HIERARCHIES OF CONDITIONAL BELIEFS AND INTERACTIVE EPISTEMOLOGY IN DYNAMIC GAMES

Pierpaolo Battigalli Marciano Siniscalchi*

Abstract

The epistemic analysis of solution concepts for dynamic games involves statements about the players' beliefs conditional upon different histories of play, their conditional beliefs about each other's conditional beliefs, etc. To represent such statements, we construct a space of infinite (coherent) hierarchies of conditional probability systems, defined with respect to a fixed collection of relevant hypotheses concerning an external state (*e.g.,* the strategy profile being played.) As an application, we derive results about common certainty of the opponent's rationality conditional on an arbitrary collection of histories in multistage games with observed actions and (possibly) incomplete information.

1 Introduction

A player's strategy in an extensive-form game is a complete description of her *dispositions to act* at different information sets. In a symmetric fashion, the analysis of rationality in extensive games relies (at least implicitly) on an equally complete description of a player's *dispositions to hold beliefs* about her opponents' strategic choices.

We take the view that preserving this symmetry is just as natural and desirable in any treatment of the epistemic foundations of solution concepts for extensive games.

Higher-order beliefs, *i.e.,* beliefs about beliefs ..., are key to the latter line of research. Thus, two questions arise naturally. First, is it possible to model players' dispositions to hold *hierarchical* beliefs in a complete and consistent

*We thank the Associate Editor and two anonymous referees for helpful comments. The usual disclaimer applies.

way? And, if so, can we make progress towards understanding some of the so-called "paradoxes" of extensive-form analysis by exploiting the expressive power of such a model?

Our first contribution answers the former question in the affirmative. In the standard, normal-form setting, it is well-known (see, *e.g.,* Mertens and Zamir [19], Brandenburger and Dekel [10]) that, under fairly general conditions, there exists a "universal" space of *epistemic types*. Its elements are sequences of probability measures, corresponding to progressively higher-order beliefs. Thus, essentially any (coherent) statement about players' reciprocal beliefs has a representation in the universal space.

In this paper, we extend this type of construction by considering a space whose elements are sequences of *collections* of (conditional) probabilities. In particular, we consider collections which satisfy Bayes' rule whenever possible, so that our representation of agents' dispositions to believe coincides with the notion of a conditional probability system (or CPS), due to Alfred Renyi [26],[1] and the elements of the "universal" space we construct are actually infinite hierarchies of CPSs.

We point out that, as in [19] and [10], our framework applies to more general situations where a natural "basic" (or external) domain of uncertainty exists, and agents hold interacting beliefs conditional on a fixed collection of (relevant) hypotheses about the prevailing external state.

As in the normal-form case, a single (coherent) hierarchical sequence of CPSs may be regarded as an epistemic type — that is, a complete and *explicit* description of an agent's conditional beliefs of arbitrary order. However, in the spirit of Harsanyi [15], one may also list a set of epistemic types for each agent, and associate with each type a CPS over the Cartesian product of the set of external states and the collection of types listed for the other agents. Each type thus defined generates an infinite hierarchy of CPSs; indeed, extending analogous results due to Mertens and Zamir [19], we show that every such *implicit* description of a type space corresponds to a (beliefs-closed) subset of the space of (coherent) hierarchies of CPSs, which can thus be rightfully deemed universal. This is the second contribution of this paper.

We would like to suggest that the model we propose may be usefully employed to further our understanding of some of the puzzles and paradoxes of extensive-form analysis. To support this claim, we provide a collection of results related to the notion of (conditional) common certainty of rationality in two-player multistage games with observed actions and (possibly) incomplete information.

Common knowledge or certainty of rationality are central ideas in the literature on the epistemic foundations of normal-form solution concepts (*e.g.,* Tan and Werlang [30]); they have also been employed in connection with extensive games (*e.g.,* Ben-Porath [7]), albeit often engendering much controversy (see, for example, Aumann [1, 2], Binmore [9] and Reny [24] on backward induction.)

[1] Myerson [20] pioneered the use of CPSs in game theory.

In an effort to at least partially clarify some of the controversial issues involved, we propose a notion of common certainty of the opponent's rationality (CCOR) given an arbitrary collection of histories. Our definition formalizes the following sequence of assumptions: for $i = 1, 2$ and $j \neq i$, $(0.i)$ player i is rational, $(1.i)$ player i is disposed to believe, after each history h in an arbitrarily specified collection $\mathcal{F}$, that $(0.j)$ is true, $(2.i)$ player i is disposed to believe, after each history $h \in \mathcal{F}$, that $(1.j)$ is true, etc.

We show that, for any collection of histories $\mathcal{F}$, these assumptions characterize an iterative elimination procedure which is reminiscent of rationalizability (Bernheim [8], Pearce [22]), but incorporates stronger, extensive-form-motivated restrictions. More specifically:

- In normal-form games, our procedure coincides with rationalizability; hence, our results formally extend those of Tan and Werlang [30].
- In extensive-form games, if one takes the collection of relevant conditioning events to be the (singleton) *initial* history, one obtains a characterization of initial CCOR, as defined and analyzed by Ben-Porath [7] in the more restricted class of generic perfect information games.
- However, we can also characterize CCOR at any *subsequent* history. This allows us to provide a simple and transparent answer to questions such as whether or not there can be CCOR in the "Centipede" game if Player 1 does not choose "down" at the initial history.
- Finally, in our opinion, the most interesting and novel applications of our result involves a non-singleton collection of conditioning events. For instance:
  - one may verify whether there can be CCOR conditional on the set of histories comprising a given path of play;
  - imposing CCOR given a collection of histories comprising a path of play as well as select off-path histories may capture elements of *forward induction* (see Example 1 in Section 5);
  - using results due to Reny [24], we can show that, in generic perfect information games, CCOR is possible given the collection $\mathcal{R}$ of all histories that are (i) consistent with the rationality of both players, and (ii) such that the player moving at an history $h \in \mathcal{R}$ does not have a dominant action, if and only if every $h \in \mathcal{R}$ is on the backward induction path.

We have already mentioned and briefly discussed the literature more immediately related to the present paper. Other relevant contributions on the foundations of (extensive form) game theory include Aumann [1, 2], Balkenborg and Winter [3], Dekel and Gul [11], Samet [27] and Stalnaker [28, 29]. A more detailed discussion of some of these papers will be deferred to the concluding section.

The paper is organized as follows. Section 2 contains the construction of the (universal) space of infinite hierarchies of CPSs. Section 3 discusses implicit (and typically finite) representations of type spaces and relates the latter to the universal space constructed in Section 2. Belief operators are the subject of Section 4. All game-theoretic results, as well as illustrative examples, appear in Section 5. Finally, Section 6 concludes. Some of the less instructive proofs are collected in an appendix. Omitted proofs and examples of peripheral facts mentioned in the paper are available upon request.[2]

2 Infinite Hierarchies of Conditional Beliefs

2.1 Conditional Probability Systems and Higher Order Beliefs

For a given Polish (separable, and completely metrizable) space X, let $\mathcal{A}$ be the Borel sigma-algebra on X and $\mathcal{B} \subset \mathcal{A}$ a non-empty, finite or countable collection such that $\emptyset \notin \mathcal{B}$ and each $B \in \mathcal{B}$ is both *closed and open*. The interpretation is that a certain individual i is uncertain about the "true" element $x \in X$, and $\mathcal{B}$ represents a collection of observable events, or "relevant hypotheses." In particular, we will mostly be interested in the following situation: there is a set Σ of basic "external" states, and a set Z consisting of (some representation) of *another* individual's beliefs about Σ; then each point (state) in the set $X = \Sigma \times Z$ provides a description of "external" as well as "epistemic" features of the situation at hand. In a game, this could comprise a description of the *strategy profile* being played, and a representation of the *beliefs* held by individual i's opponent. The set $\mathcal{B}$ could consist of hypotheses concerning the "external" state only, *i.e.,* sets of the form $B \times Z$ for $B \subset \Sigma$; as long as the latter is finite, the elements of $\mathcal{B}$ will be guaranteed to be both closed and open.[3]

For a different example, X may be the set of sample paths in a repeated experiment with finitely many outcomes, or the set of complete histories in a supergame with a finite stage game, while elements of $\mathcal{B}$ may be equivalence classes of histories sharing a common initial subhistory. In this case, too, the conditioning events may be shown to be both closed and open.

A *conditional probability system* (or CPS) on $(X, \mathcal{A}, \mathcal{B})$ is a mapping

$$\mu(\cdot|\cdot) : \mathcal{A} \times \mathcal{B} \to [0, 1]$$

satisfying the following axioms:

Axiom 1. For all $B \in \mathcal{B}$, $\mu(B|B) = 1$.

Axiom 2. For all $B \in \mathcal{B}$, $\mu(\cdot|B)$ is a probability measure on $(X, \mathcal{A})$.

[2] They are also available in electronic format at `http://www.princeton.edu/~marciano`.

[3] This fact is used in the proof of Lemma 1.

Axiom 3. For all $A \in \mathcal{A}$, $B, C \in \mathcal{B}$, if $A \subset B \subset C$ then $\mu(A|B)\mu(B|C) = \mu(A|C)$.[4]

The set of probability measures on $(X, \mathcal{A})$ is denoted by $\Delta(X)$; the set of conditional probability systems on $(X, \mathcal{A}, \mathcal{B})$ can be regarded as a subset of $[\Delta(X)]^{\mathcal{B}}$ (the set of mappings from $\mathcal{B}$ to $\Delta(X)$) and it is denoted by $\Delta^{\mathcal{B}}(X)$. Accordingly, we often write $\mu = (\mu(\cdot|B))_{B\in\mathcal{B}} \in \Delta^{\mathcal{B}}(X)$. The topology on X and $\mathcal{A}$ (the smallest sigma-algebra containing this topology) are always understood and need not be explicit in our notation. Thus we simply say "conditional probability system (or CPS) on $(X, \mathcal{B})$."

We regard $\Delta^{\mathcal{B}}(X)$ as the space of possible conditional beliefs of an individual, say j, and we wish to define the higher order beliefs of another individual i about the beliefs of j. We argue below that it is conceptually appropriate to define such higher order beliefs over the Borel sigma-algebra generated by the product topology of weak convergence of measures.

Fix a Borel set A, a relevant hypothesis B and a real number $p \in [0, 1]$. The informal statement, "conditional on B, individual j would assign probability at least p to A" corresponds to the set $\beta^p_B(A) \equiv \{\mu \in \Delta^{\mathcal{B}}(X) : \mu(A|B) \geq p\} \subset \Delta^{\mathcal{B}}(X)$.

In order to formalize more complex statements such as, "conditional on $C \in \mathcal{B}$, i would assign a subjective conditional probability to the event 'j would assign probability at least p to A conditional on B' ", we must endow $\Delta^{\mathcal{B}}(X)$ with a sigma-algebra including all sets $\beta^p_B(A)$, for all Borel-measurable $A \subset X$, $B \in \mathcal{B}$ and $p \in [0, 1]$. It is then natural to consider the sigma-algebra generated by such sets, which we denote by $\mathcal{A}^{+1}$ (cf. Heifetz and Samet [17].)

