^k_{\Theta_0\times X\times A,i} \to H^k_{\Theta,i}$ be the pushforward mapping given by

$$(\theta_0, \theta_{-i}, \xi_{-i}, a_{-i}, \delta^1_{-i}, \dots, \delta^{k-1}_{-i}) \mapsto (\theta_0, \theta_{-i}, \boldsymbol{m}^1_{-i}(\delta^1_{-i}), \dots, \boldsymbol{m}^{k-1}_{-i}(\delta^{k-1}_{-i})).$$ [14]

Then $\boldsymbol{m}_i^* : T^*_{\Theta\times X\times A,i} \to T^*_{\Theta,i}$ is defined as $(\theta_i, \xi_i, a_i, \delta_i^1, \delta_i^2, \dots) \mapsto (\theta_i, \boldsymbol{m}_i^1(\delta_i^1), \boldsymbol{m}_i^2(\delta_i^2), \dots)$.

3 Epistemic characterization of solution concepts

In this section we characterize solution concepts for asymmetric information games in terms of expressible assumptions. First we define rationality and common belief and we present the logical structure of the expressible assumptions we are going to consider (section 3.1). Then we define and

[14] Recall that we identify k-order belief hierarchies with k-order beliefs.

characterize Δ-rationalizability (section 3.2), interim correlated rationalizability (ICR, section 3.3), and interim (independent) rationalizability (IIR, section 3.4).

3.1 Expressible assumptions on rationality and beliefs

We define an *expressible assumption* as an event in (i.e. measurable subset of) the space of *states of the world*,

$$\Omega = \Theta_0 \times T^*_{\Theta_0\times X\times A,1} \times T^*_{\Theta_0\times X\times A,2}.$$

Interpreting events in Ω as expressible assumptions is justified by the equivalence between the topological construction and the measure-theoretic construction of σ-algebras mentioned in section 2.2. Indeed, an expressible assumption concerns terms that are either primitive (external states) or derived using primitive terms and other derived terms.

To make the latter claim more precise, consider the notion of expressibility introduced in Heifetz and Samet (1998): every subset of external states $S \subseteq \Theta_0 \times X \times A$ is an *expression*, and if e and f are expressions, then $\neg e$, $e \cap f$ and $B^p_i(e)$ are also expressions — for each $i \in I$ and $p \in [0,1]$ — which we read as "not e", "e and f" and "player i attaches probability at least p to e," respectively. Given any $(\Theta_0 \times X \times A)$-based type space $\mathsf{T} = \langle \Theta_0 \times X \times A, (T_i, \boldsymbol{\pi}_i, \boldsymbol{x}_i, \boldsymbol{a}_i)_{i\in I}\rangle$, every expression e can be viewed as a measurable subset $[e] \subseteq \Theta_0 \times T_1 \times T_2$: indeed, we can identify any subset S of external states with the set

$$[S] = \Big\{(\theta_0, t_1, t_2) \in \Theta_0 \times T_1 \times T_2 \;:\; (\theta_0, \boldsymbol{x}_1(t_1), \boldsymbol{a}_1(t_1), x_2(t_2), \boldsymbol{a}_2(t_2)) \in S\Big\},$$

and for any expressions d, e and f for which $[d]$, $[e]$ and $[f]$ are defined and $[f]$ has the form $[f] = F\times T_i$ for some $F \subseteq \Theta_0\times T_{-i}$, we can identify $\neg d$, $d\cap e$ and $B^p_i(f)$ with $[\neg d] = (\Theta_0\times T_1\times T_2)\setminus[d]$, $[d \cap e] = [d] \cap [e]$, and

$$[B^p_i(f)] = \Big\{(\theta_0, t_1, t_2) \in \Theta_0 \times T_1 \times T_2 \;:\; \boldsymbol{\pi}_i(t_i)[F] \geq p\Big\},$$

respectively. An event in T is *expressible* if it belongs to the σ-algebra generated by the expressions, when the latter are themselves viewed as events in T as explained above. It can be shown that expressibility of *every* event in T is equivalent to non-redundancy of T.[15] As we know, the latter is in turn equivalent to T being isomorphic to a belief-closed subset of the universal $(\Theta_0 \times X \times A)$-based type space $\mathsf{T}^*_{\Theta_0\times X\times A}$. It follows that $\mathsf{T}^*_{\Theta_0\times X\times A}$ is the *unique* (up to isomorphism) type space where all events can be seen as expressions and, conversely, every expression corresponding to some (nonempty) event in some type space, can be seen as a (nonempty) event in $\mathsf{T}^*_{\Theta_0\times X\times A}$.

The solution concepts we consider, although all related to each other in interesting ways as we shall soon see, belong to two different families. Δ-rationalizability is a type-space-free notion in that it takes as given only the environment $\mathcal{E}$ and, possibly, some information-dependent restrictions

[15] Recall from Mertens and Zamir (1985) that a $(\Theta_0 \times X \times A)$-based type space $\mathsf{T} = \langle \Theta_0 \times X \times A, (T_i, \boldsymbol{\pi}_i, \boldsymbol{x}_i, \boldsymbol{a}_i)_{i\in I}\rangle$ is non-redundant if and only if, for each player i, the smallest σ-algebra on T_i such that the mappings $\boldsymbol{\pi}_i$, $\boldsymbol{x}_i$, and $\boldsymbol{a}_i$ are all measurable separates every two distinct elements of T_i. This smallest σ-algebra is precisely the one generated by the expressions.

on first-order beliefs. As we shall see, in some cases these restrictions can be derived from those embodied in a type space. This is what links Δ-rationalizability with ICR and IIR, which are instead defined for the *Bayesian game*

$$\langle \Theta, (A_i, T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i, \boldsymbol{g}_i)_{i\in I}\rangle$$

induced by (the environment $\mathcal{E}$ and) some Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$. Indeed, the *solution set* delivered by Δ-rationalizability has the form $\boldsymbol{S}_1 \times \boldsymbol{S}_2 \subseteq (X_1 \times A_1) \times (X_2 \times A_2)$, whereas the solution set corresponding to ICR or IIR has the form $\boldsymbol{S}_1 \times \boldsymbol{S}_2 \subseteq (T_1 \times A_1) \times (T_2 \times A_2)$. In both cases, we would like to relate $\boldsymbol{S}_1 \times \boldsymbol{S}_2$ to an expressible event, so as to spell out the different (expressible) assumptions that the various notions rely on. We now define rationality and common belief, which play a prominent role in our results, and then we briefly sketch the general form of our expressible epistemic characterizations.

All the epistemic characterizations we provide below involve *rationality* of all players, which is the expressible assumption that each player chooses an action maximizing his expected payoff given his payoff-relevant information and first-order beliefs, i.e. $RAT = \Theta_0 \times RAT_1 \times RAT_2$, where

$$RAT_i = \left\{(\theta_i, \xi_i, a_i, \delta_i^1, \delta_i^2, \dots) \in T^*_{\Theta_0\times X\times A,i} \;\middle|\; a_i \in \arg\max_{a_i'} \boldsymbol{g}_i(\theta_i, a_i', \mathrm{marg}_{\Theta_0\times\Theta_{-i}\times A_{-i}}\, \delta_i^1)\right\}.$$[16]

Indeed, our characterizations involve not only rationality, but also *common belief* in this and possibly other assumptions. We define common belief in assumptions that, like RAT, take the form of a rectangular event $E = \Theta_0' \times E_1 \times E_2 \subseteq \Omega$, where $\Theta_0' \subseteq \Theta_0$ and each $E_i \subseteq T^*_{\Theta\times X\times A,i}$ is measurable. Given any such E, for every $i \in I$ define

$$B_i(E) = \left\{t_i \in T^*_{\Theta_0\times X\times A,i} \;\middle|\; \boldsymbol{\pi}^*_{\Theta_0\times X\times A,i}(t_i)[\Theta_0' \times E_{-i}] = 1\right\} \qquad \text{and} \qquad B(E) = \Theta_0 \times B_1(E) \times B_2(E).$$[17]

Now let $B^0(E) = E$ and recursively define $B^k(E) = B(B^{k-1}(E))$ for all $k \geq 1$. Then the assumptions of *(correct) k-order mutual belief* in E and *(correct) common belief* in E are, respectively,

$$MB^k(E) = \bigcap_{\ell=0}^{k} B^\ell(E) \qquad \text{and} \qquad CB(E) = \bigcap_{k\geq 0} B^k(E).$$

For each player i the projections of these events on $T^*_{\Theta_0\times X\times A,i}$ will be denoted $MB_i^k(E)$ and $CB_i(E)$, respectively. Note that $MB_i^0(E) = E_i$, $MB_i^k(E) = E_i \cap B_i(MB^{k-1}(E))$, and $CB_i(E) = \cap_{k\geq 0} MB_i^k(E)$.

The logical structure of our epistemic characterization of a solution set $\boldsymbol{S}_1 \times \boldsymbol{S}_2$ is as follows: we take an event $E \subseteq \Omega$, typically representing some basic assumption on primitives and first-order beliefs (such as RAT), and for each player i we relate $CB_i(E)$ to $\boldsymbol{S}_i$ using an appropriate, natural mapping. For example, in the case $\boldsymbol{S}_1 \times \boldsymbol{S}_2 \subseteq (X_1 \times A_1) \times (X_2 \times A_2)$ we characterize $\boldsymbol{S}_1 \times \boldsymbol{S}_2$ by showing that $\boldsymbol{S}_i = \mathrm{proj}_{X_i\times A_i} CB_i(E)$ for every player i, which means that $a_i \in \boldsymbol{S}_i(x_i)$ if and only if the pair (x_i, a_i) is consistent with the assumption that E is the case and there is common belief in E. As we shall see, in the case $\boldsymbol{S}_1 \times \boldsymbol{S}_2 \subseteq (T_1 \times A_1) \times (T_2 \times A_2)$ the expressible characterization does

[16] Following the similar slight abuses of notation often found in the game theory literature, here and in what follows $\boldsymbol{g}_i(\theta_i, a_i, \cdot)$ denotes also its linear extension to $\Delta(\Theta_0 \times \Theta_{-i} \times A_{-i})$.

[17] Note that $B(\cdot)$ maps rectangular events into rectangular events. For our purposes it is sufficient to define the *mutual belief operator* on this restricted class of events (see Battigalli and Siniscalchi, 2002).

not have this simple form unless the assumed type space has information types, so that $T_i = X_i$ for all i. This is because in general the types of a Bayesian game are not part of the primitives, and are not necessarily expressible starting from the primitives. Instead, we will then refer to the Θ-based explicit types (which are expressible) induced by the types of the Bayesian game.

3.2 Δ-rationalizability

The specification of a Θ-based type space is necessary to obtain a standard definition of equilibrium, but is not needed for Δ-rationalizability, a solution concept that is meant to capture strategic reasoning in the assumed economic environment with no reference to type spaces.[18] The solution set delivered by Δ-rationalizability has the form $\boldsymbol{R}^{\Delta} = \boldsymbol{R}_1^{\Delta} \times \boldsymbol{R}_2^{\Delta} \subseteq X \times A$ and it is parametrized by some assumed information-dependent restrictions Δ on first-order beliefs: formally, $\Delta = (\Delta_{x_i})_{i \in I, x_i \in X_i}$, where $\Delta_{x_i} \subseteq \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ is a closed set for all $i \in I$ and $x_i \in X_i$. Before presenting the formal definition and the epistemic characterization of Δ-rationalizability, it is useful to list the following two special cases, which later on will help us establish the connection with ICR and IIR, respectively:

- Exogenous beliefs derived from a type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i \in I} \rangle$ with information types: in this case $T_i = X_i$ for all $i \in I$, the restrictions take the form

$$\Delta_{x_i} = \{ \mu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i}) \mid \mathrm{marg}_{\Theta_0 \times X_{-i}} \mu_i = \boldsymbol{\pi}_i(x_i) \}$$

  for all $i \in I$ and $x_i \in X_i$, and we say that Δ is *derived from* T.

- A belief $\mu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ satisfies *information-based conditional independence* if

$$\mu_i[x_{-i}] > 0 \Rightarrow \mu_i[\theta_0, a_{-i} | x_{-i}] = \mu_i[\theta_0 | x_{-i}] \mu_i[a_{-i} | x_{-i}] \qquad \forall (\theta_0, x_{-i}, a_{-i}) \in \Theta_0 \times X_{-i} \times A_{-i}.$$

  Let $\Delta_{i,\mathrm{CI}} \subseteq \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ denote this set of first-order beliefs. We will consider the case where exogenous beliefs are derived from a Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i \in I} \rangle$ with information types, *and* information-based conditional independence holds:

$$\Delta_{x_i} = \{ \mu_i \in \Delta_{i,\mathrm{CI}} \mid \mathrm{marg}_{\Theta_0 \times X_{-i}} \mu_i = \boldsymbol{\pi}_i(x_i) \} \qquad \forall i \in I,\ x_i \in X_i.$$

  In this case we say that Δ is *CI-derived from* T.