It turns out that, since X is assumed to be a Polish space, the rather intuitive measure-theoretic structure just described is entirely consistent with a particularly convenient topological structure on the set of conditional probability systems.

More specifically, endow $\Delta(X)$ with the topology of weak convergence of measures, and $[\Delta(X)]^{\mathcal{B}}$ with the product topology. Consider the Borel sigma-algebra on $[\Delta(X)]^{\mathcal{B}}$. Lemma 1 below states that $\Delta^{\mathcal{B}}(X)$ is a closed subset of $[\Delta(X)]^{\mathcal{B}}$. Thus the collection of Borel subsets of $\Delta^{\mathcal{B}}(X)$ is the Borel sigma-algebra of $\Delta^{\mathcal{B}}(X)$ viewed as a topological subspace of $[\Delta(X)]^{\mathcal{B}}$. This Borel (sub–) sigma-algebra is precisely the "natural" sigma-algebra $\mathcal{A}^{+1}$ defined above.[5]

[4] The tuple $(X, \mathcal{A}, \mathcal{B}, \mu)$ is called *conditional probability space* by Rènyi [26]. When X is finite, $\mathcal{A} = 2^X$, $\mathcal{B} = 2^X \backslash \{\emptyset\}$, we obtain Myerson's [20] conditional probability systems.

[5] Since X is a Polish space every probability measure in $\Delta(X)$ is regular. Therefore the Borel sigma-algebra on the topological space $\Delta(X)$ coincides with the sigma-algrebra generated by the base of subsets $\beta^p(A) = \{m \in \Delta(X) : m(A) \geq p\}$, A measurable, $p \in [0, 1]$ (see, *e.g.,* Kechris [18], Theorem 17.24.) Since $\Delta(X)$ is Polish, it is second countable. Therefore the product sigma-algebra on $[\Delta(X)]^{\mathcal{B}}$ coincides with the Borel sigma-algebra generated by the product topology (*e.g.,* Kechris [18], p. 68.) This implies the result stated above. Similar arguments justify defining higher order beliefs on Borel sigma algebras in other papers on hierarchies of beliefs where the set of external states has a "nice" topological structure. (Kim Border and Aviad Heifetz kindly provided the relevant mathematical references.)

Lemma 1. *The set $\Delta^{\mathcal{B}}(X)$ of conditional probability systems on $(X, \mathcal{B})$ is a closed subset of $[\Delta(X)]^{\mathcal{B}}$. Therefore $\Delta^{\mathcal{B}}(X)$ (endowed with the relative topology inherited from $[\Delta(X)]^{\mathcal{B}}$) and $X \times \Delta^{\mathcal{B}}(X)$ (endowed with the product topology) are Polish spaces.*[6]

Proof: See the Appendix. ■

Let $X^{+1} = X \times \Delta^{\mathcal{B}}(X)$ and let $\mathcal{C} : \mathcal{A} \to 2^{X^{+1}}$ be defined by $\mathcal{C}(A) = A \times \Delta^{\mathcal{B}}(X)$. Thus $\mathcal{C}(\mathcal{B}) = \{C \subset X : \exists B \in \mathcal{B}, C = B \times \Delta^{\mathcal{B}}(X)\}$ is a set of "cylinders" generated by $\mathcal{B}$ and represents a copy of $\mathcal{B}$ in X^{+1}. Then we can define the set of "second order" CPSs $\Delta^{\mathcal{C}(\mathcal{B})}(X^{+1})$. Since X^{+1} is a Polish space, it follows from Lemma 1 that also $\Delta^{\mathcal{C}(\mathcal{B})}(X^{+1})$ (endowed with the appropriate topology as above) is a Polish space. Each element $\mu_i^{+1} \in \Delta^{\mathcal{C}(\mathcal{B})}(X)$ is a countable collection of individual i's conditional joint beliefs about the true value of $x \in X$ and $\mu_j \in \Delta^{\mathcal{B}}(X)$ — individual j's conditional beliefs about $x \in X$, — whereby the conditioning events, or hypotheses, are essentially the same as in $\mathcal{B}$.

Note that $\Delta^{\mathcal{C}(\mathcal{B})}(X^{+1})$ can be regarded as a subset of $\left[\Delta(X^{+1})\right]^{\mathcal{B}}$. Thus, we are somewhat justified in adopting the simpler notation $\Delta^{\mathcal{B}}(X^{+1})$ whenever the precise structure of the conditioning events is clear from the context and/or need not be specified, even though $\mathcal{B}$ is not a collection of subsets of X^{+1}. More generally, let $Y = X \times Z$, $\mathcal{B} \subset 2^X$, $\mathcal{B}_Y = \{C \subset Y : \exists B \in \mathcal{B}, C = B \times Z\}$; then the set of CPSs on $(Y, \mathcal{B}_Y)$ will be equivalently denoted by $\Delta^{\mathcal{B}_Y}(Y)$ or $\Delta^{\mathcal{B}}(Y)$.

Finally, for any probability measure ν on the product space $Y = X \times Z$ let $mrg_X \nu \in \Delta(X)$ denote the marginal measure on X. In what follows it is useful to note that, if $\mu = (\mu(\cdot|B \times Z))_{B \in \mathcal{B}} \in \Delta^{\mathcal{B}}(Y)$, then $(mrg_X \mu(\cdot|B \times Z))_{B \in \mathcal{B}} \in \Delta^{\mathcal{B}}(X)$.

2.2 Inductive Construction

We are now ready for the inductive construction of the space of infinite hierarchies of conditional beliefs and the universal type space. For the sake of simplicity, we assume that there are only two individuals i and j sharing a common space Σ of external states (about which they are uncertain) and a common collection of relevant hypotheses $\mathcal{B}$. The individuals have conditional beliefs about Σ and about each other for every hypothesis $B \in \mathcal{B}$. However, we do not explicitly represent the beliefs of an individual about her own beliefs. The implicit assumption is that an individual always assigns probability one to her true beliefs. As before we assume that Σ is a Polish space and $\mathcal{B}$ is a finite or countable collection of its non-empty subsets which are both closed and open. Define recursively X^n and $\mathcal{B}^n$ as follows:

$X^0 = \Sigma$, $\mathcal{B}^0 = \mathcal{B}$;

for all $n \geq 0$,

$X^{n+1} = \mathcal{C}(X^n) := X^n \times \Delta^{\mathcal{B}^n}(X^n)$,

$\mathcal{B}^{n+1} = \mathcal{C}(\mathcal{B}^n) := \{C \subset X^{n+1} : \exists B \in \mathcal{B}^n, C = B \times \Delta^{\mathcal{B}^n}(X^n)\}$.

[6] If some $B \in \mathcal{B}$ is either non-open or non-closed, $\Delta^{\mathcal{B}}(X)$ may fail to be closed.

An element $\mu^{n+1} \in \Delta^{\mathcal{B}^n}(X^n)$ is an $(n+1)^{th}$-order CPS with elements $\mu^{n+1}(\cdot|B) \in \Delta(X^n)$, $B \in \mathcal{B}^n$. It can be easily verified that in our notation

$$\Delta^{\mathcal{B}^n}(X^n) = \Delta^{\mathcal{B}}(X^n),\ X^{n+1} = \Sigma \times \prod_{k=0}^{k=n} \Delta^{\mathcal{B}}(X^k.)$$

The set of infinite hierarchies of CPSs is $H = \prod_{n=0}^{\infty} \Delta^{\mathcal{B}}(X^n)$. An infinite hierarchy represents an epistemic type and is therefore typically denoted by $t = (\mu^1, \mu^2, ..., \mu^n, ...)$. Lemma 1 implies that for all $n \geq 0$, X^n and $\Delta^{\mathcal{B}}(X^n)$ are Polish spaces. It follows that also H and $\Delta^{\mathcal{B}}(\Sigma \times H)$ are Polish spaces. Note also that for all $k \geq 0$, $\Sigma \times H$ can be decomposed as follows:

$$\Sigma \times H = X^k \times \prod_{n=k}^{\infty} \Delta^{\mathcal{B}}(X^n).$$

2.3 Coherent Hierarchies

We have not yet imposed any coherency condition relating beliefs of different order. Of course, we want to assume that, conditional on any relevant hypothesis, beliefs of different order assign the same probability to the same event. For all integers $k \geq 0$, $n \geq 1$ and subsets $A \subset X^k$ let $\mathcal{C}^n(A)$ denote the subset of X^{k+n} corresponding to A, that is,

$$\mathcal{C}^n(A) = A \times \prod_{m=k}^{m=k+n-1} \Delta^{\mathcal{B}}(X^m).$$

Note that, as the notation suggests, $\mathcal{C}(\mathcal{C}^{n-1}(A)) = \mathcal{C}^n(A)$. Similarly, $\mathcal{C}^\infty(A)$ is the subset of $\Sigma \times H$ corresponding to A (replace n and $k+n-1$ with ∞ in the formula above.) In particular, for any $B \in \mathcal{B}$, $\mathcal{C}^n(B)$ (or $\mathcal{C}^\infty(B)$) is the subset of X^n (or $\Sigma \times H$) corresponding to B. Recall that, for any probability measure ν on a product space $X \times Z$, $mrg_X \nu \in \Delta(X)$ denotes the marginal measure on X.

Definition 1. *An infinite hierarchy of CPS's $t = (\mu^1, \mu^2, ..., \mu^n, ...)$ is* coherent *if for all $B \in \mathcal{B}$, $n = 1, 2, ...,$*

$$mrg_{X^{n-1}} \mu^{n+1}(\cdot|\mathcal{C}^n(B)) = \mu^n(\cdot|\mathcal{C}^{n-1}(B)). \tag{1}$$

The set of coherent hierarchies is denoted by H_c.

The following proposition establishes that we can equivalently describe events concerning the conditional beliefs of a coherent individual i as (measurable) subsets of coherent hierarchies of conditional beliefs or (measurable) subsets of conditional beliefs about the external state and the (coherent or incoherent) infinite hierarchy of individual j.

Proposition 1. *(cf. [10], Proposition 1) There exists a "canonical" homeomorphism $f : H_c \to \Delta^{\mathcal{B}}(\Sigma \times H)$ such that if $\mu = f(\mu^1, \mu^2, ..., \mu^n, ...)$, then for all $B \in \mathcal{B}$, $n = 1, 2, ...$,*

$$mrg_{X^{n-1}} \mu(\cdot | \mathcal{C}^\infty(B)) = \mu^n(\cdot | \mathcal{C}^{n-1}(B)). \tag{2}$$

We first prove the following lemma:

Lemma 2. *Consider the following set:*

$$D = \left\{ (\delta^1, \delta^2, ...) : \forall n \geq 1, \delta^n \in \Delta(X^{n-1}), mrg_{X^{n-1}} \delta^{n+1} = \delta^n \right\}.$$

There is a homeomorphism $h : D \to \Delta(\Sigma \times H)$ such that

$$\forall n \geq 1, \ mrg_{X^{n-1}} h(\delta^1, \delta^2, ...) = \delta^n.$$

Proof: Let $Z^0 = X^0 = \Sigma$, $\forall n \geq 1, Z^n = \Delta^{\mathcal{B}}(X^{n-1})$. Each Z^n is a Polish space and

$$D = \left\{ (\delta^1, \delta^2, ...) : \forall n \geq 1, \delta^n \in \Delta(Z^0 \times ... \times Z^{n-1}), mrg_{X^{n-1}} \delta^{n+1} = \delta^n \right\}.$$

The result then follows from Lemma 1 in [10]. ■

Proof of Proof of Proposition 1: For each $B \in \mathcal{B}$, let $\pi_B : H_c \to D$ be the following projection function:

$$\pi_B(\mu^1, ..., \mu^n, ...) = \left(\mu^1(\cdot \mid B), ..., \mu^n(\cdot \mid \mathcal{C}^{n-1}(B), ... \right).$$

π_B is clearly continuous. By Lemma 2 the mapping

$$f_B = h \circ \pi_B : H_c \to \Delta(\Sigma \times H)$$

is also continuous. Let $\mu(\cdot \mid \mathcal{C}^\infty(B)) = f_B(\mu_1, \mu_2, ...)$. Clearly, $\mu(\mathcal{C}^\infty(B) \mid \mathcal{C}^\infty(B)) = 1$ and for all $n = 1, 2, ...$, eq. (2) is satisfied. Thus the mapping

$$f = (f_B)_{B \in \mathcal{B}} : H_c \to [\Delta(\Sigma \times H)]^{\mathcal{B}}$$

is continuous and satisfies eq. (2). The latter fact implies that f is $1 - 1$ and the restriction of f^{-1} to $f(H_c)$ is continuous. We only have to show that $f(H_c) = \Delta^{\mathcal{B}}(\Sigma \times H)$.

$\left(\Delta^{\mathcal{B}}(\Sigma \times H) \subset f(H_c) \right)$. Take $\mu \in \Delta^{\mathcal{B}}(\Sigma \times H)$ and for all $B \in \mathcal{B}$, $n \geq 1$ define $\mu^n(\cdot | \mathcal{C}^n(B))$ using eq. (2). If $t = (\mu^1, ..., \mu^n, ...) \in H_c$, then $f(t) = \mu \in f(H_c)$. Thus it is sufficient to show that $t = (\mu^1, ..., \mu^n, ...) \in H_c$; in order to do this we only have to verify that each μ^n satisfies Axiom 3 (coherency of t is satisfied by construction.) For each n, let $A^n \subset X^n$ be measurable, $B, C \in \mathcal{B}$ and suppose that $A^n \subset \mathcal{C}^n(B) \subset \mathcal{C}^n(C)$ (thus $B \subset C$.) Since $\Sigma \times H$ is a (countable) product of second-countable spaces, the Borel sigma-algebra generated by the product topology coincides with the (product) sigma-algebra generated by cylinders with

finitely many nontrivial[7] factors (see Kechris [18], p. 68.), so in particular $\mathcal{C}^\infty(A^n)$ is measurable. Also, $\mathcal{C}^\infty(A^n) \subset \mathcal{C}^\infty(B) \subset \mathcal{C}^\infty(C)$. Thus we can use Axiom 3 for μ and eq. (2) to show that $\mu^{n+1}(A^n|\mathcal{C}^n(B))\mu^{n+1}(\mathcal{C}^n(B)|\mathcal{C}^n(C)) = \mu^{n+1}(A^n|\mathcal{C}^n(C))$.

$\left(f(H_c) \subset \Delta^{\mathcal{B}}(\Sigma \times H)\right)$. Take $t = (\mu^1, ..., \mu^n, ...) \in H_c$ and let $\mu = f(t)$. We must verify that Axiom 3 holds for μ. Choose $B, C \in \mathcal{B}$ such that $B \subset C$ and $n \geq 0$. Consider a set $A^n \subset X^n$, measurable in the Borel sigma-algebra generated by the product topology on X^n. Applying Axiom 3 to μ^{n+1} for all $n = 0, 1, ...$, we obtain

$$\mu^{n+1}(A^n|\mathcal{C}^n(B))\mu^{n+1}(\mathcal{C}^n(B)|\mathcal{C}^n(C)) = \mu^{n+1}(A^n|\mathcal{C}^n(C)).$$

Then eq. (2) yields

$$\mu(\mathcal{C}^\infty(A^n)|\mathcal{C}^\infty(B))\mu(\mathcal{C}^\infty(B)|\mathcal{C}^\infty(C)) = \mu(\mathcal{C}^\infty(A^n)|\mathcal{C}^\infty(C))$$

This implies that μ satisfies Axiom 3 on the collection $\mathbf{C}^{<\infty}$ of cylinders, *i.e.,* Cartesian products of measurable sets of which at most finitely many are nontrivial. Again, since each factor space in the Cartesian product $\Sigma \times \Pi_{n\geq 0}\Delta^{\mathcal{B}}(X^n)$ is second-countable, $\mathbf{C}^{<\infty}$ generates $\mathcal{A}$, the sigma-algebra generated by the product topology.

Now let $\mathbf{B}(B, C) \subset \mathcal{A}$ be the collection of measurable sets for which Axiom 3 holds for fixed $B, C \in \mathcal{B}$. By sigma-additivity of μ, $\mathbf{B}(B, C)$ is a monotone class, and it contains the algebra $\mathbf{C}^{<\infty}$. Hence the smallest monotone class containing $\mathbf{C}^{<\infty}$ is also a sigma-algebra, which cannot be smaller than the sigma–algebra generated by $\mathbf{C}^{<\infty}$, *i.e.,* $\mathcal{A}$. Also, it must be contained in $\mathbf{B}(B, C)$, which completes the proof. ■

2.4 Common Certainty of Coherency

Even if i's hierarchy of CPSs t_i is coherent, some elements of $f(t_i)$ (*i.e.,* some $f_B(t_i)$, $B \in \mathcal{B}$) may assign positive probability to sets of incoherent hierarchies of the other individual j. We now consider the case in which there is common certainty of coherency conditional on every $B \in \mathcal{B}$. Observe that $\mathcal{B}$ is a collection of "external" events; conditioning on any $B \in \mathcal{B}$ does not restrict each individual's beliefs about each other's beliefs — only her beliefs about the prevailing external state. In particular, no event in $\mathcal{B}$ conveys information about an individual's coherency. It follows that there cannot be any inconsistency in assuming that there is common certainty of coherency conditional on any $B \in \mathcal{B}$: that is, we do not run the risk of formally requiring that an individual be (conditionally) certain of something that must necessarily be false, given the relevant conditional.

Formally, we shall say that individual i, endowed with a coherent hierarchy of CPSs t_i, is *certain* of some (measurable) event $E \subset \Sigma \times H$ *given* $B \in \mathcal{B}$ if

[7] That is, strictly included in the corresponding factor space.

$f_B(t_i)(E) = 1$. Common certainty of coherency given every $B \in \mathcal{B}$ can thus be inductively defined as follows:

$H_c^1 = H_c$,

for all $k \geq 2$,

$H_c^k = \{t \in H_c^{k-1} : \forall B \in \mathcal{B}, f_B(t)(\Sigma \times H_c^{k-1}) = 1\}$,

$T = \bigcap_{k \geq 1} H_c^k$.

$T \times T$ is the set of pairs of hierarchies satisfying common certainty of coherency conditional on every relevant hypothesis.

Proposition 2. *(cf. [10], Proposition 2) The restriction of $f = (f_B)_{B \in \mathcal{B}}$ to $T \subset H_c$ induces an homeomorphism $g = (g_B)_{B \in \mathcal{B}} : T \to \Delta^{\mathcal{B}}(\Sigma \times T)$ (defined by $g_B(t)(E) = f_B(t)(E)$ for all $B \in \mathcal{B}$, $t \in T$, $E \subset \Sigma \times T$ measurable.)*

Proof: First note that $T = \{t \in H_c : \forall B \in \mathcal{B}, f_B(t)(B \times T) = 1\}$. In fact, let $t \in H_c$ and suppose that, for all $B \in \mathcal{B}$, $f_B(t)(\Sigma \times T) = 1$. Then $t \in \bigcap_{k \geq 1} H_c^k = T$. Conversely, for each $t \in T$, $B \in \mathcal{B}$ and k, $f_B(t)(\Sigma \times H_c^k) = 1$. Since the measure $f_B(t)$ is sigma-additive

$$f_B(t)(\Sigma \times T) = f_B(t)\left(\Sigma \times \left(\bigcap_{k \geq 1} H_c^k\right)\right) = \lim_{k \to \infty} f_B(t)(\Sigma \times H_c^k) = 1.$$

It follows that $f(T) = \{\mu \in \Delta^{\mathcal{B}}(\Sigma \times H) : \forall B \in \mathcal{B}, \mu(B \times T | B \times H) = 1\}$, T is homeomorphic to $f(T)$, and each $f_B(T)$ is homeomorphic to $\Delta(B \times T)$. Given the definition of g in terms of f, one can check that for all $t \in T$, $g(t)$ satisfies Axioms 1, 2 and 3, and thus g is a homeomorphism between T and $\Delta^{\mathcal{B}}(\Sigma \times T)$. ■

Proposition 2 shows that each element $t \in T$ corresponds to an epistemic type in the usual sense, except that here a type is uniquely associated with a conditional probability system on $(\Sigma \times T, \mathcal{B})$ instead of a single probability measure on $\Sigma \times T$. Thus an epistemic type $t_i \in T$ represents the beliefs that individual i *would* have about the external state and about individual j's epistemic type conditional on every relevant hypothesis $B \in \mathcal{B}$.

The construction carried out above (in particular, Lemma 2) exploits the topological structure of the sets $X^0, X^1, \ldots, X^n, \ldots$. We conjecture that an alternative "topology-free" construction *à la* Heifetz and Samet [17] is possible in the present context. The resulting set of epistemic types $\tilde{T}$ could then be shown to be equivalent to the set of CPSs on $\Sigma \times \tilde{T}$, up to a measurable isomorphism (as opposed to "up to a homeomorphism.") However, the topological structure additionally enables one to associate closeness of epistemic types with closeness of beliefs — and conclude, for instance, that an individual's best reply correspondence is upper semi-continuous as a (composite) function of her type. It also enables one to conclude that, if a (coarse) subset of T approximates a finer subset of epistemic types, the same holds true for the corresponding subsets of beliefs (see, *e.g.,* Mertens and Zamir [19].)

Thus, introducing a topological structure in the analysis allows one to derive a richer theory. Moreover, as was argued above, it can be done without prejudice to the "natural" measure-theoretic structure on the space of beliefs.

3 Type Spaces

Each element $t = (\mu^1, \mu^2, \ldots)$ of the set T defined in the previous Section is by construction a complete list of an individual's hierarchical beliefs. That is, each $t \in T$ provides an "explicit" representation of the individual's epistemic type.

Alternatively, one may choose to start with an "implicit" representation, which closely mimics Harsanyi's original formulation of incomplete information games (see [15] and Mertens and Zamir [19]; for extensive games, see also Ben-Porath [7].)