The solution set $\boldsymbol{R}^{\Delta} = \boldsymbol{R}_1^{\Delta} \times \boldsymbol{R}_2^{\Delta} \subseteq (X_1 \times A_1) \times (X_2 \times A_2)$ is defined as follows: let $\boldsymbol{R}_i^{\Delta,0} = X_i \times A_i$ and, recursively for all $k \geq 0$, let

$$\boldsymbol{R}_i^{\Delta,k+1} = \left\{ (\theta_i, \xi_i, a_i) \in X_i \times A_i \;\middle|\; \begin{array}{ll} \exists\, \mu_i \in \Delta_{(\theta_i, \xi_i)} : & \\ (\Delta 1) & \mathrm{supp}\, \mu_i \subseteq \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k} \\ (\Delta 2) & a_i \in \arg\max_{a_i' \in A_i} \boldsymbol{g}_i(\theta_i, a_i', \mathrm{marg}_{\Theta_0 \times \Theta_{-i} \times A_{-i}} \mu_i) \end{array} \right\}.$$

Finally, let $\boldsymbol{R}_i^{\Delta} = \cap_{k \geq 0} \boldsymbol{R}_i^{\Delta,k}$ and $\boldsymbol{R}^{\Delta} = \boldsymbol{R}_1^{\Delta} \times \boldsymbol{R}_2^{\Delta}$. For every $i \in I$, $x_i \in X_i$, and $k \geq 0$ let

$$\boldsymbol{R}_i^{\Delta,k}(x_i) = \{ a_i \in A_i \mid (x_i, a_i) \in \boldsymbol{R}_i^{\Delta,k} \}$$

[18] See Battigalli (2003) and Battigalli and Siniscalchi (2003, 2007). Of course, there are other type-space-free solution concepts, like ex post equilibrium.

and $\boldsymbol{R}_i^{\Delta}(x_i) = \cap_{k\geq 0}\boldsymbol{R}_i^{\Delta,k}(x_i)$. Battigalli and Siniscalchi (2003) provide general conditions (satisfied by the special cases introduced above) under which $\boldsymbol{R}_i^{\Delta}(x_i)$ is nonempty for all $i \in I$ and $x_i \in X_i$. They show that $\mathrm{proj}_{\Theta_i\times A_i}\boldsymbol{R}_i^{\Delta}$ yields the set of pairs (θ_i, a_i) that are realizable in some Bayesian equilibrium model consistent with the restrictions Δ.

Remark 1. *In the case of no restrictions, that is, $\Delta_{x_i} = \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ for all $i \in I$ and $x_i \in X_i$, the payoff-irrelevant information ξ_i plays no role. In this case we can drop the superscript Δ, write $\boldsymbol{R}_i^k(\theta_i) = \boldsymbol{R}_i^k(\theta_i, \xi_i)$ for some arbitrary ξ_i, and redefine $\boldsymbol{R}_i^k$ as a subset of $\Theta_i \times A_i$.*[19] *Then $a_i \in \boldsymbol{R}_i^k(\theta_i)$ if and only if (θ_i, a_i) survives k rounds of the following elimination procedure: for every $k > 0$ the pair $(\theta_i, a_i) \in \boldsymbol{R}_i^{k-1}$ is deleted at round k, so that $a_i \notin \boldsymbol{R}_i^k(\theta_i)$, if there exists $\alpha_i \in \Delta(\boldsymbol{R}_i^{k-1}(\theta_i))$ such that $\boldsymbol{g}_i(\theta_0, \theta_i, \theta_{-i}, \alpha_i, a_{-i}) > \boldsymbol{g}_i(\theta_0, \theta_i, \theta_{-i}, a_i, a_{-i})$ for all $(\theta_0, \theta_{-i}, a_{-i}) \in \Theta_0 \times \boldsymbol{R}_{-i}^{k-1}$. This extends the classical iterated dominance characterization of rationalizability in complete information games due to Pearce (1984). See Battigalli (2003).*

Let $[\Delta] \subseteq \Omega$ denote the event that all players' first-order beliefs satisfy the restrictions, that is,

$$[\Delta] = \Theta_0 \times [\Delta_1] \times [\Delta_2], \quad \text{where} \quad [\Delta_i] = \left\{(\theta_i, \xi_i, a_i, \delta_i^1, \delta_i^2, \dots) \in T^*_{\Theta_0\times X\times A, i} \;\middle|\; \delta_i^1 \in \Delta_{x_i}\right\} \qquad \forall i \in I.$$

The following result, which is a special case of Proposition 1 in Battigalli and Siniscalchi (2007) and whose proof is therefore omitted, says that Δ-rationalizability is characterized by the expressible assumption that there is common belief in the players' rationality and in their first-order beliefs satisfying the restrictions.

Lemma 1. *For all $i \in I$ and $k \geq 1$,*

$$\boldsymbol{R}_i^{\Delta,k} = \mathrm{proj}_{X_i\times A_i} MB_i^{k-1}(RAT \cap [\Delta]) \qquad \textit{and} \qquad \boldsymbol{R}_i^{\Delta} = \mathrm{proj}_{X_i\times A_i} CB_i(RAT \cap [\Delta]).$$

3.3 Interim correlated rationalizability

Fix a Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$. Interim correlated rationalizability (ICR) yields a solution set $\boldsymbol{ICR}^{\mathsf{T}} \subseteq (T_1 \times A_1) \times (T_2 \times A_2)$ for the Bayesian game induced by T (see Dekel, Fudenberg, and Morris, 2007) defined recursively as follows:[20] $\boldsymbol{ICR}_i^{\mathsf{T},0} = T_i \times A_i$ and

$$\boldsymbol{ICR}_i^{\mathsf{T},k+1} = \left\{ (t_i, a_i) \in T_i \times A_i \;\middle|\; \begin{array}{ll} \exists\, \nu_i \in \Delta(\Theta_0 \times T_{-i} \times A_{-i}) : & \\ \text{(ICR1)} & a_i \in \arg\max_{a_i' \in A_i} \boldsymbol{g}_i(\boldsymbol{\theta}_i(t_i), a_i', \boldsymbol{\mu}_i(\nu_i)), \\ \text{(ICR2)} & \mathrm{supp}\, \nu_i \subseteq \Theta_0 \times \boldsymbol{ICR}_{-i}^{\mathsf{T},k} \\ \text{(ICR3)} & \mathrm{marg}_{\Theta_0\times T_{-i}}\, \nu_i = \boldsymbol{\pi}_i(t_i) \end{array} \right\},$$

where $\boldsymbol{\mu}_i(\nu_i) \in \Delta(\Theta_0 \times \Theta_{-i} \times A_{-i})$ is the belief induced by ν_i in the obvious way, i.e. the pushforward of ν_i given by the mapping $(\theta_0, t_{-i}, a_{-i}) \mapsto (\theta_0, \boldsymbol{\theta}_{-i}(t_{-i}), a_{-i})$. Finally, $\boldsymbol{ICR}_i^{\mathsf{T}} = \cap_{k\geq 0}\boldsymbol{ICR}_i^{\mathsf{T},k}$ and

[19] This is the procedure used by Bergemann and Morris (2009) to define *iterative implementation*, which is shown to be equivalent to robust (or type-space-independent) implementation.

[20] The sets Θ_i and Ξ_i are singletons (and hence do not appear at all) in Dekel, Fudenberg, and Morris (2007). However, their definitions and results extend seamlessly to the more general framework of this paper.

$\boldsymbol{ICR}^{\mathsf{T}} = \boldsymbol{ICR}_1^{\mathsf{T}} \times \boldsymbol{ICR}_2^{\mathsf{T}}$. Given any $i \in I$ and $t_i \in T_i$, we denote the sets of k-order ICR actions and ICR actions of type t_i as

$$\boldsymbol{ICR}_i^{\mathsf{T},k}(t_i) = \left\{ a_i \in A_i \,\middle|\, (t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T},k} \right\}$$

and $\boldsymbol{ICR}_i^{\mathsf{T}}(t_i) = \cap_{k \geq 0} \boldsymbol{ICR}_i^{\mathsf{T},k}(t_i)$. The qualification "correlated" in ICR is due to the possibility that, according to the justifying belief ν_i, θ_0 is correlated with a_{-i} even after conditioning on t_{-i}.

We report an alternative definition of ICR to facilitate comparison to other solution concepts. The intuition is that type t_i of player i forms a probabilistic conjecture $\boldsymbol{\sigma}_{-i} : \Theta_0 \times T_{-i} \to \Delta(A_{-i})$ of how the behavior of $-i$ depends on t_{-i} and θ_0 (possibly via some implicit correlation device). Given $\boldsymbol{\pi}_i(t_i)$, the conjecture σ_{-i} then induces the belief $\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})$ used to compute the expected payoff:

$$\boldsymbol{ICR}_i^{\mathsf{T},k+1} = \left\{ (t_i, a_i) \in T_i \times A_i \,\middle|\, \begin{array}{ll} \multicolumn{2}{l}{\exists \text{ a measurable } \boldsymbol{\sigma}_{-i} : \Theta_0 \times T_{-i} \to \Delta(A_{-i}) \text{ such that:}} \\ \text{(ICR1a)} & a_i \in \arg\max_{a_i' \in A_i} \boldsymbol{g}_i(\boldsymbol{\theta}_i(t_i), a_i', \boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})), \\ \text{(ICR2a)} & \qquad \forall\ (\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}, \\ & \operatorname{supp} \boldsymbol{\sigma}_{-i}(\theta_0, t_{-i}) \subseteq \boldsymbol{ICR}_{-i}^{\mathsf{T},k}(t_{-i}) \end{array} \right\},$$

where $\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i}) \in \Delta(\Theta_0 \times \Theta_{-i} \times A_{-i})$ is defined by

$$\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})[\theta_0, \theta_{-i}, a_{-i}] = \int_{(\boldsymbol{\theta}_{-i})^{-1}(\theta_{-i})} \boldsymbol{\sigma}_{-i}(\theta_0, t_{-i})[a_{-i}] \cdot \boldsymbol{\pi}_i(t_i)[\theta_0, \mathrm{d}t_{-i}].$$

It can be shown that the two definitions are equivalent — see Dekel, Fudenberg, and Morris (2007).

The following proposition relates ICR and Δ-rationalizability in the important special case of a type space T with information types. Note that this indirectly provides an expressible characterization of ICR for this special case, via Lemma 1.

Proposition 1. *Let T be a type space with information types and let Δ be derived from T. Then $\boldsymbol{ICR}^{\mathsf{T},k} = \boldsymbol{R}^{\Delta,k}$ for every $k \geq 0$ and hence $\boldsymbol{ICR}^{\mathsf{T}} = \boldsymbol{R}^{\Delta}$.*

Proof. By our definitions, $\boldsymbol{ICR}_i^{\mathsf{T},0} = \boldsymbol{R}_i^{\Delta,0} = X_i \times A_i$ for all $i \in I$. Now suppose by way of induction that, for some $k \geq 0$, we have $\boldsymbol{ICR}_i^{\mathsf{T},k} = \boldsymbol{R}_i^{\Delta,k}$ for all $i \in I$. Pick any $i \in I$, $x_i \in X_i$, $a_i \in A_i$, and $\nu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i})$. By the inductive hypothesis, $\operatorname{supp} \nu_i \subseteq \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k}$ is equivalent to $\operatorname{supp} \nu_i \subseteq \Theta_0 \times \boldsymbol{ICR}_{-i}^{\mathsf{T},k}$. Moreover, $\nu_i \in \Delta_{x_i}$ is equivalent to $\operatorname{marg}_{\Theta_0 \times X_{-i}} \nu_i = \boldsymbol{\pi}_i(x_i)$, because Δ is derived from T. It follows that the conditions for $a_i \in \boldsymbol{ICR}_i^{\mathsf{T},k+1}(x_i)$ are equivalent to those for $a_i \in \boldsymbol{R}_i^{\Delta,k+1}(x_i)$. Since this is true for all $i \in I$, $x_i \in X_i$, and $a_i \in A_i$, the induction step follows and the proof is complete. □

Combined with a result in Dekel, Fudenberg, and Morris (2007), Proposition 1 generalizes as follows.

Corollary 1. *Let T and T' be Θ-based type spaces. Assume that T has information types and let Δ be derived from T. If there is a belief morphism $(\boldsymbol{m}_i)_{i \in I}$ from T' to T, then*

$$\boldsymbol{ICR}_i^{\mathsf{T}',k}(t_i') = \boldsymbol{R}_i^{\Delta,k}(\boldsymbol{m}_i(t_i')) \qquad \forall i \in I,\ \forall k \geq 1.$$

Proof. Dekel, Fudenberg, and Morris (2007, Corollary 2) prove that the set of ICR actions of a type only depends on the Θ-based beliefs generated by it. If $(\boldsymbol{m}_i)_{i \in I}$ is a belief morphism, $\boldsymbol{\tau}_i^{\mathsf{T}'} = \boldsymbol{\tau}_i^{\mathsf{T}} \circ \boldsymbol{m}_i$ for all $i \in I$, hence $\boldsymbol{ICR}_i^{\mathsf{T}'}(\cdot) = \boldsymbol{ICR}_i^{\mathsf{T}}(\boldsymbol{m}_i(\cdot))$ for all $i \in I$. Combining this fact with Proposition 1 gives the result. □

3.3.1 Expressible epistemic characterization of ICR

Lemma 1, Proposition 1 and Corollary 1 entail an expressible epistemic characterization of ICR for a class of Θ-based type spaces that encompasses many economic applications — the class of type spaces with information types. Here we provide an expressible epistemic characterization that holds for all Θ-based type spaces.

Following the notation introduced in section 2.2, let $\boldsymbol{a}_i^* : T^*_{\Theta\times X\times A,i} \to A_i$ denote the natural projection. The following theorem says that, given any type space T, the set $\boldsymbol{ICR}_i^{\mathsf{T}}$ comprises all and only those pairs $(t_i, a_i) \in T_i \times A_i$ such that a_i is consistent with rationality and common belief in rationality, given that the Θ-based explicit type of i is the one induced by t_i via $\boldsymbol{\tau}_i^{\mathsf{T}}$. (Dekel, Fudenberg, and Morris, 2007 prove a related result.)