Definition 2. *A* type space *on* $(\Sigma, \mathcal{B})$ *is a tuple* $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ *such that for each* $i = 1, 2$, T_i *is a Polish space and* g_i *is a continuous function*

$$g_i = (g_{i,B})_{B \in \mathcal{B}} : T_i \to \Delta^{\mathcal{B}}(\Sigma \times T_j),$$

where $i \neq j$.

There are obvious parallels between the definition of a type space and Proposition 2.[8]

Remark 1. *By Proposition 2, if we set* $T_1 = T_2 = T$ *and* $g_1 = g_2 = g$ *we obtain a (symmetric) type space which is denoted by* $\mathcal{T}^u$.

Moreover, given a type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ on $(\Sigma, \mathcal{B})$, it is possible to associate to every "implicit" description $t_i \in T_i$ an "explicit" hierarchy of beliefs, *i.e.,* a point in the set H constructed in the previous Section. A canonical procedure, which we presently illustrate, achieves this.

3.1 From Implicit to Explicit Representation

The following notation is essential. For any given measurable function $\varphi_{-i} : \Sigma \times T_j \to \Sigma \times T'_j$, let $\widehat{\varphi_{-i}} = (\widehat{\varphi_{-i,B}})_{B \in \mathcal{B}} : \Delta^{\mathcal{B}}(\Sigma \times T_j) \to \Delta^{\mathcal{B}}(\Sigma \times T'_j)$ be the corresponding function associating to each CPS μ_i on $(\Sigma \times T_j, \mathcal{B})$ the induced CPS $\mu'_i = \widehat{\varphi_{-i}}(\mu_i)$ on $(\Sigma \times T'_j, \mathcal{B})$. More specifically, for all $\mu_i \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$, $A' \subset \Sigma \times T'$ (measurable), $B \in \mathcal{B}$,

$$\widehat{\varphi_{-i,B}}(\mu_i)(A') = \mu_i((\varphi_{-i})^{-1}(A') | B \times T_j).$$

Our objective is to construct a pair of functions (φ_1, φ_2) associating to each type $\tau_i \in T_i$ a corresponding hierarchy of CPSs $t_i = \varphi_i(\tau_i) \in H$. The mappings $\varphi_i = \left(\varphi_i^1, \varphi_i^2, ...\right) = \left[(\varphi_{i,B}^1)_{B \in \mathcal{B}}, (\varphi_{i,B}^2)_{B \in \mathcal{B}}, ...\right]$, $i = 1, 2$ are obtained with a canonical inductive construction: the first order beliefs $\varphi_i^1(\tau_i)$ are derived by marginalization on Σ; the second order beliefs $\varphi_i^2(\tau_i)$ are obtained using g_j and φ_j^1, and so on. More precisely:

[8] Note however that the maps g_i in the definition of a type space are *not* required to be homeomorphisms.

- (1) For each $i = 1, 2$, $\tau_i \in T_i$, $B \in \mathcal{B}$,

$$\varphi^1_{i,B}(\tau_i) = mrg_\Sigma g_{i,B}(\tau_i).$$

For each $i, j = 1, 2$, $i \neq j$, $\tau_j \in T_j$, $\sigma \in \Sigma$,

$$\psi^1_{-i}(\sigma, \tau_j) = (\sigma, \varphi^1_j(\tau_j)),$$

that is, $\psi^1_{-i} = (Id_\Sigma, \varphi^1_j)$ (Id_Σ is the identity function on Σ.) Thus we have $\varphi^1_i : T_i \to \Delta^{\mathcal{B}}(X^0)$ and $\psi^1_{-i} : \Sigma \times T_j \to X^1$ (recall that $X^0 = \Sigma$ and $X^{n+1} = X^n \times \Delta^{\mathcal{B}}(X^n)$.)

- (n+1, n≥1) Let $\varphi^n_i : T_i \to \Delta^{\mathcal{B}}(X^{n-1})$ and $\psi^n_{-i} : \Sigma \times T_j \to X^n$ ($i, j = 1, 2$, $i \neq j$) be given. For each $i = 1, 2$, $\tau_i \in T_i$, $B \in \mathcal{B}$, $A^n \subset X^n$ (measurable),

$$\varphi^{n+1}_{i,B}(\tau_i)(A^n) = g_{i,B}(\tau_i)\left((\psi^n_{-i})^{-1}(A^n)\right),$$

that is, $\varphi^{n+1}_i = \widehat{\psi^n_{-i}} \circ g_i$. For each $i, j = 1, 2$, $i \neq j$, $\tau_j \in T_j$, $\sigma \in \Sigma$,

$$\psi^{n+1}_{-i}(\sigma, \tau_j) = \left(\psi^n_{-i}(\sigma, \tau_j), \varphi^{n+1}_j(\tau_j)\right),$$

that is, $\psi^{n+1}_{-i} = (\psi^n_{-i}, \varphi^{n+1}_j)$. Thus we have $\varphi^{n+1}_i : T_i \to \Delta^{\mathcal{B}}(X^n)$ and $\psi^{n+1}_{-i} : \Sigma \times T_j \to X^{n+1}$.

Note that $\psi^{n+1}_{-i}(\sigma, \tau_j) = \left(\sigma, \varphi^1_j(\tau_j), ..., \varphi^n_j(\tau_j), \varphi^{n+1}_j(\tau_j)\right)$.
This completes the inductive step.

3.2 Type-Morphisms and Universality

The preceding construction shows that, for any type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$, there exists a canonical embedding of each T_i in H. This subsection addresses the question whether the sets T_i can actually be embedded in T, the collection of infinite hierarchies of beliefs satisfying coherency and common certainty of coherency conditional on every hypothesis. If this is the case, then any type space may essentially be regarded as a (belief-closed) subspace of the symmetric type space $\mathcal{T}^u$.

In order to formalize these ideas, we need to develop an adequate notion of embedding for type spaces. The central ingredient is again the map $\widehat{\varphi_{-i}} : \Delta^{\mathcal{B}}(\Sigma \times T_j) \to \Delta^{\mathcal{B}}(\Sigma \times T'_j)$ induced by a continuous function $\varphi_{-i} : \Sigma \times T_j \to \Sigma \times T'_j$, where T_j and T'_j are sets of epistemic types.

Definition 3. *Let $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ and $\mathcal{T}' = (\Sigma, \mathcal{B}, T'_1, T'_2, g'_1, g'_2)$ be two type spaces on $(\Sigma, \mathcal{B})$. A* type-morphism *from $\mathcal{T}$ to $\mathcal{T}'$ is a triple of functions $\varphi = (\varphi_0, \varphi_1, \varphi_2)$ whereby φ_0 is the identity function on Σ and for each $i = 1, 2$, $\varphi_i : T_i \to T'_i$ is a continuous function such that*

$$g'_i \circ \varphi_i = \widehat{\varphi_{-i}} \circ g_i$$

(where $\varphi_{-i} = (\varphi_0, \varphi_j) : \Sigma \times T_j \to \Sigma \times T'_j$.) If φ is a homeomorphism between $\Sigma \times T_1 \times T_2$ and $\Sigma \times T'_1 \times T'_2$, then we say that $\mathcal{T}$ and $\mathcal{T}'$ are isomorphic.

The intuition is as follows. Fix a type $t_i \in T_i$; the function φ_i maps t_i to some $t'_i \in T'_i$, and $g'_i(t'_i)$ then retrieves a CPS μ' on $\Sigma \times T'_j$. Alternatively, one can use the function g_i to obtain from t_i a CPS ν on $\Sigma \times T_j$, then $\widehat{\varphi_{-i}}$ to map ν to a CPS ν' on $\Sigma \times T'_j$. Intuitively, μ' and ν' should coincide, because both originate from the same epistemic type $t_i \in T_i$. Equivalently, the embedding $\varphi_i : T_i \to T'_i$ and the (derived) embedding $\widehat{\varphi_{-i}} : \Delta^{\mathcal{B}}(\Sigma \times T_j) \to \Delta^{\mathcal{B}}(\Sigma \times T'_j)$ should be *consistent* with each other. This is precisely what the above definition requires.

Type-morphisms satisfy an intuitively appealing closure property:

Remark 2. *Let let $\varphi = (\varphi_0, \varphi_1, \varphi_2)$ be a type-morphism between the type spaces $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ and $\mathcal{T}' = (\Sigma, \mathcal{B}, T'_1, T'_2, g'_1, g'_2)$ on $(\Sigma, \mathcal{B})$. Then*

$$\forall i = 1, 2, \forall \tau'_i \in \varphi_i(T_i), \forall B \in \mathcal{B} : \quad g'_{i,B}(\tau'_i)(\Sigma \times \varphi_j(T_j)) = 1$$

That is, $\Sigma \times \varphi_1(T_1) \times \varphi_2(T_2)$ is a belief-closed *subset of $\Sigma \times T'_1 \times T'_2$.*

This property is consistent with the proposed interpretation of type-morphisms as a way to view one type space as a subset of another (up to renaming and deletion of redundant types.)

Another useful (and natural) property of type-morphisms follows.

Remark 3. *Suppose φ is a type-morphism from $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ to $\mathcal{T}' = (\Sigma, \mathcal{B}, T'_1, T'_2, g'_1, g'_2)$ let $E \subset \Sigma \times T_1 \times T_2$ and $E' \subset \Sigma \times T'_1 \times T'_2$ be measurable subsets such that $\varphi(E) \subset E'$. Then for all $i \in \{1, 2\}$, $\tau_i \in T_i$, $B \in \mathcal{B}$,*

$$g_{i,B}(\tau_i)\left(\{(\sigma, \tau_j) : (\sigma, \tau_i, \tau_j) \in E\}\right) \leq$$

$$g'_{i,B}(\varphi(\tau_i))\left(\left\{(\sigma, \tau'_j) : \tau'_j = \varphi_j(\tau_j), \varphi(\sigma, \tau_i, \tau_j) \in E'\right\}\right).$$

We are finally able to tackle the issue of "universality".[9]

The formal definition of this property should be by now entirely transparent:

Definition 4. *A type space $\mathcal{T}'$ on $(\Sigma, \mathcal{B})$ is* universal *if for every other type space $\mathcal{T}$ on $(\Sigma, \mathcal{B})$ there is unique type-morphism from $\mathcal{T}$ to $\mathcal{T}'$.*

Remark 4. *Any two universal type spaces are isomorphic.*

We are ready to state the main result of this Section.

Proposition 3. *Let $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ be an arbitrary type space on $(\Sigma, \mathcal{B})$ and, for $i = 1, 2$, let $\varphi_i : T_i \to H$ be the functions defined in Subsection 3.1. Then, for each $i = 1, 2$, $\varphi_i(T_i) \subset T$ and $\varphi = (Id_\Sigma, \varphi_1, \varphi_2)$ is the unique type-morphism from $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ to $\mathcal{T}^u = (\Sigma, \mathcal{B}, T, T, g, g)$. Thus $\mathcal{T}^u$ is the unique universal type space (up to isomorphisms.)*

Proof: See the Appendix.

[9] See Mertens and Zamir [19] and Heifetz and Samet [17, 16]. Heifetz and Samet show that, if we drop the topological structure, the space of hierarchies of beliefs (satisfying coherency and common certainty of coherency) is "larger" than the set of hierarchies generated by some type space. The latter is a universal type space.

3.3 Independence

As was suggested above, the set Σ represents a collection of possible external states which are relevant to the individuals' decision problems. Apart from certain topological properties, the construction of the universal type space $\mathcal{T}^u$ and the definition of a type space do not require that Σ exhibit any particular structure. However, in game-theoretic applications, Σ is the Cartesian product of the two players' strategy spaces[10]. Thus, we may wish to require that an individual's conditional beliefs satisfy a (weak) form of independence: informally, her beliefs about her *own* strategy should be separable from her beliefs about *her opponent's* strategy and epistemic type.

In general, suppose that $\Sigma = \Sigma_1 \times \Sigma_2$, where each Σ_i, $i = 1, 2$, is a Polish space. Derive from $\mathcal{B}$ two collections $\mathcal{B}_1$, $\mathcal{B}_2$ of *marginal* conditioning events as follows:

$$\mathcal{B}_1 = \{B_1 \subset \Sigma_1 : \exists B_2 \subset \Sigma_2, \exists B \in \mathcal{B}, B = B_1 \times B_2\},$$

$\mathcal{B}_2$ is similarly defined. Note that each $\mathcal{B}_i$ is a finite or countable collection of subsets which are both closed and open. Finally, suppose that

$$\mathcal{B} \subset \{B \subset \Sigma \ : \ \exists B_1 \in \mathcal{B}_1, \ B_2 \in \mathcal{B}_2 \text{ such that } B = B_1 \times B_2\} \tag{3}$$

For any $i = 1, 2$, the set of CPSs on Σ_i and $\Sigma_i \times T_i$ will be denoted by $\Delta^{\mathcal{B}_i}(\Sigma_i)$ and $\Delta^{\mathcal{B}_i}(\Sigma_i \times T_i)$ respectively.

Definition 5. *Fix a type space $\mathcal{T} = (\Sigma = \Sigma_1 \times \Sigma_2, \mathcal{B}, T_1, T_2, g_1, g_2)$ satisfying 3. Player i's CPS $\mu_i \in \Delta^{\mathcal{B}}(\Sigma_1 \times \Sigma_2 \times T_2)$ has the* independence property *if there are two CPSs $\mu_{ii} \in \Delta^{\mathcal{B}_i}(\Sigma_i)$ and $\mu_{ij} \in \Delta^{\mathcal{B}_j}(\Sigma_j \times T_j)$ such that for all $B = B_1 \times B_2 \in \mathcal{B}$,*

$$\mu_i(\cdot|B_1 \times B_2 \times T_j) = \mu_{ii}(\cdot|B_i) \otimes \mu_{ij}(\cdot|B_j \times T_j),$$

where $\otimes$ denotes the product of measures.

The set of CPSs for player i with the independence property is denoted by $I\Delta^{\mathcal{B}}(\Sigma_i, \Sigma_j \times T_j)$. Similarly, the set of CPSs on Σ satisfying the independence property is denoted $I\Delta^{\mathcal{B}}(\Sigma_i, \Sigma_j)$. Note that for all $\mu_i \in I\Delta^{\mathcal{B}}(\Sigma_i, \Sigma_j \times T_j)$ the CPSs μ_{ii} and μ_{ij} mentioned in Definition 5 are uniquely determined. We call μ_{ii} and μ_{ij} the *marginals* of μ_i on Σ_i and $\Sigma_j \times T_j$ respectively.

As one should expect, type-morphisms preserve the independence property:

Lemma 3. *Suppose that $\Sigma = \Sigma_1 \times \Sigma_2$ and $\mathcal{B}$ satisfies 3. Fix two type spaces $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ and $\mathcal{T}' = (\Sigma, \mathcal{B}, T_1', T_2', g_1', g_2')$ on $(\Sigma, \mathcal{B})$ and a type-morphism $\varphi = (\varphi_0, \varphi_1, \varphi_2)$ between $\mathcal{T}$ and $\mathcal{T}'$. For all $i = 1, 2$, $t_i \in T_i$,*

$$g_i(t_i) \in I\Delta^{\mathcal{B}}(\Sigma_i, \Sigma_j \times T_j) \quad \Rightarrow \quad g_i'(\varphi_i(t_i)) \in I\Delta^{\mathcal{B}}(\Sigma_i, \Sigma_j \times T_j')$$

Proof: Omitted.

[10] Or, for incomplete-information games, the Cartesian product of the players' sets of strategy — payoff type pairs: see Section 5 for details.

4 Conditional Belief Operators

Fix an arbitrary type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$. A point $(\sigma, \tau_1, \tau_2) \in \Sigma \times T_1 \times T_2$ comprises a description of the external state σ, and (perhaps via the canonical maps $\varphi_i : T_i \to T$) a complete list of both individuals' hierarchical beliefs. Thus, we refer to any such point as a *state of the world*; similarly, measurable sets $E \subset \Sigma \times T_1 \times T_2$ will be called *events*.

The next order of business is to define the notions of *probability p belief* and *certainty* (*i.e.,* probability one belief.)

For each $\tau_i \in T_i$, $E_{\tau_i} \subset \Sigma \times T_j$ denotes the set of pairs (σ, τ_j) consistent with the event E and the epistemic type τ_i ($E_{\tau_1} = \{(\sigma, \tau_2) \in \Sigma \times T_2 : (\sigma, \tau_1, \tau_2) \in E\}$, E_{τ_2} is similarly defined.) Type τ_i assigns to E a probability of at least p conditional on each hypothesis $B \in \mathcal{F} \subset \mathcal{B}$ if $\forall B \in \mathcal{F}$, $g_{i,B}(\tau_i)(E_{\tau_i}) \geq p$. Note that we are implicitly assuming that i is certain of her epistemic type. For every $E \subset \Sigma \times T_1 \times T_2$ and collection of relevant hypotheses $\emptyset \notin \mathcal{F} \subset \mathcal{B}$, the event "$i$ would be *certain* of E conditional on every $B \in \mathcal{F}$" is

$$\beta_{i,\mathcal{F}}(E) := \{(\sigma, \tau_1, \tau_2) : \forall B \in \mathcal{F}, g_{i,B}(\tau_i)(E_{\tau_i}) = 1\}$$

(note that $\beta_{i,\mathcal{F}}(E)$ is measurable for each (measurable) E.) If $\mathcal{F}$ is a singleton, we replace it with its unique element as a subscript. If we have to emphasize the type space $\mathcal{T}$, we add $\mathcal{T}$ as a subscript to the belief operators, that is, we write $\beta_{i,\mathcal{F},\mathcal{T}}(E)$.

It is easily shown that each $\beta_{i,\mathcal{F}}$ has all the standard properties of belief operators.[11] In particular, each $\beta_{i,\mathcal{F}}$ satisfies:

- *Monotonicity*: $E \subset F$ implies $\beta_{i,\mathcal{F}}(E) \subset \beta_{i,\mathcal{F}}(F)$,
- *Conjunction*: $\beta_{i,\mathcal{F}}(E \cap F) = \beta_{i,\mathcal{F}}(E) \cap \beta_{i,\mathcal{F}}(F)$.

In the following, we will often consider events pertaining to the realization of the external state and the individuals' *first-order* beliefs about Σ; we presently develop the required notation and note a related property of type morphisms.

Let $E \subset \Sigma \times \Delta^{\mathcal{B}}(\Sigma) \times \Delta^{\mathcal{B}}(\Sigma)$ be measurable. The event corresponding to subset E in type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ is denoted $E_{\mathcal{T}}$, *i.e.,*

$$E_{\mathcal{T}} := \left\{(\sigma, \tau_1, \tau_2) : \left(\sigma, \left(mrg_{\Sigma} g_{1,B}(\tau_1)\right)_{B \in \mathcal{B}}, \left(mrg_{\Sigma} g_{2,B}(\tau_2)\right)_{B \in \mathcal{B}}\right) \in E\right\}.$$

The following lemma states that if there is a type morphism φ from $\mathcal{T}$ to $\mathcal{T}'$ and at some state (σ, τ_1, τ_2) of $\mathcal{T}$ player i would believe $E_{\mathcal{T}}$ conditional on each $B \in \mathcal{F}$, then at the corresponding state $(\sigma, \varphi_1(\tau_1), \varphi_2(\tau_2))$ in $\mathcal{T}'$ player i would believe $E_{\mathcal{T}'}$ conditional on each $B \in \mathcal{F}$. By induction, the result also holds for higher-order beliefs about $E_{\mathcal{T}'}$.

[11] See, for example, axioms K2-K6 in Osborne and Rubinstein [21], pp 69-70. Axiom K1 obviously does not hold because $\beta_{i,\mathcal{F}}$ is not a knowledge operator. K1 is replaced by the weaker axiom that player i does not have contradictory conditional beliefs: $\beta_{i,\mathcal{F}}(\emptyset) = \emptyset$.

Lemma 4. *Suppose that φ is a type morphism from $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ to $\mathcal{T}' = (\Sigma, \mathcal{B}, T_1', T_2', g_1', g_2')$ and let $E \subset \Sigma \times \Delta^{\mathcal{B}}(\Sigma) \times \Delta^{\mathcal{B}}(\Sigma)$ be measurable. Then*

$$\varphi(E_{\mathcal{T}}) \subset E_{\mathcal{T}'}$$

and for all integers $n \geq 1$, for all collections $\{i_1, ..., i_n\}$ and $\{\mathcal{F}_1, ..., \mathcal{F}_n\}$ with $i_k \in \{1, 2\}$ and $\emptyset \neq \mathcal{F}_k \subset \mathcal{B}$ for all $k = 1, \ldots, n$,

$$\varphi\left((\beta_{i_1,\mathcal{F}_1,\mathcal{T}} \circ ... \circ \beta_{i_n,\mathcal{F}_n,\mathcal{T}})(E_{\mathcal{T}})\right) \subset (\beta_{i_1,\mathcal{F}_1,\mathcal{T}'} \circ ... \circ \beta_{i_n,\mathcal{F}_n,\mathcal{T}'})(E_{\mathcal{T}'}).$$

Proof: Since φ is a type-morphism from $\mathcal{T}$ to $\mathcal{T}'$, $mrg_{\Sigma} g_{i,B}(\tau_i) = mrg_{\Sigma} g'_{i,B}(\varphi_i(\tau_i))$ for all i, τ_i, B. This implies $\varphi(E_{\mathcal{T}}) \subset E_{\mathcal{T}'}$. Thus the first statement is true. Remark 3 implies that the second statement is true for $n = 1$. An obvious induction argument (again using Remark 3) implies that the second statement is true for all n. ■

5 Interactive Epistemology and Rationality in Dynamic Games

We now apply the foregoing analysis to the theory of dynamic games. For the sake of simplicity we only consider *finite* games with *observed actions*. On the other hand, we allow for incomplete information because this does not alter the analysis in any significant way.