Theorem 1. *Fix a Θ-based type space* $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$. *For all* $i \in I$ *and* $k \geq 1$,

$$\boldsymbol{ICR}_i^{\mathsf{T},k} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists\, t_i^* \in MB_i^{k-1}(RAT) \text{ s.t. } \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \text{ and } \boldsymbol{a}_i^*(t_i^*) = a_i\right\}.$$

Furthermore,

$$\boldsymbol{ICR}_i^{\mathsf{T}} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists\, t_i^* \in CB_i(RAT) \text{ s.t. } \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \text{ and } \boldsymbol{a}_i^*(t_i^*) = a_i\right\},$$

or equivalently,

$$\boldsymbol{ICR}_i^{\mathsf{T}}(t_i) = \mathrm{proj}_{A_i}\left(CB(RAT) \cap \{(\theta_0, t_1^*, t_2^*) \in \Omega \mid \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i)\}\right).$$

Proof. See Appendix A. □

3.4 Interim independent rationalizability

Fix a Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$. Interim independent rationalizability (IIR) yields a solution set $\boldsymbol{IIR}^{\mathsf{T}} \subseteq \boldsymbol{ICR}^{\mathsf{T}}$ for the Bayesian game induced by T (Dekel, Fudenberg, and Morris, 2007). Given $i \in I$ and $t_i \in T_i$, a measurable function $\boldsymbol{\sigma}_{-i} : T_{-i} \to \Delta(A_{-i})$ induces a measure $\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i}) \in \Delta(\Theta_0 \times \Theta_{-i} \times A_{-i})$ as follows:

$$\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})[\theta_0, \theta_{-i}, a_{-i}] = \int_{(\boldsymbol{\theta}_{-i})^{-1}(\theta_{-i})} \boldsymbol{\sigma}_{-i}(t_{-i})[a_{-i}] \cdot \boldsymbol{\pi}_i(t_i)[\theta_0, \mathrm{d}t_{-i}]. \tag{1}$$

Note that we can write $\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})[\theta_0, a_{-i}|t_{-i}] = \boldsymbol{\sigma}_{-i}(t_{-i})[a_{-i}] \cdot \boldsymbol{\pi}_i(t_i)[\theta_0|t_{-i}]$ for any t_{-i} in the support of $\mathsf{marg}_{T_{-i}}\, \pi_i(t_i)$. (This is obvious when T_{-i} is finite.) Thus, $\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})$ satisfies the conditional independence property that θ_0 and a_{-i} are independent conditional on t_{-i}. The recursive definition of interim independent rationalizability is as follows: $\boldsymbol{IIR}_{i,0}^{\mathsf{T}} = T_i \times A_i$,

$$\boldsymbol{IIR}_i^{\mathsf{T},k+1} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \begin{array}{ll} \multicolumn{2}{l}{\exists \text{ measurable } \boldsymbol{\sigma}_{-i} : T_{-i} \to \Delta(A_{-i}) \;:} \\ \text{(IIR1)} & a_i \in \arg\max_{a_i' \in A_i} \boldsymbol{g}_i(\boldsymbol{\theta}_i(t_i), a_i', \boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})) \\ \text{(IIR2)} & \forall t_{-i} \in T_{-i},\ \mathrm{supp}\, \boldsymbol{\sigma}_{-i}(t_{-i}) \subseteq \boldsymbol{IIR}_{-i}^{\mathsf{T},k}(t_{-i}) \end{array}\right\}.$$

Finally, $\boldsymbol{IIR}^{\mathsf{T},k} = \boldsymbol{IIR}_1^{\mathsf{T},k} \times \boldsymbol{IIR}_2^{\mathsf{T},k}$ and $\boldsymbol{IIR}^{\mathsf{T}} = \cap_{k\geq 0} \boldsymbol{IIR}^{\mathsf{T}}$.

Remark 2. *For each i and t_i in T, $a_i \in \boldsymbol{IIR}_i^{\mathsf{T}}(t_i)$ if and only if a_i is rationalizable for t_i in the interim strategic form of the Bayesian game induced by T (see Appendix B).*

The alternative definition of ICR makes it clear that, as anticipated, $\boldsymbol{IIR}^{\mathsf{T}} \subseteq \boldsymbol{ICR}^{\mathsf{T}}$. By the argument in the proof of Corollary 1, it follows that all IIR outcomes obtaining across all Θ-based type spaces T' that belief-morphically map into T, are ICR outcomes of T. The following result shows that the converse also holds. Note that this equivalence suggests the following robustness interpretation: the predictions of ICR are all and only the predictions that the analyst could make using IIR, but without committing to the assumption that the assumed type space is the "true" one. (This may be because, for instance, he suspects he is disregarding some payoff-irrelevant variable observed by the players.)

Proposition 2. *For every Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$ there exists a Θ-based type space $\mathsf{T}' = \langle \Theta, (T_i', \boldsymbol{\pi}_i', \boldsymbol{\theta}_i')_{i\in I}\rangle$ and a belief morphism $(\boldsymbol{m}_i)_{i\in I}$ from T' onto T such that*

$$\boldsymbol{ICR}_i^{\mathsf{T}}(t_i) = \bigcup_{t_i' \in (\boldsymbol{m}_i)^{-1}(t_i)} \boldsymbol{ICR}_i^{\mathsf{T}'}(t_i') = \bigcup_{t_i' \in (\boldsymbol{m}_i)^{-1}(t_i)} \boldsymbol{IIR}_i^{\mathsf{T}'}(t_i') \qquad \forall i \in I,\ \forall t_i \in T_i. \tag{2}$$

Proof. For every $i \in I$ and $(t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T}}$, let $V_i(t_i, a_i)$ designate the set of all $\nu_i \in \Delta(\Theta_0 \times \boldsymbol{ICR}_{-i}^{\mathsf{T}})$ such that a_i is a best response to ν_i for t_i. Define $\mathsf{T}' = \langle \Theta, (T_i', \boldsymbol{\pi}_i', \boldsymbol{\theta}_i')_{i\in I}\rangle$ as follows: for all $i \in I$, $T_i' = \boldsymbol{ICR}_i^{\mathsf{T}}$, $\boldsymbol{\theta}_i'(t_i, a_i) = \boldsymbol{\theta}_i(t_i)$ for all $(t_i, a_i) \in T_i'$, and $\boldsymbol{\pi}_i' : T_i' \to \Delta(\Theta_0 \times T_{-i}')$ is an arbitrary measurable selector from the correspondence V_i.[21]

Let $\boldsymbol{m}_i : T_i' \to T_i$ be the natural projection for each $i \in I$, and let us verify that $(\boldsymbol{m}_i)_{i\in I}$ is a belief morphism from T' onto T. Indeed, for each $i \in I$ and $(t_i, a_i) \in T_i'$ we have $\boldsymbol{\theta}_i(\boldsymbol{m}_i(t_i, a_i)) = \boldsymbol{\theta}_i(t_i) = \boldsymbol{\theta}_i'(t_i, a_i)$, and $\mathsf{marg}_{\Theta_0 \times T_{-i}} \boldsymbol{\pi}_i'(t_i, a_i)$ is the pushforward of $\boldsymbol{\pi}_i(t_i)$ given by $(\theta_0, t_i, a_i) \mapsto (\theta_0, t_i)$. This proves that $(\boldsymbol{m}_i)_{i\in I}$ is a belief morphism, which is onto because $\mathsf{proj}_{T_i} \boldsymbol{ICR}_i^{\mathsf{T}} = T_i$.

We will prove (2) now, thus completing the proof of the theorem. Since $(\boldsymbol{m}_i)_{i\in I}$ is a morphism, $\boldsymbol{\tau}_i^{\mathsf{T}'} = \boldsymbol{\tau}_i^{\mathsf{T}} \circ \boldsymbol{m}_i$ and hence $\boldsymbol{ICR}_i^{\mathsf{T}'}(\cdot) = \boldsymbol{ICR}_i^{\mathsf{T}}(\boldsymbol{m}_i(\cdot))$ (Dekel, Fudenberg, and Morris, 2007, Corollary 2). Therefore it suffices to show that $a_i \in \boldsymbol{IIR}_i^{\mathsf{T}'}(t_i, a_i)$ for all $i \in I$ and $(t_i, a_i) \in T_i'$. For each $i \in I$ define W_i as the set of triplets of the form (t_i, a_i, a_i), where $(t_i, a_i) \in T_i'$. By construction, for every $i \in I$ and $(t_i, a_i) \in T_i'$ action a_i is a best response to $\boldsymbol{\pi}_i'(t_i, a_i)$ for type $t_i \in T_i$, therefore a_i is also a best response for type $(t_i, a_i) \in T_i'$ to the pushforward of $\boldsymbol{\pi}_i'(t_i, a_i)$ given the mapping $(t_{-i}, a_{-i}) \mapsto (t_{-i}, a_{-i}, a_{-i})$. Thus $(W_i)_{i\in I}$ has the independent best-response property as in the fixed-point definition of IIR (see Ely and Pęski, 2006) and hence we obtain $a_i \in \boldsymbol{IIR}_i^{\mathsf{T}'}(t_i, a_i)$ for all $i \in I$ and $(t_i, a_i) \in T_i'$. □

Although ICR is weaker than IIR, the two notions coincide in the important special case where ($\mathcal{E}$ is such that) there is distributed knowledge of the payoff state:

Remark 3. *Suppose that there is distributed knowledge of the payoff state, that is, assume that Θ_0 is a singleton. Then $\boldsymbol{IIR}^{\mathsf{T}} = \boldsymbol{ICR}^{\mathsf{T}}$.*

[21] To verify that T' is a well defined Θ-based type space, note that T_i' is standard Borel because $\boldsymbol{ICR}_i^{\mathsf{T}} \subseteq T_i \times A_i$ is closed and both T_i and A_i are standard Borel. Moreover, $\boldsymbol{\theta}_i'$ is clearly measurable. Finally, $\boldsymbol{\pi}_i'$ exists and is measurable, because V_i is a nonempty-valued, closed-graph correspondence between compact spaces. (This is by the Kuratowski-Ryll-Nardzewski selection theorem — see e.g. Aliprantis and Border, 1999.)

Proof. If Θ_0 is a singleton we can suppress it in the notation and write conjectures of player i in the alternative definition of ICR and in the definition of IIR in the same way, i.e. as measurable mappings from T_{-i} to $\Delta(A_{-i})$. Thus, in this case the two definitions coincide. □

Remark 3 implies that an expressible characterization of IIR is possible under distributed knowledge of the payoff state. Though almost trivial, the remark is important because many economic applications feature distributed knowledge of θ. Models with private values are an obvious example, but also many models with interdependent values satisfy this property.[22]

We can provide an expressible epistemic characterization of IIR in another important special case, namely when T has information types. We do this using a preliminary result that relates IIR to Δ-rationalizability, when Δ is derived from T under the conditional independence assumption. This provides an indirect epistemic characterization via Lemma 1. Note the parallel between Proposition 1 and the next result: the former establishes equivalence between Δ-rationalizability and ICR when we require Δ to be derived from the assumed type space; the latter establishes the analogous equivalence with IIR when, in addition, we require Δ to be CI-derived from the type space.

Proposition 3. *Let* T *be a type space with information types (so that* $T_i = X_i$*,* $i = 1, 2$*) and let* Δ *be CI-derived from* T*. Then* $\boldsymbol{IIR}^{\mathsf{T},k} = \boldsymbol{R}^{\Delta,k}$ *for all* $k \geq 0$ *and, therefore,* $\boldsymbol{IIR}^{\mathsf{T}} = \boldsymbol{R}^{\Delta}$*.*

Proof. By our definitions, $\boldsymbol{IIR}_i^{\mathsf{T},0} = \boldsymbol{R}_i^{\Delta,0} = X_i \times A_i$ for all $i \in I$. Now suppose by way of induction that, for some $k \geq 0$, we have $\boldsymbol{IIR}_i^{\mathsf{T},k} = \boldsymbol{R}_i^{\Delta,k}$ for all $i \in I$. Pick any $i \in I$, any $x_i \in X_i$, any $a_i \in A_i$, and any $\nu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i})$. By the inductive hypothesis, $\operatorname{supp} \nu_i \subseteq \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k}$ is equivalent to $\operatorname{supp} \nu_i \subseteq \Theta_0 \times \boldsymbol{IIR}_{-i}^{\mathsf{T},k}$. Moreover, $\nu_i \in \Delta_{x_i}$ is equivalent to $\operatorname{marg}_{\Theta_0 \times X_{-i}} \nu_i = \boldsymbol{\pi}_i(x_i)$ and $\nu_i[\theta_0, a_{-i} | x_{-i}] = \nu_i[\theta_0 | x_{-i}] \nu_i[a_{-i} | x_{-i}]$ for each x_{-i} with $\nu_i[x_{-i}] > 0$, as Δ is CI-derived from T.