5.1 Games of Incomplete Information with Observed Actions

Consider a finite, two-person, multistage game with observed actions and incomplete information (see, *e.g.,* [13], Chapter 8, or [21], Chapter 12) without the probabilistic structure. Let Θ_i the set of payoff-relevant types for player i. A payoff-relevant type $\theta_i \in \Theta_i$ corresponds to i's private information about payoff-relevant aspects of the game and has to be distinguished from the epistemic type which specifies i's attitudes to have certain conditional beliefs given certain events. Players' beliefs about the opponent's payoff-relevant type will be specified within an epistemic model. We will omit the adjective "payoff-relevant" whenever no confusion can arise. $\mathcal{H}$ denotes the set of partial histories, which includes the *empty history* ϕ, and $\mathcal{Z}$ denotes the set of terminal histories. The set of strategies for player i (functions from $\mathcal{H}$ to feasible actions) is denoted S_i. Player i preferences over lotteries are represented by a VNM utility function $u_i : \mathcal{Z} \times \Theta_1 \times \Theta_2 \to \mathbb{R}$. Static games, games of complete information and games of perfect information are included in this class of games as special cases.[12]

[12] A game is *static* if $\mathcal{H} = \{\phi\}$, has *complete information* if $\Theta_1 \times \Theta_2$ is a singleton, and has *perfect information* if for each $h \in \mathcal{H}$ either player 1 or player 2 has only one feasible action. We are assuming that the set of feasible actions of each player may be history-dependent,

The basic elements of our analysis are strategy-type pairs $(s_i, \theta_i) \in S_i \times \Theta_i$, $i = 1, 2$. A generic pair for player i is denoted σ_i and the set of such feasible pairs is $\Sigma_i := S_i \times \Theta_i$. The external state space is $\Sigma := \Sigma_1 \times \Sigma_2$ with generic element $\sigma = (\sigma_1, \sigma_2) = (s_1, \theta_1, s_2, \theta_2)$. When there is complete information each Θ_i is a singleton and Σ simply represents set of strategy pairs. For each history h, $S_i(h)$ denotes the set of player i's strategies consistent with h, $\Sigma_i(h) = S_i(h) \times \Theta_i$ is the set of σ_i consistent with h, and $\Sigma(h) = \Sigma_1(h) \times \Sigma_2(h)$ is the set of external states inducing h. $\mathcal{H}(s_i)$ is the set of partial histories consistent with s_i, that is, $\mathcal{H}(s_i) := \{h \in \mathcal{H} : s_i \in S_i(h)\}$. For every partial history h, $\Sigma_j(h)$ is a strategic form representation of i's information about j at h. We can obtain a strategic form payoff function $U_i : \Sigma \to \mathbb{R}$ in the usual way: for all $(z, \theta_1, \theta_2) \in \mathcal{Z} \times \Theta_1 \times \Theta_2$ and $(s_1, \theta_1, s_2, \theta_2) \in \Sigma(z)$, $U_i(s_1, \theta_1, s_2, \theta_2) = u_i(z, \theta_1, \theta_2)$.

To illustrate our game-theoretic notation, consider the signalling game depicted in Figure 1. We have $\Theta_1 = \{\theta_1', \theta_1''\}$, $\Theta_2 = \{\theta_2\}$ is a singleton, thus the set of pairs of types is $\Theta_1 \times \Theta_2 = \{\theta', \theta''\}$. The set of partial histories is $\mathcal{H} = \{\phi, (R)\}$ and the set of outcomes is $\{(L), (R, u), (R, d)\} \times \{\theta', \theta''\}$. The set of external states is $(S_1 \times \Theta_1) \times (S_2 \times \Theta_1)$, where $S_1 = \{L, R\}$ and $S_2 = \{\mathbf{u}, \mathbf{d}\}$ ($\mathbf{a}$ means "choose action a if R is observed," $a \in \{u, d\}$.) The "strategic representation" of partial history (R) is $\Sigma(R) = \{(R, \theta_1'), (R, \theta_1'')\} \times \{(\mathbf{u}, \theta_2), (\mathbf{d}, \theta_2)\}$. To draw the picture we rely on the fact that the set of triples $(h, \theta_1, \theta_2) \in (\mathcal{H} \cup \mathcal{Z}) \times \Theta_1 \times \Theta_2$ can be regarded as an arborescence with initial nodes $(\phi, \theta_1, \theta_2) \in \{\phi\} \times \Theta_1 \times \Theta_2$ and terminal nodes $(z, \theta_1, \theta_2) \in \mathcal{Z} \times \Theta_1 \times \Theta_2$.[13] For each type θ_i and partial history h, $\{h\} \times \{\theta_i\} \times \Theta_i$ corresponds to an information set for player i in the graphical representation. For example, $\{(\theta_1', \theta_2, (R)), (\theta_1'', \theta_2, (R))\}$ corresponds to the information set for player 2 depicted in Figure 1.

We are interested in players' (mutual) conditional beliefs at each (commonly observable) partial history h. Thus the collection of relevant hypotheses in this context is $\mathcal{B} = \{B : \exists h \in \mathcal{H}, B = \Sigma(h)\}$. Note that $\Sigma \in \mathcal{B}$, because $\Sigma = \Sigma(\phi)$, where ϕ is the *empty history*. In order to complete the model we have to introduce a(n) (epistemic) type-space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$. A complete type for player i is a pair $(\theta_i, \tau_i) \in \Theta_i \times T_i$ corresponding to a vector $(\theta_i, g_i(\tau_i)) \in \Theta_i \times \Delta^{\mathcal{B}}(\Sigma \times T_j)$.[14] This description of an interactive epistemic model based on a dynamic game is consistent with several papers about the theory of extensive form games. In particular, it can be regarded as a generalization of the epistemic model put forward by Ben Porath [7].

Since each element of $\mathcal{B}$ represents the event that some history h occurs, we

but not type-dependent. The extension to the case of type-dependent feasibility constraints is conceptually straightforward, but requires a more complex notation.

[13] The precedence relation is: (θ_1, θ_2, h) precedes $(\theta_1', \theta_2', h')$ if and only if $(\theta_1, \theta_2) = (\theta_1', \theta_2')$ and h is a prefix (initial subhistory) of h'. Clearly, to obtain the standard graph-theoretic representation simultaneous moves have to be ordered in some arbitrary way adding information sets appropriately.

[14] In static games $\Theta_i \times T_i$ corresponds to the set of types in the sense of Harsanyi [15]. In most applications of the theory of games with incomplete information Θ_i is assumed to coincide with T_i and the functions g_i, $i = 1, 2$, are derived from a common prior on $\Theta_1 \times \Theta_2$ and a Bayesian equilibrium profile.

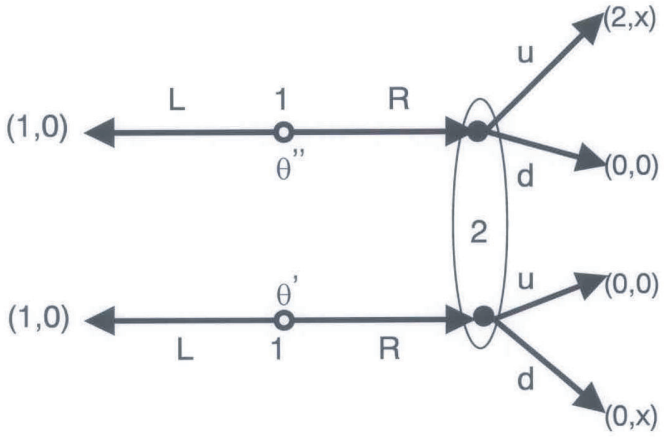

Figure 1

simplify our notation for CPSs on Σ or $\Sigma \times T_i$ $(i = 1, 2)$ and replace $\mathcal{B}$ with $\mathcal{H}$. Indeed, we shall denote strategic form events $B = \Sigma(h) \in \mathcal{B}$ by $h \in \mathcal{H}$ whenever needed (and in particular, in subscripts denoting conditioning events.)

Note that $\mathcal{B}$ satisfies the product condition 3 of Section 3: since $\Sigma(h) = \Sigma_1(h) \times \Sigma_2(h)$ for all $h \in \mathcal{H}$, we have

$$\mathcal{B} \subset \{B \subset \Sigma : \exists B_1 \in \mathcal{B}_1, \exists B_2 \in \mathcal{B}_2, B = B_1 \times B_2\},$$

where

$$\mathcal{B}_i = \{\Sigma_i(h) \ : \ h \in \mathcal{H}\} \quad i = 1, 2.$$

For notational simplicity, we shall write $\Delta^{\mathcal{B}_i}(\Sigma_i)$ and $\Delta^{\mathcal{B}_i}(\Sigma_i \times T_i)$ as $\Delta^{\mathcal{H}}(\Sigma_i)$ and $\Delta^{\mathcal{H}}(\Sigma_i \times T_i)$ for $i = 1, 2$, and no confusion will arise.

Finally, we continue to identify singletons with their unique elements. For example, given $h \in \mathcal{H}$ or $\mathcal{F} \subset \mathcal{H}$, we write $(g_{i,h}(\tau_i))_{h \in \mathcal{H}} \in \Delta^{\mathcal{H}}(\Sigma \times T_j)$ and $\beta_{i,\mathcal{F}}(E)$ instead of $(g_{i,\Sigma(h)}(\tau_i))_{\Sigma(h) \in \mathcal{B}} \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$ and $\beta_{i,\{\Sigma(h):h \in \mathcal{F}\}}(E)$.

We are formally assuming that a player has beliefs about her own strategy and payoff-relevant type. Considering a player's beliefs about her own strategy is germane to extensive from analysis, because the choice of player i at a given history is motivated by her beliefs about the opponents' and *her own* behavior later on in the game. However, it is natural to focus on player i's beliefs about the opponent. We assume that a rational player i is certain of her strategy and type and that she takes a best response against her beliefs about the opponent. These beliefs are represented by a conditional probability system $\mu_{ij} \in \Delta^{\mathcal{H}}(\Sigma_j \times T_j)$ with corresponding first order beliefs $(mrg_{\Sigma_j}\mu_{ij,h})_{h \in \mathcal{H}} \in \Delta^{\mathcal{H}}(\Sigma_j)$.

Definition 6. *Let* $(s_i, \theta_i) \in \Sigma_i$, $\mu = (\mu(\cdot|\Sigma_j(h))_{h\in\mathcal{H}} \in \Delta^{\mathcal{H}}(\Sigma_j)$. *Strategy* s_i *is a* best response *to* μ *for type* θ_i, *written* $(s_i, \theta_i) \in r_i(\mu)$, *if for all* $h \in \mathcal{H}(s_i)$, $s'_i \in S_i(h)$

$$\sum_{\sigma_j \in \Sigma_j(h)} [U_i(s_i, \theta_i, \sigma_j) - U_i(s'_i, \theta_i, \sigma_j)]\, \mu(\sigma_j|\Sigma_j(h)) \geq 0.$$

Note that this is a best response property for plans of actions,[15] as maximization is required only at histories consistent with the given strategy (cf. Reny [23].) A standard dynamic programming argument shows for every $\mu \in \Delta^{\mathcal{H}}(\Sigma_j)$, $r_i(\mu) \neq \emptyset$.