Suppose that $a_i \in \boldsymbol{R}_i^{\Delta,k+1}(x_i)$ because a_i is a best reply for type x_i to a belief $\nu_i \in \Delta_{x_i}$ with $\operatorname{supp} \nu_i \subseteq \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k}$. Define $\boldsymbol{\sigma}_{-i} : X_{-i} \to \Delta(A_{-i})$ as follows: for all $x_{-i} \in X_{-i}$ and $a_{-i} \in A_{-i}$,

$$\boldsymbol{\sigma}_{-i}(x_{-i})[a_{-i}] = \begin{cases} \nu_i[a_{-i} | x_{-i}] & \text{if } \nu_i[x_{-i}] > 0, \\ 1/\left|\boldsymbol{IIR}_{-i}^{\mathsf{T},k}(x_{-i})\right| & \text{if } \nu_i[x_{-i}] = 0 \text{ and } a_{-i} \in \boldsymbol{IIR}_{-i}^{\mathsf{T},k}(x_{-i}), \\ 0 & \text{if } \nu_i[x_{-i}] = 0 \text{ and } a_{-i} \notin \boldsymbol{IIR}_{-i}^{\mathsf{T},k}(x_{-i}). \end{cases}$$

Note that $\operatorname{marg}_{\Theta_0 \times \Theta_{-i} \times A_{-i}} \nu_i = \boldsymbol{\mu}_i(x_i, \boldsymbol{\sigma}_{-i})$ and, by the inductive hypothesis, for all $x_{-i} \in X_{-i}$ we have $\operatorname{supp} \boldsymbol{\sigma}_{-i}(x_{-i}) \subseteq \boldsymbol{IIR}_{-i}^{\mathsf{T},k}(x_{-i})$. It follows that the conditions for $a_i \in \boldsymbol{IIR}_i^{\mathsf{T},k+1}(x_i)$ are satisfied. Next suppose that $a_i \in \boldsymbol{IIR}_i^{\mathsf{T},k+1}(x_i)$ and hence that a_i is a best response for x_i to a conjecture $\boldsymbol{\sigma}_{-i}$ with $\operatorname{supp} \boldsymbol{\sigma}_{-i}(x_{-i}) \subseteq \boldsymbol{IIR}_{-i}^{\mathsf{T},k}(x_{-i})$ for all $x_{-i} \in X_{-i}$. Let $\nu_i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ de defined as follows: for all $\theta_0 \in \Theta_0$, $x_{-i} \in X_{-i}$, and $a_{-i} \in A_{-i}$, $\nu_i[\theta_0, x_{-i}, a_{-i}] = \boldsymbol{\pi}_i(x_i)[\theta_0, x_{-i}] \boldsymbol{\sigma}_{-i}(x_{-i})[a_{-i}]$. Since a_i is a best response to $\boldsymbol{\mu}_i(x_i, \boldsymbol{\sigma}_{-i})$, it is also a best response to ν_i. Moreover, $\nu_i \in \Delta_{x_i}$ and, by the inductive hypothesis, $\operatorname{supp} \nu_i \subseteq \Theta_0 \times \boldsymbol{IIR}_{-i}^{\mathsf{T},k} = \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k}$. Thus, $a_i \in \boldsymbol{R}_i^{\Delta,k+1}(x_i)$. Since this is true for all $i \in I$, $x_i \in X_i$, and $a_i \in A_i$, the induction step follows and the proof is complete. □

[22] For example, consider "wallet games" (Klemperer, 1998), or any model where θ_i specifies player i's characteristics such as ability or riskiness, and the consequences for each player of an action profile depend on all players' characteristics.

3.4.1 Expressible epistemic characterization of IIR

Since types à la Harsanyi need not be expressible, the conditional independence assumption underlying the general definition of IIR need not be expressible, either. Thus, we are unable to provide a general characterization of IIR via expressible assumptions. But expressible characterizations can be given in interesting special cases.

Our first result is an immediate consequence of Theorem 1 and Remark 3. If there is distributed knowledge of the payoff state — a property of the fixed environment $\mathcal{E}$ — then for any Θ-based type space T, $\boldsymbol{IIR}_i^{\mathsf{T}}$ is the set of all pairs $(t_i, a_i) \in T_i \times A_i$ such that a_i is consistent with rationality and common belief in rationality, given that the Θ-based explicit type of i is the one induced by t_i.

Corollary 2. *Suppose that there is distributed knowledge of the payoff state, that is, assume that Θ_0 is a singleton. Fix a Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$. For all $i \in I$ and $k \geq 1$,*

$$\boldsymbol{IIR}_i^{\mathsf{T},k} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists t_i^* \in MB_i^{k-1}(RAT) \quad s.t. \quad \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}^{\mathsf{T}}(t_i) \quad and \quad \boldsymbol{a}_i^*(t_i^*) = a_i\right\}.$$

Furthermore,

$$\boldsymbol{IIR}_i^{\mathsf{T}} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists t_i^* \in CB_i(RAT) \quad s.t. \quad \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}^{\mathsf{T}}(t_i) \quad and \quad \boldsymbol{a}_i^*(t_i^*) = a_i\right\},$$

or equivalently,

$$\boldsymbol{IIR}_i^{\mathsf{T}}(t_i) \;=\; \mathrm{proj}_{A_i}\left(CB(RAT) \cap \{(\theta_0, t_1^*, t_2^*) \in \Omega \mid \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i)\}\right).$$

To state the next result, consider the following expressible assumption of *information-based conditional independence*,

$$ICI = \left\{(\theta_0, (x_i, a_i, h_i^*)_{i\in I}) \in \Omega \;\middle|\; \mathrm{marg}_{\Theta_0\times X_{-i}\times A_{-i}} \boldsymbol{\phi}_{\Theta_0\times X\times A,i}(h_i^*) \in \Delta_{i,\mathrm{CI}} \;\; \forall i \in I\right\}.$$

(Recall that $\Delta_{i,\mathrm{CI}}$ is the set of probability measures in $\Delta(\Theta_0 \times X_{-i} \times A_{-i})$ such that θ_0 and a_{-i} are independent conditional on x_{-i}.) The result follows at once from Lemma 1 and Proposition 3. If T is a Θ-based type space with information types, then $\boldsymbol{IIR}_i^{\mathsf{T}}$ is the set of all pairs $(t_i, a_i) \in T_i \times A_i$ such that a_i is consistent with rationality, information-based conditional independence, and common belief in these two assumptions, given that the private information and the Θ-hierarchy of i are the ones induced by t_i.

Corollary 3. *Fix a Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$ with information types. For all $k \geq 1$ and $i \in I$,*

$$\boldsymbol{IIR}_i^{\mathsf{T},k} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists t_i^* \in MB_i^{k-1}(RAT \cap ICI) \quad s.t. \quad \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}^{\mathsf{T}}(t_i) \quad and \quad \boldsymbol{a}_i^*(t_i^*) = a_i\right\}.$$

Furthermore,

$$\boldsymbol{IIR}_i^{\mathsf{T}} = \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists t_i^* \in CB_i(RAT \cap ICI) \quad s.t. \quad \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}^{\mathsf{T}}(t_i) \quad and \quad \boldsymbol{a}_i^*(t_i^*) = a_i\right\},$$

or equivalently,

$$\boldsymbol{IIR}_i^{\mathsf{T}}(t_i) \;=\; \mathrm{proj}_{A_i}\left(CB(RAT \cap ICI) \cap \{(\theta_0, t_1^*, t_2^*) \in \Omega \mid \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i)\}\right).$$

We conjecture that, under non-redundancy, an analogous result holds for arbitrary Θ-based type spaces, once we require independence conditional on the Θ-based explicit type rather than ICI.

4 Ex ante and interim rationalizability

In this section we show that the differences between rationalizability in the ex ante and interim strategic form of a Bayesian game are due to the different independence restrictions implicit in these solution concepts. This follows from a preliminary result about Δ-rationalizability that helps clarifying the conceptual issue. We consider a set Δ of information-dependent restrictions on beliefs, define a notion of ex ante correlated Δ-rationalizability, and show that it is in a strong sense equivalent to the interim notion of Δ-rationalizability introduced earlier.

We take the point of view of player i in an ex ante stage at which he has not yet received his information x_i. Let $\boldsymbol{F}_i$ be the set of all functions from X_i to A_i, for every $i \in I$. Then we can define a "structural" ex ante strategic form with two real players, 1 and 2, choosing strategies in $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$, and a fictitious player choosing an exogenous external state in $\Theta_0 \times X$.

Let $\overline{\boldsymbol{\theta}}_i : X_i \to \Theta_i$ for each player i denote the natural projection. Then the payoff function $\overline{\boldsymbol{g}}_i : \Theta_0 \times X \times \boldsymbol{F}_1 \times \boldsymbol{F}_2 \to \mathbb{R}$ of player i is defined by the formula

$$\overline{\boldsymbol{g}}_i(\theta_0, x_1, x_2, \boldsymbol{f}_1, \boldsymbol{f}_2) = \boldsymbol{g}_i(\theta_0, \overline{\boldsymbol{\theta}}_1(x_1), \overline{\boldsymbol{\theta}}_2(x_2), \boldsymbol{f}_1(x_1), \boldsymbol{f}_2(x_2))$$

Player i forms an ex ante belief $\mu_i \in \Delta(\Theta_0 \times X \times \boldsymbol{F}_{-i})$ about the choice of the fictitious player and the strategy of $-i$. Again slightly abusing our notation, we write $\overline{\boldsymbol{g}}_i(\boldsymbol{f}_i, \mu_i)$ for the corresponding ex ante expected payoff when i chooses strategy $\boldsymbol{f}_i$.

Now fix some restrictions on interim beliefs $\Delta = (\Delta_1, \Delta_2)$, where $\Delta_i = (\Delta_{x_i})_{x_i \in X_i}$ and $\Delta_{x_i} \subseteq \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ for each $i \in I$ and $x_i \in X_i$. Clearly, Δ_i implies restrictions on the ex ante beliefs of i. Thus we say that $\mu_i \in \Delta(\Theta_0 \times X \times \boldsymbol{F}_{-i})$ is *consistent with* Δ_i if it assigns positive probability to each x_i and if it yields interim beliefs consistent with Δ_i, that is, if for all $x_i \in X_i$ we have:

- $\mu_i[x_i] := \sum_{\theta_0, x_{-i}, \boldsymbol{f}_{-i}} \mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}] > 0$;[23]
- $(\mu_i[\theta_0, x_{-i}, a_{-i}|x_i])_{\theta_0 \in \Theta_0, x_{-i} \in X_{-i}, a_{-i} \in A_{-i}} \in \Delta_{x_i}$, where

$$\mu_i[\theta_0, x_{-i}, a_{-i}|x_i] := \sum_{\boldsymbol{f}_{-i} : \boldsymbol{f}_{-i}(x_{-i}) = a_{-i}} \frac{\mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}]}{\mu_i[x_i]}.$$

Note that we allow i's ex ante beliefs to exhibit *correlation* between the fictitious player and the real opponent. The set of *ex Ante Correlated Δ-Rationalizable strategies* is recursively defined as follows: $\boldsymbol{ACR}_i^{\Delta,0} = \boldsymbol{F}_i$ and

$$\boldsymbol{ACR}_i^{\Delta,k+1} = \left\{ \boldsymbol{f}_i \in \boldsymbol{F}_i \;\middle|\; \begin{array}{ll} \exists\, \mu_i \in \Delta(\Theta_0 \times X \times \boldsymbol{F}_{-i}) \ : & \\ (\mathrm{A}\Delta 1) & \boldsymbol{f}_i \in \arg\max_{\boldsymbol{f}'_i \in \boldsymbol{F}_i} \overline{\boldsymbol{g}}_i(\boldsymbol{f}'_i, \mu_i) \\ (\mathrm{A}\Delta 2) & \operatorname{supp} \mu_i \subseteq \Theta_0 \times X \times \boldsymbol{ACR}_{-i}^{\Delta,k} \\ (\mathrm{A}\Delta 3) & \mu_i \text{ is consistent with } \Delta_i \end{array} \right\};$$

[23] We impose this weak requirement to derive well-defined interim beliefs and avoid tedious issues concerning the differences between ex ante and interim expected payoff maximization. Alternatively, we could impose a perfection requirement (see Brandenburger and Dekel, 1987). This discussion would distract the reader's attention from the important issues.

$$\boldsymbol{ACR}_i^\Delta = \bigcap_{k\geq 0}\boldsymbol{ACR}_i^{\Delta,k}; \qquad \boldsymbol{ACR}^\Delta = \boldsymbol{ACR}_1^\Delta \times \boldsymbol{ACR}_2^\Delta.$$

Adapting the argument used by Battigalli and Siniscalchi (2007) to prove their Proposition 1, one can show that $\boldsymbol{ACR}^\Delta$ is the set of strategy profiles of the ex ante structural strategic form that are consistent with (correct) common belief of rationality and the restrictions Δ.

We now relate $\boldsymbol{ACR}^\Delta$ to (interim) Δ-rationalizability. Recall that interim notions of rationalizability yield solution sets made of type-action pairs. A set $\boldsymbol{S}_i \subseteq X_i \times A_i$ whose projection on X_i is X_i itself, is equivalent to a nonempty-valued correspondence $x_i \mapsto \boldsymbol{S}_i(x_i) \subseteq A_i$, and we can look at the selections from this correspondence. Then, given a set $\boldsymbol{F}_i' \subseteq \boldsymbol{F}_i$, it makes sense to write $\boldsymbol{F}_i' \approx \boldsymbol{S}_i$ whenever $\boldsymbol{F}_i'$ is precisely the set of such selections:

$$\boldsymbol{F}_i' \approx \boldsymbol{S}_i \quad \text{if and only if} \quad \boldsymbol{F}_i' = \left\{\boldsymbol{f}_i \in \boldsymbol{F}_i \;\middle|\; \forall x_i \in X_i,\ \boldsymbol{f}_i(x_i) \in \boldsymbol{S}_i(x_i)\right\}.$$

The following result shows that ex ante Δ-rationalizability is fully equivalent to Δ-rationalizability.

Proposition 4. *For all $k \geq 0$, $\boldsymbol{ACR}^{\Delta,k} \approx \boldsymbol{R}^{\Delta,k}$. Furthermore, $\boldsymbol{ACR}^\Delta \approx \boldsymbol{R}^\Delta$.*

Proof. We prove by induction that $\boldsymbol{ACR}_i^{\Delta,k} \approx \boldsymbol{R}_i^{\Delta,k}$ for each $i = 1,2$ and $k \geq 0$. This is trivially true for $k = 0$, so suppose by way of induction that $\boldsymbol{ACR}_i^{\Delta,k} \approx \boldsymbol{R}_i^{\Delta,k}$ for each $i = 1,2$. We shall prove that $\boldsymbol{ACR}_i^{\Delta,k+1} \approx \boldsymbol{R}_i^{\Delta,k+1}$ for each $i = 1,2$.

Let $\boldsymbol{f}_i \in \boldsymbol{ACR}_i^{\Delta,k+1}$. We must show that $\boldsymbol{f}_i$ is a selection from $\boldsymbol{R}_i^{\Delta,k+1}(\cdot)$. Let μ_i be a belief that justifies $\boldsymbol{f}_i$ as in the definition of ex ante Δ-rationalizability. By condition (AΔ3), for each $x_i \in X_i$ we can derive interim beliefs $\nu_{x_i} \in \Delta_{x_i}$ by letting $\nu_{x_i}[\cdot] = \mu_i[\cdot|x_i]$. The inductive hypothesis and condition (AΔ2) then imply $\operatorname{supp}\nu_{x_i} \subseteq \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k}$. The ex ante maximization condition is

$$\boldsymbol{f}_i \in \operatorname*{arg\,max}_{\boldsymbol{f}_i' \in \boldsymbol{F}_i} \overline{\boldsymbol{g}}_i(\boldsymbol{f}_i', \mu_i) = \operatorname*{arg\,max}_{\boldsymbol{f}_i' \in \boldsymbol{F}_i} \sum_{x_i \in X_i} \mu_i[x_i] \sum_{\theta_0, x_{-i}, \boldsymbol{f}_{-i}} \frac{\mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}]}{\mu_i[x_i]}\, \overline{\boldsymbol{g}}_i(\theta_0, x_i, x_{-i}, \boldsymbol{f}_i', \boldsymbol{f}_{-i}),$$

where the latter expression is well defined since μ_i is consistent with Δ_i. Thus, $\boldsymbol{f}_i$ must maximize each x_i-term in the summation. The inductive hypothesis and conditions (AΔ2) and (AΔ3) imply that for all $a_i \in A_i$ and $x_i \in X_i$ we have

$$\sum_{\theta_0, x_{-i}, \boldsymbol{f}_{-i}} \frac{\mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}]}{\mu_i[x_i]}\, \boldsymbol{g}_i(\theta_0, \overline{\boldsymbol{\theta}}_i(x_i), \overline{\boldsymbol{\theta}}_{-i}(x_{-i}), a_i, \boldsymbol{f}_{-i}(x_{-i})) =$$

$$= \sum_{\theta_0, x_{-i}, a_{-i}} \sum_{\boldsymbol{f}_{-i} : \boldsymbol{f}_{-i}(x_{-i}) = a_{-i}} \frac{\mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}]}{\mu_i[x_i]}\, \boldsymbol{g}_i(\theta_0, \overline{\boldsymbol{\theta}}_i(x_i), \overline{\boldsymbol{\theta}}_{-i}(x_{-i}), a_i, a_{-i}) =$$

$$= \sum_{\theta_0, x_{-i}, a_{-i}} \mu_i[\theta_0, x_{-i}, a_{-i}|x_i]\, \boldsymbol{g}_i(\theta_0, \overline{\boldsymbol{\theta}}_i(x_i), \overline{\boldsymbol{\theta}}_{-i}(x_{-i}), a_i, a_{-i}) = \boldsymbol{g}_i(\overline{\boldsymbol{\theta}}_i(x_i), a_i, \nu_{x_i}),$$

so we conclude that $\boldsymbol{f}_i(x_i) \in \operatorname{arg\,max}_{a_i} \boldsymbol{g}_i(\overline{\boldsymbol{\theta}}_i(x_i), a_i, \nu_{x_i})$ and hence that ν_{x_i} satisfies all the conditions in the definition of $\boldsymbol{R}_i^{\Delta,k+1}(x_i)$. This proves that $\boldsymbol{f}_i$ is a selection from $\boldsymbol{R}_i^{\Delta,k+1}(\cdot)$.

Conversely, let $\boldsymbol{f}_i$ be a selection from $\boldsymbol{R}_i^{\Delta,k+1}(\cdot)$. We must show that $\boldsymbol{f}_i \in \boldsymbol{ACR}_i^{\Delta,k+1}$. By definition of $\boldsymbol{R}_i^{\Delta,k+1}$ and by the inductive assumption for each $x_i \in X_i$ there exists $\nu_{x_i} \in \Delta_{x_i}$ such that $\operatorname{supp}\nu_{x_i} \subseteq \Theta_0 \times \boldsymbol{R}_{-i}^{\Delta,k}$ and $\boldsymbol{f}_i(x_i) \in \operatorname{arg\,max}_{a_i} \boldsymbol{g}_i(\overline{\boldsymbol{\theta}}_i(x_i), a_i, \nu_{x_i})$. We construct a measure ν_i on $\Theta_0 \times X_i \times X_{-i} \times A_{-i}$ and we derive an appropriate measure $\mu_i \in \Delta(\Theta_0 \times X \times \boldsymbol{F}_{-i})$. Let

$\lambda_i \in \Delta(X_i)$ be an arbitrary full-support probability measure on X_i, and for each $(\theta_0, x_i, x_{-i}, a_{-i})$ let $\nu_i[\theta_0, x_i, x_{-i}, a_{-i}] = \lambda_i[x_i]\nu_{x_i}[\theta_0, x_{-i}, a_{-i}]$. Then

$$\sum_{\theta_0, x_i, x_{-i}, a_{-i}} \nu_i[\theta_0, x_i, x_{-i}, a_{-i}] = \sum_{x_i} \lambda_i[x_i] \sum_{\theta_0, x_{-i}, a_{-i}} \nu_{x_i}[\theta_0, x_{-i}, a_{-i}] = 1$$

and, moreover, $\nu_i \in \Delta(\Theta_0 \times X_i \times X_{-i} \times A_{-i})$. Clearly $\nu_{x_i}[\theta_0, x_{-i}, a_{-i}] = \nu_i[\theta_0, x_{-i}, a_{-i}|x_i]$. Define marginal conditional probabilities in the usual way when possible and arbitrarily when the conditioning event has zero probability. Now define $\mu_i \in \Delta(\Theta_0 \times X \times \boldsymbol{F}_{-i})$ by

$$\mu_i\left[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}\right] = \nu_i[\theta_0, x_i, x_{-i}] \prod_{x'_{-i}} \nu_i[\boldsymbol{f}_{-i}(x'_{-i})|\theta_0, x_i, x'_{-i}]. \tag{3}$$

To verify that this is indeed a well defined probability distribution, assume without loss of generality that $X_{-i} = \{x^1_{-i}, \ldots, x^n_{-i}\}$ and note that a strategy $\boldsymbol{f}_{-i}$ can be equivalently represented as an n-tuple $(a^1_{-i}, \ldots, a^n_{-i})$, so that

$$\begin{aligned}
\sum_{\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}} \mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}] &= \sum_{\theta_0, x_i, x_{-i}} \nu_i[\theta_0, x_i, x_{-i}] \sum_{\boldsymbol{f}_{-i}} \prod_{x'_{-i}} \nu_i[\boldsymbol{f}_{-i}(x'_{-i})|\theta_0, x_i, x'_{-i}] \\
&= \sum_{\theta_0, x_i, x_{-i}} \nu_i[\theta_0, x_i, x_{-i}] \sum_{a^1_{-i}, \ldots, a^n_{-i}} \prod_{k=1}^{n} \nu_i[a^k_{-i}|\theta_0, x_i, x^k_{-i}] \\
&= \sum_{\theta_0, x_i, x_{-i}} \nu_i[\theta_0, x_i, x_{-i}] \sum_{a^1_{-i}} \nu_i[a^1_{-i}|\theta_0, x_i, x^1_{-i}] \cdots \sum_{a^n_{-i}} \nu_i[a^n_{-i}|\theta_0, x_i, x^n_{-i}] = 1.
\end{aligned}$$

By construction, μ_i is an ex ante belief consistent with Δ_i. Suppose that $\boldsymbol{f}_{-i} \notin \boldsymbol{ACR}^{\Delta,k}_{-i}$. By the inductive assumption, $\boldsymbol{f}_{-i}$ is not a selection from $\boldsymbol{R}^{\Delta,k}_{-i}(\cdot)$, so $(x_{-i}, \boldsymbol{f}_{-i}(x_{-i})) \notin \boldsymbol{R}^{\Delta,k}_{-i}$ for some x_{-i} and hence for each θ_0 and x_i we have

$$\lambda_i[x_i]\nu_{x_i}[\theta_0, x_{-i}, \boldsymbol{f}_{-i}(x_{-i})] = \nu_i[\theta_0, x_i, x_{-i}]\nu_i[\boldsymbol{f}_{-i}(x_{-i})|\theta_0, x_i, x_{-i}] = 0,$$

which by (3) implies $\mu_i[\theta_0, x_i, x_{-i}, \boldsymbol{f}_{-i}] = 0$. This shows that $\operatorname{supp} \mu_i \subseteq \Theta_0 \times X \times \boldsymbol{ACR}^{\Delta,k}_{-i}$. Furthermore, note that for every $\boldsymbol{f}'_i \in \boldsymbol{F}_i$ we have

$$\overline{\boldsymbol{g}}_i(\boldsymbol{f}'_i, \mu_i) = \sum_{x_i} \lambda_i[x_i]\, \boldsymbol{g}_i(\overline{\boldsymbol{\theta}}_i(x_i), \boldsymbol{f}'_i(x_i), \nu_{x_i}).$$

As $\boldsymbol{f}_i(x_i) \in \arg\max_{a_i} \lambda_i[x_i]\boldsymbol{g}_i(\overline{\boldsymbol{\theta}}_i(x_i), a_i, \nu_{x_i})$ for all $x_i \in X_i$, we have $\boldsymbol{f}_i \in \arg\max_{\boldsymbol{f}'_i \in \boldsymbol{F}_i} \overline{\boldsymbol{g}}_i(\boldsymbol{f}'_i, \mu_i)$. Thus μ_i satisfies the conditions in the definition of ACR, hence $\boldsymbol{f}_i \in \boldsymbol{ACR}^{\Delta,k+1}_i$. □

Propositions 1 and 4 yield an equivalence result for ex ante and interim correlated rationalizability in Bayesian games with information types. Before stating the result formally, let us first review the standard notion of ex ante rationalizability. Since ex ante rationalizability makes sense only when Harsanyi types represent information that can be learned, we restrict our attention to Bayesian games with information types. However, we remark that an equivalence result like the one stated below can be proved for *every* Bayesian game.

A strategy for the Bayesian game induced by a type space T with information types is ex ante rationalizable if it is rationalizable in the ex ante strategic form of the game. To define the ex ante strategic form, we must first specify ex ante beliefs on $\Theta_0 \times X$ consistent with the type space T. Say that a prior $\Pi_i \in \Delta(\Theta_0 \times X)$ is *consistent with* T if for each x_i we have:

- $\Pi_i[x_i] := \sum_{\theta_0, x_{-i}} \Pi_i[\theta_0, x_i, x_{-i}] > 0$;[24]

- $\Pi_i[\theta_0, x_{-i}|x_i] := \Pi_i[\theta_0, x_i, x_{-i}]/\Pi_i[x_i] = \boldsymbol{\pi}_i(x_i)[\theta_0, x_{-i}] \quad \forall \theta_0 \in \Theta_0, \ \forall x_{-i} \in X_{-i}$.

Once we fix priors $\Pi = (\Pi_1, \Pi_2)$ consistent with T, the *ex ante strategic form* of the Bayesian game induced by T is given by the expected payoff functions

$$\boldsymbol{g}_i^{\Pi}(\boldsymbol{f}_1, \boldsymbol{f}_2) = \sum_{\theta_0 \in \Theta_0} \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} \Pi_i[\theta_0, x_1, x_2] \, \boldsymbol{g}_i(\theta_0, \overline{\boldsymbol{\theta}}_1(x_1), \overline{\boldsymbol{\theta}}_2(x_2), \boldsymbol{f}_1(x_1), \boldsymbol{f}_2(x_2)).$$

It can be verified that a strategy is rationalizable in the strategic game $(\boldsymbol{g}_1^{\Pi}, \boldsymbol{g}_2^{\Pi})$ if and only if it is rationalizable in every other strategic game $(\boldsymbol{g}_1^{\Pi'}, \boldsymbol{g}_2^{\Pi'})$ where prior beliefs Π' are also consistent with T. It is also standard to show that ex ante rationalizability implicitly relies on an *ex ante independence* assumption: a player's beliefs about (θ_0, x) and $\boldsymbol{f}_{-i}$ are given by a product measure $\Pi_i \times \mu_i$ where $\mu_i \in \Delta(\boldsymbol{F}_{-i})$. Ex ante independence implies interim independence; therefore ex ante rationalizability implies interim independent rationalizability, which is equivalent to rationalizability in the interim strategic form of the Bayesian game (Remark 2).