Definition 7. *Fix a type space* $\mathcal{T} = (\Sigma, \mathcal{H}, T_1, T_2, g_1, g_2)$. *Player* i *is* rational *at state* $(s_i, \theta_i, \sigma_j, \tau_i, \tau_j)$ *in* $\mathcal{T}$ *if*
(1) epistemic type τ_i *is always certain of* θ_i *and is certain of* s_i *whenever possible, that is, for all* $h \in \mathcal{H}$, $g_{i,h}(\tau_i)\,(S_i \times \{\theta_i\} \times \Sigma_j \times T_j) = 1$ *and if* $s_i \in S_i(h)$, $g_{i,h}(\tau_i)\,(\{(s_i, \theta_i)\} \times \Sigma_j \times T_j) = 1$,
(2) $g_i(\tau_i) \in I\Delta^{\mathcal{H}}(\Sigma_i, \Sigma_j \times T_j)$ *(see Definition 5 in Section 3),*
(3) $(s_i, \theta_i) \in r_i\left((mrg_{\Sigma_j} g_{i,h}(\tau_i))_{h\in\mathcal{H}}\right)$.

Condition (1) says that a rational player knows her type and chooses actions according to a specific plan she intends to implement. In most epistemic models for games a property like (1) is assumed to hold globally, while we only require that it holds at states where player i is rational. Condition (2) says that the beliefs of a rational player can be decomposed into marginal beliefs about herself and about the opponent. In a static model (2) is implied by (1), but in a dynamic model player i might change her beliefs about the opponents simply because she deviated from her own plan. A condition similar to (2) is assumed explicitly in Reny [24, 25] and implicitly in Ben Porath [7].[16] Note that the natural extension of (2) to an n-person game would *not* require that player i' beliefs be uncorrelated. In fact, the marginal $\mu_{i,-i} \in \Delta^{\mathcal{H}}(\Sigma_{-i} \times T_{-i})$ might exhibit correlation. Independence of beliefs about the opponents should be studied as a separate assumption (see [6].)

In the following, we shall discuss a number of (finite) epistemic models for games. Our analysis will focus on states in which Conditions (1) and (2) above hold and there is common certainty of this fact. This allows us to represent a (finite) type space in a compact tabular form. Consider for example the following table, which refers to the game in Figure 1 above; we use the notation $g_{ij,h}(\tau_i) = mrg_{\Sigma_j \times T_j} g_{i,h}(\tau_i)$.

[15] Two strategies s_i and s'_i are *realization equivalent* if $\mathcal{H}(s_i) = \mathcal{H}(s'_i)$ and $s_i(h) = s'_i(h)$ for all $h \in \mathcal{H}(s_i)$. A *plan of action* is a maximal set of realization equivalent strategies.

[16] Condition (2) is not really essential for our analysis and in the previous version of this paper (2) was not used. In fact, it can be shown that for every state (σ, t_i, t_j) in the universal type space where (1) and (3) are satisfied, there is a state (σ, t'_i, t_j) where (1), (2) and (3) are satisfied and the beliefs of t'_i about j coincide with the beliefs of t_i at each h consistent with σ_i. However, assuming (2) facilitates the comparison with the literature.

(σ_1, τ_1)	$g_{12,\phi}(\tau_1)$	$g_{12,(R)}(\tau_1)$
$((L, \theta'), \tau_1^1)$	1,0	1,0
$((L, \theta''), \tau_1^2)$	0,1	0,1
$((R, \theta'), \tau_1^3)$	1,0	1,0
$((R, \theta''), \tau_1^4)$	1,0	1,0

(σ_2, τ_2)	$g_{21,\phi}(\tau_2)$	$g_{21,(R)}(\tau_2)$
(u, τ_2^1)	p, 0, 0, $1-p$	0, 0, 0, 1
(d, τ_2^2)	q, $1-q$, 0, 0	0, 0, 1, 0

Table 1: A type space for the game in Figure 1

Although Table 1 does not display an exhaustive list of states and shows only the marginal beliefs about the opponent, it contains all the essential information. A type space corresponding to a table like Table 1 can be constructed according to the following conventions:

- (i) For each player i, T_i is the set of epistemic types τ_i^k listed in the table. A similar convention holds for Θ_i (θ_i is omitted if Θ_i is a singleton.) But not all strategies need be listed in the table (see Table 2 below.)
- (ii) For each row k of player i in the table, the pair $(s_i^k, \theta_i^k, \tau_i^k)$ satisfies Conditions (1) and (2) in Definition 7. This completely determines the conditional beliefs $g_{i,h}(\tau_i^k)$ at histories $h \in \mathcal{H}(s_i)$. Otherwise, the CPS $g_i(\tau_i^k)$ is completed so as to satisfy (1) and (2).
- (iii) The set of states is completed by taking, for each i, all the combinations $(s_i, \theta_i, \tau_i) \in S_i \times \Theta_i \times T_i$. (Note that, by convention (i), all the states not listed in the table violate (1).)
- (iv) Only the probabilities of (coordinates of) states listed in the table are shown (the kth number is the probability of opponent's row k.) The probabilities of other states are always zero. Thus the set of states listed in the table is a belief-closed subset of $\Sigma \times T_1 \times T_2$.

¿From a substantive viewpoint, the following remarks are in order:

- Player 1's beliefs about her opponent are the same at the beginning of the game and after the history (R). This is a consequence of the independence assumption, together with Bayes' rule.
- Choosing R is strictly dominated for Player 1's payoff-type θ'. Hence, at any state $((R, \theta'), \sigma_2, \tau_1^3, \tau_2)$, for any $\sigma_2 \in \{u, d\}$ and $\tau_2 \in \{\tau_2^1, \tau_2^2\}$, Player 1 is not rational.
- If $x > 0$, type τ_2^1 (resp. τ_2^2) of Player 2 justifies choice u (resp. d.) Therefore, in any of the explicitly described states, Player 1 is certain of Player 2's rationality.

Further comments on this game will be provided below.

The set of states in $\mathcal{T}$ where player i is rational is denoted $R_{i,\mathcal{T}}$. The event that every player is rational is $R_{\mathcal{T}} = R_{1,\mathcal{T}} \cap R_{2,\mathcal{T}}$. Whenever no confusion arises, we drop the reference to the given type space in our notation and simply write R_i for the event "player i is rational" and $\beta_{i,\mathcal{F}}(E)$ for the event "player i is certain of E conditional on each $h \in \mathcal{F}$."

5.2 Common Certainty of the Opponent's Rationality

We are interested in the following question (among others): "What might player i do if (1) she is rational and (2) for all $h \in \mathcal{F}$, she believes that her opponent is rational, (3) for all $h \in \mathcal{F}$, she believes that, for all $h' \in \mathcal{F}$, her opponent believes that she is rational, (4) ...?" In other words we ask for the consequences of rationality and common certainty of the *opponent*'s rationality conditional on a given collection of histories (cf. Reny [24].)

Formally, the statement "There is common certainty of the opponent's rationality given $\mathcal{F}$ from the point of view of player i" corresponds to the following event:

$$\begin{aligned} CCOR_{i,\mathcal{F}} &:= \beta_{i,\mathcal{F}}(R_j) \cap \beta_{i,\mathcal{F}}(\beta_{j,\mathcal{F}}(R_i)) \cap \beta_{i,\mathcal{F}}(\beta_{j,\mathcal{F}}(\beta_{i,\mathcal{F}}(R_j))) \cap ... \\ &= \bigcap_{n \geq 1} \left(\beta_{i_1,\mathcal{F}} \circ ... \circ \beta_{i_n,\mathcal{F}}\right)(R_{i_{n+1}}),\ i_1 = i,\ i_{k+1} \neq i_k. \end{aligned}$$

Hence, the statement "There is common certainty of the opponent's rationality given $\mathcal{F}$" corresponds to the event

$$CCOR_{\mathcal{F}} := CCOR_{1,\mathcal{F}} \cap CCOR_{2,\mathcal{F}}$$

Finally, let $R := R_1 \cap R_2$.

Definition 8. *We say that σ is* consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$ *if there are a type space $\mathcal{T}$ and a pair of types (τ_1, τ_2) such that*

$$(\sigma, \tau_1, \tau_2) \in R \cap CCOR_{\mathcal{F}}$$

If $\mathcal{F}$ is a singleton ($\mathcal{F} = \{h\}$ for some h) we obtain a notion of common certainty of the opponent's rationality at a given history (cf. Reny [25].) In particular, we may be interested in the consequences of common certainty of the opponent's rationality at the beginning of the game, that is, given the empty history $h = \phi$ (cf. Ben Porath [7].)[17]

[17] In games with observed actions, a relevant hypothesis $B = \Sigma(h)$ represents an event that becomes common knowledge when history h occurs. Hence, an event such as $CCOR_h \cap (\Sigma(h) \times T_1 \times T_2)$ may be interpreted as saying that history h occurs, and as soon as this becomes common knowledge, there is common certainty of the opponent's rationality.

Information sets in general extensive games do not correspond to common knowledge events. This is not problematic for certain applications (see, *e.g.,* Battigalli and Siniscalchi [5].) However, Battigalli and Bonanno [4]) show how to enrich the conventional formalization of general extensive games to fully describe the players' information at every node.

There is a compact and convenient way to express event $R_i \cap CCOR_{i,\mathcal{F}}$. Let $R^1_{i,\mathcal{F}} := R_i$, $i = 1, 2$. For all $n > 1$, $i \neq j$, define

$$R^n_{i,\mathcal{F}} = R_i \cap \beta_{i,\mathcal{F}}(R^{n-1}_{j,\mathcal{F}}).$$

Clearly,

$$R^2_{i,\mathcal{F}} = R_i \cap \beta_{i,\mathcal{F}}(R_j).$$

Since $\beta_{i,\mathcal{F}}$ satisfies conjunction, an easy induction argument shows that

$$R^n_{i,\mathcal{F}} = R_i \cap \beta_{i,\mathcal{F}}(R_j) \cap ... \cap (\beta_{i,\mathcal{F}} \circ ... \circ \beta_{j,\mathcal{F}})(R_i),\ n > 2 \text{ odd},$$

$$R^n_{i,\mathcal{F}} = R_i \cap \beta_{i,\mathcal{F}}(R_j) \cap ... \cap (\beta_{i,\mathcal{F}} \circ ... \circ \beta_{j,\mathcal{F}} \circ \beta_{i,\mathcal{F}})(R_j),\ n > 2 \text{ even}.$$

Therefore

$$CCOR_{i,\mathcal{F}} = \bigcap_{n \geq 1} R^n_{i,\mathcal{F}}\ .$$

5.2.1 Examples

We begin with an analysis of the game in Figure 1, along with the epistemic model defined by the tables in Subsection 5.1. Assume that $x > 0$.

Consider the state $((L, \theta'), d, \tau^2_1, \tau^2_2)$ corresponding to the second row in the left-hand table and the second row in the right-hand table. Clearly, both players choose best responses to their beliefs. In particular, Player 2 is certain at ϕ that Player 1, regardless of her type, will choose L, so that the history (R) should not occur. However, if it does, Player 2 revises his beliefs and becomes convinced that Player 1's payoff-relevant type is θ', which justifies his own choice of d.