We now define a notion of ex ante correlated rationalizability that removes the ex ante independence assumption. Fix arbitrarily a profile of priors Π consistent with T. The set $\boldsymbol{ACR}^{\mathsf{T}}$ of *ex ante correlated rationalizable* strategy profiles is given by the following recursive definition: $\boldsymbol{ACR}_i^{\mathsf{T},0} = \boldsymbol{F}_i$,

$$\boldsymbol{ACR}_i^{\mathsf{T},k+1} = \left\{ \boldsymbol{f}_i \in \boldsymbol{F}_i \;\middle|\; \begin{array}{ll} \exists \mu_i \in \Delta(\Theta_0 \times X \times \boldsymbol{F}_{-i}): & \\ \text{(ACR1)} & \boldsymbol{f}_i \in \arg\max_{\boldsymbol{f}'_i \in \boldsymbol{F}_i} \overline{\boldsymbol{g}}_i(\boldsymbol{f}'_i, \mu_i), \\ \text{(ACR2)} & \operatorname{supp} \mu_i \subseteq \Theta_0 \times X \times \boldsymbol{ACR}_{-i}^{\mathsf{T},k} \\ \text{(ACR3)} & \operatorname{marg}_{\Theta_0 \times X} \mu_i = \Pi_i \end{array} \right\},$$

Finally, $\boldsymbol{ACR}_i^{\mathsf{T}} = \cap_{k>0} \boldsymbol{ACR}_i^{\mathsf{T},k}$, $\boldsymbol{ACR}^{\mathsf{T}} = \boldsymbol{ACR}_1^{\mathsf{T}} \times \boldsymbol{ACR}_2^{\mathsf{T}}$. It can be shown that the definition of $\boldsymbol{ACR}^{\mathsf{T}}$ is independent of the priors Π that we choose, as long as they are consistent with T.

Remark 4. *If T has information types and Δ is derived from T, then $\boldsymbol{ACR}^{\mathsf{T}} = \boldsymbol{ACR}^{\Delta}$.*

Theorem 2. *Ex ante correlated rationalizability is equivalent to interim correlated rationalizability: for every type space T with information types, $\boldsymbol{ACR}^{\mathsf{T}} \approx \boldsymbol{ICR}^{\mathsf{T}}$.*

Proof. By the remark above, $\boldsymbol{ACR}^{\mathsf{T}} = \boldsymbol{ACR}^{\Delta}$. By Proposition 4, $\boldsymbol{ACR}^{\Delta} \approx \boldsymbol{R}^{\Delta}$. By Proposition 1, $\boldsymbol{R}^{\Delta} = \boldsymbol{ICR}^{\mathsf{T}}$. Therefore $\boldsymbol{ACR}^{\mathsf{T}} \approx \boldsymbol{ICR}^{\mathsf{T}}$. □

Thus, looking deeper into the discrepancy between ex ante and interim rationalizability, we see that it is due to the different conditional independence restrictions, not to different types being allowed or not to hold different conjectures. Indeed, once these restrictions are removed, the discrepancy disappears: ex ante correlated rationalizability treats different types just as different information sets of the same player, and yet it is fully equivalent to ICR.

[24] As before, we include this mild requirement to avoid distracting the reader.

5 Discussion

5.1 A summary of our approach and results

In this paper we provide a unified framework to elucidate and relate to each other different notions of rationalizability for two-person, static games with asymmetric information. Our guiding principle is that a solution concept should allow an *expressible characterization*, that is, it should describe the implications for players' behavior of expressible assumptions about rationality and interactive beliefs. Based on Heifetz and Samet's (1998) work, we argue that these assumptions can be identified by events (measurable subsets) in the canonical space of external states and infinite hierarchies of beliefs about them. The external states are the primitive terms of the language used to express assumptions about rationality and interactive beliefs, and form what we call the economic environment. Infinite hierarchies of beliefs are obtained as derived elements. Our starting point is the observation that rationalizability for complete information games is characterized by rationality and common belief in rationality, therefore notions of rationalizability for games with incomplete/asymmetric information should be obtained by appropriate modifications of these basic expressible assumptions.

We use as a "glue solution concept" Δ-rationalizability, a procedure that iteratively deletes private information-action pairs, defined on the economic environment without reference to Harsanyi types, and parametrized by restrictions Δ on first-order beliefs. More standard notions of rationalizability, in particular IIR and ICR, are defined for Bayesian games, which obtain from the economic environment by appending to it a type space à la Harsanyi. When Harsanyi types can be interpreted as private information (both payoff relevant and payoff irrelevant), then we obtain ICR and IIR as special cases of Δ-rationalizability (Propositions 1,3). Since the latter admits an expressible epistemic characterization (Lemma 1), we obtain as corollaries expressible characterizations of ICR and IIR (Corollary 3) for this special case, which is very common in economic applications.

We cannot provide a general expressible characterization of IIR, as this seems to rely on a notion of conditional independence that refers to non-expressible features of Harsanyi types: each type of each player believes that, conditional on the type of the opponent, there is no residual correlation between his action and the payoff state. On the other hand, ICR drops conditional independence and only relies on expressible features of Harsanyi types (private payoff-relevant information and Θ-hierarchy), hence it admits an expressible characterization (Theorem 1). We point out that IIR and ICR coincide when the environment features distributed knowledge of the payoff state (no residual uncertainty about θ), a common situation in economics (Remark 3). This yields, by Theorem 1, an expressible characterization of IIR for this interesting special case (Corollary 2).

Besides characterizing and exploring the relationships between interim notions of rationalizability, relevant when asymmetric information is interpreted as genuine incomplete information, we analyze ex ante notions of rationalizability, relevant when asymmetric information concerns an actual initial chance move. We show that allowing for correlated conjectures about chance and the opponent, ex ante and interim Δ-rationalizability are equivalent (Proposition 4). As for Bayesian games, this implies that when Harsanyi types can be interpreted as private information, and hence the ex ante interpretation of the game makes sense, a correlated version of ex ante rationalizability is equivalent to ICR (Theorem 2).

To conclude, independence assumptions are responsible not only for the differences between interim independent rationalizability (i.e. rationalizability on the interim strategic form, see Remark 2) and interim correlated rationalizability, but also for the differences between ex ante and interim rationalizability. Removing the independence assumptions we obtain equivalent ex ante and interim solution concepts that allow an expressible characterization: the (correlated) rationalizable actions of a type t_i are the actions consistent with rationality, common belief in rationality, and the expressible features of t_i.

5.2 Extensions

***n* players.** The most natural extension of IIR to static games with more than two players assumes that each type of each player believes that, conditional on the opponents' types, the payoff state and all the opponents' actions are mutually independent, whereas the natural extension of ICR allows for general correlation. All our equivalence and expressible characterization results have straightforward generalizations, except for those based on Remark 3. Indeed, for this natural extension of IIR, our remark about the equivalence between IIR and ICR under distributed knowledge of the payoff state does not hold, for the same reasons why independent rationalizability is a refinement of correlated rationalizability in games with complete information.

Dynamic games. Δ-rationalizability in dynamic games with incomplete information has been studied by Battigalli (2003) and Battigalli and Siniscalchi (2003, 2007). These papers discuss also how to model independence assumptions in dynamic games. They study two versions of the solution concept, one that features a forward induction principle in the spirit of Pearce (1984) and Battigalli (1997), and a weaker one that does not. Battigalli and Siniscalchi (2007) provide expressible characterizations of both versions, thus extending Lemma 1. Proposition 4 on ex ante and interim Δ-rationalizability can also be extended. Similarly, one can define versions of ICR and IIR for dynamic Bayesian games with and without forward induction. (Penta, 2009 deals with the analogue of ICR without forward induction, defining analogues for the other notions is straightforward.) For these solution concepts, we can provide appropriate extensions of Propositions 1, 3 and Theorem 2; we conjecture that we can prove extensions of Theorem 1 and Corollary 3 as well.

5.3 Related literature

We already mentioned the relationship with the work of Battigalli (2003) and Battigalli and Siniscalchi (2003, 2007) on Δ-rationalizability. Here we just notice that none of these papers makes the difference between payoff relevant and payoff irrelevant information explicit; actually, their notation and language suggest that only payoff relevant information is considered, although this is not a formal assumption. Furthermore, Battigalli (2003) and Battigalli and Siniscalchi (2003) assume distributed knowledge of the payoff state, although their results do not depend on this assumption.

ICR has been introduced by Dekel, Fudenberg, and Morris (2007), who also provide some epistemic characterization results. Proposition 1 and much of our discussion rely on their important result that the ICR actions of a type only depend on its expressible features (in their paper, its Θ-based hierarchy of beliefs). This allows to restrict attention to ICR actions in the Θ-based universal type space, as Dekel, Fudenberg, and Morris (2006), Weinstein and Yildiz (2007), Chen, Di Tillio, Faingold, and Xiong (2009), and Penta (2009) do in their analysis of the continuity of rationalizable actions with respect to beliefs hierarchies. (Penta, 2009 considers an extensive form version of ICR.)

The most important differences between the approach of Dekel, Fudenberg, and Morris (2007) and ours is that they neglect private information (like Ely and Pęski, 2006) and do not state their epistemic results as expressible characterizations, i.e. by means of events in the appropriate canonical universal type space. These differences are related. One advantage of modeling private information (including the payoff irrelevant one) explicitly, is that this provides a sufficiently rich language with which we can express a property of (information based) conditional independence and a related characterization of IIR. We find the analogous characterization of Dekel, Fudenberg, and Morris (2007) less instructive because it relies on an *interpretation* of the type space as an "objective" information system that cannot be expressed in a formal language. Other advantages of our richer framework are that we can relate IIR and ICR to Δ-rationalizability, and that we can formally state the obvious but important point that ICR and IIR are equivalent in two-person environments with distributed knowledge of the payoff state.

Ely and Pęski (2006) analyze IIR. Like Dekel, Fudenberg, and Morris (2007), their starting point is the observation that IIR is not invariant to the addition/deletion of redundant types, and therefore depends on something more than the Θ-based hierarchies of beliefs of the Harsanyi types. Their approach to IIR is essentially orthogonal to ours. We look for conditions under which IIR actions admit an expressible characterization, whereas they change the notion of belief hierarchy in order to obtain one that identifies IIR actions. They show that, under some regularity conditions, Harsanyi types yield — beside the standard Θ-hierarchies — also richer Δ-*hierarchies* where i's first-order beliefs are elements of $\Delta(\Delta(\Theta_0 \times \Theta_{-i}))$.[25] Then they show that Δ-hierarchies identify IIR actions. It is not clear to us whether Δ-hierarchies are expressible in a meaningful sense. To elaborate further, take any Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i \in I} \rangle$. As Ely and Pęski (2006) point out, letting $\boldsymbol{\pi}_i(t_i|\cdot) : T_{-i} \to \Delta(\Theta_0 \times \{\boldsymbol{\theta}_{-i}(\cdot)\})$ for each $i \in I$ and $t_i \in T_i$ be a version of the conditional probability given $-i$'s type, we obtain Δ-hierarchies: in particular, the first-order belief in the Δ-hierarchy corresponding to type t_i of player i is defined by

$$\boldsymbol{\pi}_i^{\Delta,1}(t_i)[E] = \boldsymbol{\pi}_i(t_i)[\Theta_0 \times \{t_{-i} \in T_{-i} : \boldsymbol{\pi}_i(t_i|t_{-i}) \in E\}] \quad \text{for every measurable } E \subseteq \Delta(\Theta_0 \times \Theta_{-i}).$$

Now, if T has information types, so that $T_{-i} = X_{-i}$, then one can express this first-order belief as uncertainty about the relevant probability measure in the array

$$(\boldsymbol{\pi}_i(t_i|x_{-i}))_{x_{-i} \in X_{-i}} \in [\Delta(\Theta)]^{X_{-i}},$$

which would make Δ-hierarchies expressible in some sense. But if T does not have information

[25] Ely and Pęski (2006) have no private information — in our framework, this would correspond to the case where X_i is a singleton for each player i. We translate their definitions into our framework in the obvious way.

types, then we are not allowed to identify T_{-i} and X_{-i}, and this interpretation cannot be offered.

Sadzik (2007) seems to take a similar route to Ely and Pęski (2006): he defines hierarchical beliefs that identify Bayesian equilibrium actions. But on closer inspection, we find his approach much more similar to ours. He enriches the environment by adding to the payoff state θ a countable sequence of payoff-irrelevant (and continuous) *signals* for each player. On this expanded space of exogenous primitive uncertainty, call it Z, he constructs a formal language and relates it to standard Z-based hierarchies, showing that they identify Bayesian equilibrium actions. We speculatively propose the following interpretation of the difference between our approach to modeling uncertainty and his: we assume that there is common awareness only of a finite number of signals and consequently put only those signals in the commonly known environment.[26] This justifies conditionally correlated beliefs: when i conditions on the information type x_{-i} of $-i$, he suspects that $-i$ may observe some other payoff irrelevant variable i is not aware of, which in turn may be correlated with θ_0, thus allowing correlation between θ_0 and a_{-i} conditional on x_{-i} — this is a restatement of the incomplete model interpretation of conditional correlation given by Dekel, Fudenberg, and Morris (2007). On the other hand, Sadzik (2007) puts in the environment all the "conceivable" signals, which is justified if there is common awareness of all of them.

Liu (2009) analyzes Bayesian equilibrium predictions and the role of redundant types using an approach similar to ours. In particular, he distinguishes between redundant and non-redundant Θ-based type spaces, arguing that redundant types should be used only to represent hidden uncertainty entertained by players that the modeler does not explicitly take into account. Coherently with this approach, he suggests the modeler should always use a non-redundant type space unless he is aware there may be some additional strategically relevant information he is unaware of.[27] In our framework, the additional uncertainty is represented by the set of payoff irrelevant states Ξ and the exogenous beliefs of players are modeled using $(\Theta \times \Xi)$-based type spaces. In addition, Liu (2009) also shows that the same Bayesian equilibrium predictions can be obtained both with a Θ-based redundant type space and with an appropriate $(\Theta \times \Xi)$-based non-redundant type space. Instead of addressing Bayesian Equilibrium predictions, we use this richer uncertainty space, to highlight the connections among different definitions of rationalizability and to investigate the role of expressible independence restrictions.

A Proof of Theorem 1

Part I

Here we prove that for all $i \in I$ and $k \geq 1$,

$$\boldsymbol{ICR}_i^{\mathsf{T},k} \supseteq \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists\, t_i^* \in MB_i^{k-1}(RAT) \text{ s.t. } \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \text{ and } \boldsymbol{a}_i^*(t_i^*) = a_i\right\},$$

[26]Of course, a player may observe payoff-irrelevant aspects of which the opponent is unaware. In this case our rationalizability analysis should (and does) neglect these aspects.

[27]He also provides a necessary and sufficient condition on the space Θ (called "separativity") to identify a Θ-based redundant type space with a $(\Theta \times \Xi)$-based non-redundant type space through a mapping that preserves Θ-hierarchies. Given the finiteness assumption, this condition is satisfied in our framework.

which clearly implies

$$\boldsymbol{ICR}_i^{\mathsf{T}} \supseteq \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists\, t_i^* \in CB_i(RAT) \text{ s.t. } \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \text{ and } \boldsymbol{a}_i^*(t_i^*) = a_i\right\}.$$

The proof is by induction in k. Fix $i \in I$, $t_i \in T_i$ and $t_i^* = (\theta_i, \xi_i, a_i, \delta_i^1, \delta_i^2, \ldots) \in MB_i^0(RAT) = RAT_i$ such that $\boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i)$. Let $\sigma_{-i}^0 : \Theta_0 \times \Theta_{-i} \to \Delta(A_{-i})$ be any conditional distribution associated to $\operatorname{marg}_{\Theta_0 \times \Theta_{-i} \times A_{-i}} \delta_i^1$.[28] Define $\sigma_{-i} : \Theta_0 \times T_{-i} \to \Delta(A_{-i})$ so that $\sigma_{-i}(\theta_0, t_{-i}) = \sigma_{-i}^0(\theta_0, \boldsymbol{\theta}_{-i}(t_{-i}))$ for all $(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}$. Then

$$\begin{aligned} a_i &\in \operatorname*{arg\,max}_{a_i'} \sum_{\theta_0, \theta_{-i}, \xi_{-i}, a_{-i}} \boldsymbol{g}_i(\theta_0, \theta_i, \theta_{-i}, a_i', a_{-i}) \delta_i^1[\theta_0, \theta_{-i}, \xi_{-i}, a_{-i}] \\ &= \operatorname*{arg\,max}_{a_i'} \sum_{\theta_0, \theta_{-i}} \boldsymbol{\pi}_i^{\mathsf{T},1}(t_i)[\theta_0, \theta_{-i}] \sum_{a_{-i}} \boldsymbol{g}_i(\theta_0, \theta_i, \theta_{-i}, a_i', a_{-i}) \sigma_{-i}^0(\theta_0, \theta_{-i})[a_{-i}] \\ &= \operatorname*{arg\,max}_{a_i'} \int_{\Theta_0 \times T_{-i}} \sum_{a_{-i}} \boldsymbol{g}_i(\theta_0, \theta_i, \theta_{-i}, a_i', a_{-i}) \sigma_{-i}(\theta_0, t_{-i})[a_{-i}] \boldsymbol{\pi}_i(t_i)[d\theta_0 \times dt_{-i}], \end{aligned}$$

where the first line follows from $t_i^* \in MB_i^0(RAT)$ and the second from $\boldsymbol{\tau}_i^{\mathsf{T}}(t_i) = \boldsymbol{m}_i^*(t_i^*)$. This proves that $a_i \in \boldsymbol{ICR}_i^{\mathsf{T},1}(t_i)$.

Now let $k \geq 2$ and assume the claim holds true for $k-1$, that is, assume that $(t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T},k-1}$ for all $i \in I$ and for all $t_i \in T_i$ and $a_i \in A_i$ such that $\boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i)$ and $\boldsymbol{a}_i^*(t_i^*) = a_i$ for some $t_i^* \in MB_i^{k-2}(RAT)$. Fix $i \in I$, $t_i \in T_i$ and $t_i^* = (\theta_i, \xi_i, a_i, \delta_i^1, \delta_i^2, \ldots) \in MB_i^{k-1}(RAT)$ such that $\boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i)$. Let $\sigma_{-i}^{k-1} : \Theta_0 \times T_{\Theta,-i}^* \to \Delta(A_{-i})$ be any conditional distribution associated to the measure on $\Theta_0 \times T_{\Theta,-i}^* \times A_{-i}$ which is the pushforward of $\boldsymbol{\pi}_i^*(t_i^*)$ given by the mapping

$$(\theta_0, t_{-i}^*) \mapsto (\theta_0, \boldsymbol{m}_{-i}^*(t_{-i}^*), \boldsymbol{a}_{-i}^*(t_{-i}^*)).$$

Define $\sigma_{-i} : \Theta_0 \times T_{-i} \to \Delta(A_{-i})$ so that $\sigma_{-i}(\theta_0, t_{-i}) = \sigma_{-i}^{k-1}(\theta_0, \boldsymbol{\tau}_{-i}^{\mathsf{T}}(t_{-i}))$ for all $(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}$. By the induction hypothesis, $\operatorname{supp} \sigma_{-i}(\theta_0, t_{-i}) \subseteq \boldsymbol{ICR}_{-i}^{\mathsf{T},k-1}(t_{-i})$ for $\boldsymbol{\pi}_i(t_i)$-almost every $(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}$. Moreover, as before, we have

$$\begin{aligned} a_i &\in \operatorname*{arg\,max}_{a_i'} \int_{\Theta_0 \times T_{\Theta_0 \times X \times A,-i}^*} \boldsymbol{g}_i(\theta_0, \theta_i, \boldsymbol{\theta}_{-i}^*(t_{-i}^*), a_i', \boldsymbol{a}_{-i}^*(t_{-i}^*)) \boldsymbol{\pi}_i^*(t_i^*)[d\theta_0 \times dt_{-i}^*] \\ &= \operatorname*{arg\,max}_{a_i'} \int_{\Theta_0 \times T_{\Theta,-i}^*} \boldsymbol{g}_i(\theta_0, \theta_i, \boldsymbol{\theta}_{-i}^*(t_{-i}^*), a_i', \sigma_{-i}^{k-1}(\theta_0, t_{-i}^*)) \boldsymbol{\pi}_i^*(t_i^*)[d\theta_0 \times dt_{-i}^*] \\ &= \operatorname*{arg\,max}_{a_i'} \int_{\Theta_0 \times T_{-i}} \boldsymbol{g}_i(\theta_0, \theta_i, \boldsymbol{\theta}_{-i}(t_{-i}), a_i', \sigma_{-i}(\theta_0, t_{-i})) \boldsymbol{\pi}_i(t_i)[d\theta_0 \times dt_{-i}], \end{aligned}$$

where, again, the first line follows from $t_i^* \in MB_i^0(RAT) = RAT_i$ and the second from $\boldsymbol{\tau}_i^{\mathsf{T}}(t_i) = \boldsymbol{m}_i^*(t_i^*)$. This proves that $a_i \in \boldsymbol{ICR}_i^{\mathsf{T},k}(t_i)$.

Part II

Here we prove that for all $i \in I$ and $k \geq 1$,

$$\boldsymbol{ICR}_i^{\mathsf{T},k} \subseteq \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists\, t_i^* \in MB_i^{k-1}(RAT) \text{ s.t. } \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \text{ and } \boldsymbol{a}_i^*(t_i^*) = a_i\right\}.$$

[28] See e.g. Dudley (1989, pp. 269-270).

Since for every $(t_i, a_i) \in T_i \times A_i$ the sequence

$$MB_i^k(RAT) \cap (\boldsymbol{m}_i^*)^{-1}(\boldsymbol{\tau}_i^{\mathsf{T}}(t_i)) \cap (\boldsymbol{a}_i^*)^{-1}(a_i) \qquad (k \geq 1)$$

is a decreasing sequence of nonempty compact sets, we will have also proved that

$$\boldsymbol{ICR}_i^{\mathsf{T}} \subseteq \left\{(t_i, a_i) \in T_i \times A_i \;\middle|\; \exists\, t_i^* \in CB_i(RAT) \text{ s.t. } \boldsymbol{m}_i^*(t_i^*) = \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \text{ and } \boldsymbol{a}_i^*(t_i^*) = a_i\right\}.$$

Fix once and for all some $i \in I$, $k \geq 1$ and $(\xi_i, \xi_{-i}) \in \Xi$. For each $(t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T},k}$ let $V_i(t_i, a_i)$ designate the set of all $\nu_i \in \Delta(\Theta_0 \times T_{-i} \times A_{-i})$ that rationalize a_i for t_i at order k. For each $(t_{-i}, a_{-i}) \in \boldsymbol{ICR}_{-i}^{\mathsf{T},k-1}$ let $V_{-i}(t_{-i}, a_{-i})$ designate the set of all $\nu_{-i} \in \Delta(\Theta_0 \times T_i \times A_i)$ that rationalize a_{-i} for t_{-i} at order $k-1$. Define a $(\Theta_0 \times X \times A)$-based type space $\overline{\mathsf{T}}$ as follows:

$$\overline{\mathsf{T}} = \langle \Theta_0 \times X \times A, (\overline{T}_j, \overline{\boldsymbol{\pi}}_j, \boldsymbol{x}_j, \boldsymbol{a}_j^*)_{j \in I} \rangle,$$

where $\overline{T}_j = T_j \times A_j$ and $(\boldsymbol{x}_j, \boldsymbol{a}_j^*)(t_j, a_j) = (\boldsymbol{\theta}_j(t_j), \xi_j, a_j)$ for all $j \in I$ and $(t_j, a_j) \in \overline{T}_j$, and $\overline{\boldsymbol{\pi}}_j : \overline{T}_j \to \Delta(\Theta_0 \times \overline{T}_{-j})$ is an arbitrary measurable extension of an arbitrary measurable selector from the correspondence V_j.[29]

It is clear that the natural projections of the spaces $(\overline{T}_j)_{j \in I}$ on the spaces $(T_j)_{j \in I}$ constitute a belief morphism from $\overline{\mathsf{T}}$ onto T. In particular, we have

$$\begin{aligned}
\boldsymbol{m}_i^*(\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i)) &= \boldsymbol{\tau}_i^{\mathsf{T}}(t_i) \quad \text{and} \quad \boldsymbol{a}_i^*(\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i)) = a_i \quad \forall (t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T},k}, \\
\boldsymbol{m}_{-i}^*(\boldsymbol{\tau}_{-i}^{\overline{\mathsf{T}}}(t_{-i}, a_{-i})) &= \boldsymbol{\tau}_{-i}^{\mathsf{T}}(t_{-i}) \quad \text{and} \quad \boldsymbol{a}_{-i}^*(\boldsymbol{\tau}_{-i}^{\overline{\mathsf{T}}}(t_{-i}, a_{-i})) = a_{-i} \quad \forall (t_{-i}, a_{-i}) \in \boldsymbol{ICR}_{-i}^{\mathsf{T},k-1}.
\end{aligned}$$

Thus, to conclude the proof we only need to show that $\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i) \in MB_i^{k-1}(RAT)$ for every $(t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T},k}$. Since a_i is a best reply to $\overline{\boldsymbol{\pi}}_i(t_i, a_i)$,

$$\begin{aligned}
a_i \;&\in\; \operatorname*{arg\,max}_{a_i'} \sum_{\theta_0, \theta_{-i}, a_{-i}} \boldsymbol{g}(\theta_0, \boldsymbol{\theta}_i(t_i), \theta_{-i}, a_i', a_{-i}) \overline{\boldsymbol{\pi}}_i(t_i, a_i)[\theta_0 \times \boldsymbol{\theta}_{-i}^{-1}(\theta_{-i}) \times a_{-i}] \\
&\in\; \operatorname*{arg\,max}_{a_i'} \sum_{\theta_0, \theta_{-i}, a_{-i}} \boldsymbol{g}(\theta_0, \boldsymbol{\theta}_i(t_i), \theta_{-i}, a_i', a_{-i}) \overline{\boldsymbol{\pi}}_i^{\overline{\mathsf{T}},1}(t_i, a_i)[\theta_0, \theta_{-i}, \xi_{-i}, a_{-i}],
\end{aligned}$$

which proves $\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i) \in MB_i^0(RAT) = RAT_i$. If $k = 1$ then the proof is complete, and if $k \geq 2$ then an analogous argument establishes that $\boldsymbol{\tau}_{-i}^{\overline{\mathsf{T}}}(t_{-i}, a_{-i}) \in MB_{-i}^0(RAT)$ for every $(t_{-i}, a_{-i}) \in \boldsymbol{ICR}_{-i}^{\mathsf{T},1}$. Thus we can assume that for some $1 \leq \ell < k$ we have

$$\boldsymbol{\tau}_j^{\overline{\mathsf{T}}}(t_j, a_j) \in MB_j^{\ell-1}(RAT) \qquad \forall j \in I,\; \forall (t_j, a_j) \in \boldsymbol{ICR}_j^{\mathsf{T},\ell}. \tag{4}$$

It remains to prove that $\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i) \in MB_i^{\ell}(RAT)$ for every $(t_i, a_i) \in \boldsymbol{ICR}_i^{\mathsf{T},\ell+1}$ and, if $k > \ell + 1$, also that the analogous claim holds for player $-i$. In effect, since (4) already guarantees that $\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i) \in MB_i^{\ell-1}(RAT)$, it suffices to prove $\boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i)[\Theta_0 \times MB_{-i}^{\ell-1}(RAT)] = 1$. Indeed, $\overline{\boldsymbol{\pi}}_i(t_i, a_i)$ rationalizes a_i for t_i at order k, so it does so at order $\ell + 1$ as well. Thus,

$$\begin{aligned}
1 = \overline{\boldsymbol{\pi}}_i(t_i, a_i)\left[\Theta_0 \times \boldsymbol{ICR}_{-i}^{\mathsf{T},\ell}\right] &\leq \overline{\boldsymbol{\pi}}_i(t_i, a_i)\left[\Theta_0 \times \left\{(t_i, a_i) \in T_{-i} \times A_{-i} \;\middle|\; \boldsymbol{\tau}_{-i}^{\overline{\mathsf{T}}}(t_{-i}, a_{-i}) \in MB_{-i}^{\ell-1}(RAT)\right\}\right] \\
&= \boldsymbol{\tau}_i^{\overline{\mathsf{T}}}(t_i, a_i)\left[\Theta_0 \times MB_{-i}^{\ell-1}(RAT)\right],
\end{aligned}$$

[29]Such selector exists because V_j is a nonempty-valued, closed-graph correspondence between compact spaces (see footnote 21). A measurable extension to $T_j \times A_j$ then exists because the codomain $\Delta(\Theta_0 \times T_{-j} \times A_{-j})$ is Polish.

where the inequality follows from (4) and the second line from the fact that $(\boldsymbol{\tau}_j^{\overline{\mathsf{T}}})_{j\in I}$ is a belief morphism from $\overline{T}$ to the universal $(\Theta_0 \times X \times A)$-based type space. The proof of the analogous claim for player $-i$ when $k > \ell + 1$ is analogous.

B Rationalizability on the interim strategic form

Fix a finite Θ-based type space $\mathsf{T} = \langle \Theta, (T_i, \boldsymbol{\pi}_i, \boldsymbol{\theta}_i)_{i\in I}\rangle$ and the corresponding Bayesian game. The player set of the interim strategic form of the Bayesian game is $T_1 \cup T_2$. Letting $\boldsymbol{B}_i$ denote the set of all mappings from T_i to A_i for all $i \in I$, an element $\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2) \in \boldsymbol{B}_1 \times \boldsymbol{B}_2$ thus specifies an action profile for this game — for each $i \in I$ and $t_i \in T_i$, the action chosen by t_i is $\boldsymbol{b}_i(t_i)$, and the corresponding payoff to player/type t_i depends only on $\boldsymbol{b}_i(t_i)$ and $\boldsymbol{b}_{-i}$:

$$U_{t_i}(\boldsymbol{b}) = \sum_{\theta_0\in\Theta_0}\sum_{t_{-i}\in T_{-i}} \boldsymbol{\pi}_i(t_i)[\theta_0, t_{-i}]\boldsymbol{g}_i(\theta_0, \boldsymbol{\theta}_i(t_i), \boldsymbol{\theta}_{-i}(t_{-i}), \boldsymbol{b}_i(t_i), \boldsymbol{b}_{-i}(t_{-i})).$$

Interim rationalizability is a process of iterated maximal elimination, for each t_i, of actions that are not best responses to conjectures of the form $\boldsymbol{\sigma}_{-i} : T_{-i} \to \Delta(A_{-i})$.[30] Let $\boldsymbol{\Sigma}_{-i}$ denote the set of all such conjectures. The strategic form expected payoff of type t_i from using action a_i given conjecture $\boldsymbol{\sigma}_{-i}$ is (once again slightly abusing notation)

$$U_{t_i}(a_i, \boldsymbol{\sigma}_{-i}) = \sum_{\boldsymbol{b}_{-i}\in\boldsymbol{B}_{-i}}\prod_{t_{-i}\in T_{-i}} \boldsymbol{\sigma}_{-i}(t_{-i})[\boldsymbol{b}_{-i}(t_{-i})]U_{t_i}(\boldsymbol{b}_i, \boldsymbol{b}_{-i}),$$

where $\boldsymbol{b}_i \in \boldsymbol{B}_i$ is any function such that $\boldsymbol{b}_i(t_i) = a_i$. Rationalizability on the interim strategic form (ISFR) is recursively defined as follows: for all $i \in I$ and $t_i \in T_i$, $\boldsymbol{ISFR}_i^{\mathsf{T},0}(t_i) = A_i$, and for all $k \geq 0$

$$\boldsymbol{ISFR}_i^{\mathsf{T},k+1}(t_i) = \left\{ a_i \in A_i \;\middle|\; \begin{array}{ll} \exists\, \boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i} \;: & \\ \text{(ISFR1)} & a_i \in \arg\max_{a'_i\in A_i} U_{t_i}(a'_i, \boldsymbol{\sigma}_{-i}) \\ \text{(ISFR2)} & \operatorname{supp} \boldsymbol{\sigma}_{-i}(t_{-i}) \subseteq \boldsymbol{ISFR}_{-i}^{\mathsf{T},k}(t_{-i}) \quad \forall t_{-i} \in T_{-i} \end{array} \right\}.$$

Proposition 5. *For every $i \in I$, $t_i \in T_i$, and $k \geq 0$, $\boldsymbol{IIR}_i^{\mathsf{T},k}(t_i) = \boldsymbol{ISFR}_i^{\mathsf{T},k}(t_i)$.*

Proof. It suffices to show that conditions (IIR1) and (ISFR1) are equivalent, as the result then follows by an obvious induction. Thus, fix $i \in I$, $t_i \in T_i$, and $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. We shall prove that

$$\boldsymbol{g}_i(\boldsymbol{\theta}_i(t_i), a_i, \boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})) = U_{t_i}(a_i, \boldsymbol{\sigma}_{-i}),$$

where $\boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})$ is defined as in (1). Indeed, for every $a_{-i} \in A_{-i}$ and $t_{-i} \in T_{-i}$ define

$$\boldsymbol{B}_{t_{-i}}^{a_{-i}} = \{\boldsymbol{b}_{-i} \in \boldsymbol{B}_{-i} : \boldsymbol{b}_{-i}(t_{-i}) = a_{-i}\},$$

[30] This is *independent* rationalizability on the interim strategic form of the Bayesian game. But, by Kuhn's (1953) equivalence result, with $I = \{1, 2\}$, correlated and independent rationalizability on the interim strategic form are equivalent (T_{-i} is like a coalition with perfect recall in the extensive form of the Bayesian game).

so that

$$\sigma_{-i}(t_{-i})[a_{-i}] = \sum_{\boldsymbol{b}_{-i} \in B^{a_{-i}}_{t_{-i}}} \prod_{t'_{-i} \in T_{-i}} \sigma_{-i}(t'_{-i})[\boldsymbol{b}_{-i}(t'_{-i})].$$

Then

$$\begin{aligned}
&\boldsymbol{g}_i(\boldsymbol{\theta}_i(t_i), a_i, \boldsymbol{\mu}_i(t_i, \boldsymbol{\sigma}_{-i})) = \\
&\quad = \sum_{\theta_0 \in \Theta_0} \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[\theta_0, t_{-i}] \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(t_{-i})[a_{-i}] \boldsymbol{g}_i(\theta_0, \boldsymbol{\theta}_i(t_i), \boldsymbol{\theta}_{-i}(t_{-i}), a_i, a_{-i}) \\
&\quad = \sum_{\theta_0 \in \Theta_0} \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[\theta_0, t_{-i}] \sum_{a_{-i} \in A_{-i}} \sum_{\boldsymbol{b}_{-i} \in \boldsymbol{B}^{a_{-i}}_{t_{-i}}} \prod_{t'_{-i} \in T_{-i}} \sigma_{-i}(t'_{-i})[\boldsymbol{b}_{-i}(t'_{-i})] \boldsymbol{g}_i(\theta_0, \boldsymbol{\theta}_i(t_i), \boldsymbol{\theta}_{-i}(t_{-i}), a_i, a_{-i}) \\
&\quad = \sum_{\theta_0 \in \Theta_0} \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[\theta_0, t_{-i}] \sum_{\boldsymbol{b}_{-i} \in \boldsymbol{B}_{-i}} \prod_{t'_{-i} \in T_{-i}} \sigma_{-i}(t'_{-i})[\boldsymbol{b}_{-i}(t'_{-i})] \boldsymbol{g}_i(\theta_0, \boldsymbol{\theta}_i(t_i), \boldsymbol{\theta}_{-i}(t_{-i}), a_i, \boldsymbol{b}_{-i}(t'_{-i})) \\
&\quad = U_{t_i}(a_i, \boldsymbol{\sigma}_{-i}).
\end{aligned}$$

□

References

Aliprantis, Charalambos D. and Kim C. Border (1999), *Infinite Dimensional Analysis*. Springer, second edition.

Battigalli, Pierpaolo (1988), "Implementable strategies, prior information and the problem of credibility in extensive games." *RISEC (International Review of Economics and Business)*, 35, 705–733.

Battigalli, Pierpaolo (1997), "On rationalizability in extensive games." *Journal of Economic Theory*, 74, 40–61.

Battigalli, Pierpaolo (2003), "Rationalizability in infinite, dynamic games with incomplete information." *Research in Economics*, 57, 1–38.

Battigalli, Pierpaolo and Marciano Siniscalchi (2002), "Strong belief and forward induction reasoning." *Journal of Economic Theory*, 106, 356–391.

Battigalli, Pierpaolo and Marciano Siniscalchi (2003), "Rationalization and incomplete information." *Advances in Theoretical Economics*, 3. Available at http://www.bepress.com/bejte/advances/vol3/iss1/art3.

Battigalli, Pierpaolo and Marciano Siniscalchi (2007), "Interactive epistemology in games with payoff uncertainty." *Research in Economics*, 61, 165–184.

Bergemann, Dirk and Stephen Morris (2005), "Robust implementation: The role of large type spaces." Mimeo, Yale University (previous version: 2003).

Bergemann, Dirk and Stephen Morris (2007), "Belief-free incomplete information games." Cowles Foundation Discussion Paper 1619.

Bergemann, Dirk and Stephen Morris (2009), "Robust implementation in direct mechanisms." *Review of Economic Studies*, 76, 1175–1204.

Bernheim, B. Douglas (1984), "Rationalizable strategic behavior." *Econometrica*, 52, 1007–1028.

Brandenburger, Adam and Eddie Dekel (1987), "Rationalizability and correlated equilibria." *Econometrica*, 55, 1391–1402.

Carlsson, Hans and Eric van Damme (1993), "Global games and equilibrium selection." *Econometrica*, 61, 989–1018.

Chen, Yi-Chun, Alfredo Di Tillio, Eduardo Faingold, and Siyang Xiong (2009), "Uniform topologies on types." Cowles Foundation Discussion Paper 1734.

Dekel, Eddie, Drew Fudenberg, and Stephen Morris (2006), "Topologies on types." *Theoretical Economics*, 1, 275–309.

Dekel, Eddie, Drew Fudenberg, and Stephen Morris (2007), "Interim correlated rationalizability." *Theoretical Economics*, 2, 15–40.

Dudley, Richard M. (1989), *Real Analysis and Probability*. Wadsworth & Brooks/Cole.

Ely, Jeffrey C. and Marcin Pęski (2006), "Hierarchies of belief and interim rationalizability." *Theoretical Economics*, 1, 19–65.

Ely, Jeffrey C. and Marcin Pęski (2008), "Critical types." Mimeo, Northwestern University.

Fagin, R. J., Joe Halpern, Yoram Moses, and M. Vardi (1995), *Reasoning About Knowledge*. MIT Press.

Harsanyi, John (1967-68), "Games with incomplete information played by bayesian players, i-iii." *Management Science*, 14, 159–182, 320–334, 486–502.

Heifetz, Aviad and Dov Samet (1998), "Topology-free typology of beliefs." *Journal of Economic Theory*, 82, 324–341.

Kechris, Alexander S. (1995), *Classical Descriptive Set Theory*. Springer-Verlag, New York.

Klemperer, Paul (1998), "Auctions with almost common values: The 'wallet game' and its applications." *European Economic Review*, 42, 757–770.

Kuhn, Harold W. (1953), "Extensive games and the problem of information." In Harold W. Kuhn and A. W. Tucker, editors, "Contributions to the Theory of Games II," Princeton University Press, pp. 193–216.

Liu, Qingmin (2009), "On redundant types and bayesian formulation of incomplete information." *Journal of Economic Theory*, 144, 2115–2145.

Mertens, Jean-Francois and Shmuel Zamir (1985), "Formulation of bayesian analysis for games with incomplete information." *International Journal of Game Theory*, 14, 1–29.

Morris, Stephen and H. S. Shin (2007), "Common belief foundations of global games." Mimeo, Princeton University.

Osborne, Martin J. and Ariel Rubinstein (1994), *A Course in Game Theory*. MIT Press.

Pearce, David G. (1984), "Rationalizable strategic behavior and the problem of perfection." *Econometrica*, 52, 1029–1050.

Penta, Antonio (2009), "Higher order beliefs in dynamic environments." Mimeo, University of Pennsylvania.

Sadzik, Tomasz (2007), "Beliefs revealed in bayesian equilibrium." Mimeo, Stanford University.

Tan, T. and S. Werlang (1988), "The bayesian foundation of solution concepts of games." *Journal of Economic Theory*, 45, 370–391.

Weinstein, Jonathan and Muhamet Yildiz (2007), "A structure theorem for rationalizability with application to robust predictions of refinements." *Econometrica*, 75, 365–400.