Observe that this implies that, after history (R), type τ^2_2 of Player 2 is no longer certain that Player 1 is rational. However, at ϕ his beliefs are concentrated on states at which Player 1 chooses optimally, and this holds for the beliefs of Player 2's type τ^1_2 as well. That is, in any explicitly described state, Player 2 is certain of Player 1's rationality at the beginning of the game.

This, in turn, implies that, in any explicitly described state, Player 1 is certain at ϕ that (Player 2 is certain at ϕ that (Player 1 is rational).)

It is easy to see that, in fact, there is common certainty of the opponent's rationality at ϕ in state $((L, \theta'), d, \tau^2_1, \tau^2_2)$. Of course, in that state common certainty of rationality *would* fail after the *counterfactual* history (R).

Consider now state $((R, \theta''), u, \tau^4_1, \tau^1_2)$. It is easy to see that here, too, there is common certainty of the opponent's rationality at ϕ. However, regardless of the value of $p = g_{21,\phi}(\tau^1_2)((L, \theta'), \tau^1_1))$, now Player 2 remains convinced that Player 1 is rational even after observing R (which is an unexpected event, if $p = 0$.) Thus, Player 2's type τ^1_2 is actually certain of Player 1's rationality given $\mathcal{F} = \{\phi, (R)\}$.

Indeed, since $g_{12,\phi}(\tau^4_1)((u, \tau^1_2)) = g_{12,\phi}(\tau^1_1)((u, \tau^1_2)) = 1$, Player 1 is certain at ϕ (hence, by independence and Bayes' rule, also after (R)) that Player 2 is certain of 1's rationality given $\mathcal{F}$. One sees easily that $((R, \theta''), u, \tau^4_1, \tau^1_2) \in R \cap CCOR_{\mathcal{F}}$.

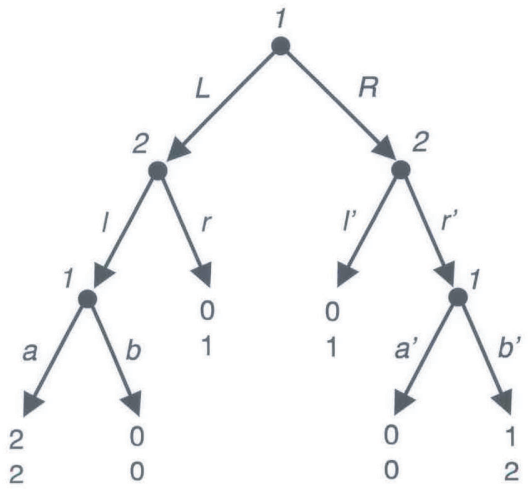

Figure 2

Thus, insisting on common certainty of the opponent's rationality at every history refines our prediction about the behavior of Player 1's type θ''. The underlying argument has the flavor of *forward induction*: faced with a deviation from his original prediction, Player 2 attempts to find an explanation for Player 1's choice of R which is consistent with the assumption that she is rational; but this implies that he has to assign conditional probability one to Player 1's payoff-relevant type θ'' — and consequently best-respond with u. Of course, Player 1 anticipates this, which makes R optimal at ϕ.

Notice however that this kind of reasoning leads to inconsistencies if $x < 0$. Intuitively, in this case if Player 1 is rational and is certain at ϕ that Player 2 (i) is rational and (ii) is certain at *both* ϕ and (R) that his opponent is rational, she should expect Player 2 to choose d, and should therefore pick L herself. But then Player 2 cannot be certain after history (R) that Player 1 is able to reason along these lines, precisely because then she would not choose R.

This informal argument *suggests* that rationality and common certainty of rationality given $\mathcal{F}$ are impossible in this game. In the next subsection we will show that rationality and common certainty of rationality given a family of histories $\mathcal{F}$ identify a simple iterative deletion procedure. We shall then be able to show that the above intuition is correct by noting that no strategy profile survives the relevant procedure in the game of Figure 1.

The game of complete and perfect information depicted in Figure 2 (cf. Figure 5 in Reny [24]) further illustrates the differences between common certainty of the opponent's rationality for a given history and for a collection of histories. Table 2 shows all the essential elements of a type space for this game with four

epistemic types for each player. (Recall that we show only the states satisfying conditions (1) and (2) of Definition 7. We do not list all the eight strategies of Player 1 and we do not specify the beliefs of Player 1 when she is irrational.)

(σ_1, τ_1)	$\phi, (R), (L)$	(L, l)	(R, r')
Lab', τ_1^1	0,1,0,0	0,1,0,0	0,1,0,0
Lbb', τ_1^2	...	...	...
Raa', τ_1^3	...	...	...
Rab', τ_1^4	0,0,0,1	0,1,0,0	0,0,0,1

(σ_2, τ_2)	ϕ	$(L), (L, l)$	$(R), (R, r')$
ll', τ_2^1	1,0,0,0	1,0,0,0	$0,0,\frac{3}{4},\frac{1}{4}$
lr', τ_2^2	1,0,0,0	1,0,0,0	0,0,0,1
rl', τ_2^3	1,0,0,0	1,0,0,0	$0,0,\frac{3}{4},\frac{1}{4}$
rr', τ_2^4	0,0,0,1	0,1,0,0	0,0,0,1

Table 2: A type space for the game in Figure 2

It can be verified that common certainty of the opponent's rationality at histories/nodes (L) and (R) is possible. For example,

$$(Lab', lr', \tau_1^1, \tau_2^2) \in R \cap CCOR_{\{\phi,(L)\}},\ (Rab', lr', \tau_1^4, \tau_2^2) \in R \cap CCOR_{\{\phi,(R)\}}.$$

Note also that state $(Lab', lr', \tau_1^1, \tau_2^2)$ satisfies the property the there *would be* common certainty of the opponent's rationality *if* (R) occurred, even though state $(Lab', lr', \tau_1^1, \tau_2^2)$ precludes history (R). Therefore,

$$(Lab', lr', \tau_1^1, \tau_2^2) \in R \cap CCOR_{\{\phi,(L)\}} \cap CCOR_{\{\phi,(R)\}}.$$

However, we argue that common certainty of the opponent's rationality given $\{\phi, (L), (R)\}$ is impossible: if player 2 is rational and believes after (L) that player 1 is rational, player 2 chooses l after (L). Anticipating this and being rational, as implied by $R^3_{1,\{\phi,(L),(R)\}} \supset CCOR_{\{\phi,(L),(R)\}}$, player 1 chooses L. But then the occurrence of history (R) would imply that either player 1 is irrational or she does not believe that player 2 would believe at (L) that player 1 is rational. Therefore player 2 could not believe event $R^3_{1,\{\phi,(L),(R)\}}$ after history (R) and this implies $CCOR_{\{\phi,(L),(R)\}} = \emptyset$. The characterization result provided below can be used to verify our claim with a simple iterative deletion procedure.

5.2.2 Characterization

We can ask the following questions about common certainty of the opponent's rationality:

(i) How can we characterize the set of σ consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$ without any reference to epistemic types?

(ii) When we consider the set of σ consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$, can we restrict our attention to *finite* type spaces (more generally, type spaces with the same cardinality of Σ)?

(iii) Can we restrict our attention to the *universal* type space $\mathcal{T}^u$ containing all the hierarchies of conditional systems satisfying common certainty of coherency?

It is known that, for *static games*, the answers to (ii) and (iii) are affirmative and that the answer to the characterization problem (i) is given by an inductive procedure equivalent to the iterated deletion of strictly dominated strategies.[18] These results can now be extended to dynamic games. Let us start from (i). By analogy with the analysis of static games the answer should rely on an inductive procedure. For any $K_j \subset \Sigma_j$, $\mathcal{F} \subset \mathcal{H}$, let

$$\Lambda_{i,\mathcal{F}}(K_j) := \left\{\mu_{ij} \in \Delta^{\mathcal{H}}(\Sigma_j) : \forall h \in \mathcal{F}, \mu_{ij}(K_j|\Sigma_j(h)) = 1\right\}.$$

(Note that, if $\mathcal{F}$ is "large" and K_j is "small," $\Lambda_{i,\mathcal{F}}(K_j)$ may well be empty.) The inductive procedure is defined as follows:

- $\Sigma^0_{i,\mathcal{F}} = \Sigma_i$, $i \in \{1,2\}$,
- for $n \geq 0$, $i,j \in \{1,2\}$, $i \neq j$, $\Sigma^{n+1}_{i,\mathcal{F}} = r_i\left[\Lambda_{i,\mathcal{F}}(\Sigma^n_{j,\mathcal{F}})\right]$.

That is, $\Sigma^{n+1}_{i,\mathcal{F}}$ is the set of (s_i,θ_i) such that s_i is a best response for θ_i to some CPS μ_{ij} satisfying $\mu_{ij}(\Sigma^n_{j,\mathcal{F}}|\Sigma_j(h)) = 1$ for all $h \in \mathcal{F}$. Note that $\Sigma^1_{i,\mathcal{F}}$ is the set of σ_i consistent with player i being rational and does not depend on $\mathcal{F}$. The natural conjecture is that the set of σ consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$ is $\Sigma^\infty_{1,\mathcal{F}} \times \Sigma^\infty_{2,\mathcal{F}} := \bigcap_{n\geq 1} \Sigma^n_{1,\mathcal{F}} \times \Sigma^n_{2,\mathcal{F}}$.

Define $\Sigma^n_{\mathcal{F}} := \Sigma^n_{1,\mathcal{F}} \times \Sigma^n_{2,\mathcal{F}}$, $n = 1,2,...,\infty$, and

$$\rho_{\mathcal{F}}(K_1 \times K_2) := r_1\left[\Lambda_{1,\mathcal{F}}(K_2)\right] \times r_2\left[\Lambda_{2,\mathcal{F}}(K_1)\right].$$

Clearly $\rho_{\mathcal{F}}$ is a monotone set to set operator,[19] $\Sigma^{n+1}_{\mathcal{F}} = \rho_{\mathcal{F}}(\Sigma^n_{\mathcal{F}})$ and the sequence $\{\Sigma^n_{\mathcal{F}}\}^\infty_{n=0}$ is (weakly) decreasing. Since Σ is finite there is some N such that, for all $n \geq N$, $\Sigma^\infty_{\mathcal{F}} = \Sigma^n_{\mathcal{F}}$. This implies that the product set $\Sigma^\infty_{\mathcal{F}}$ has the familiar fixed point property: $\Sigma^\infty_{\mathcal{F}} = \rho_{\mathcal{F}}(\Sigma^\infty_{\mathcal{F}})$. It is easy to prove (using monotonicity of $\rho_{\mathcal{F}}$) that every rectangular subset Σ^* such that $\Sigma^* \subset \rho_{\mathcal{F}}(\Sigma^*)$ is a subset of $\Sigma^\infty_{\mathcal{F}}$. In general, $\Sigma^\infty_{\mathcal{F}}$ may well be empty (cf. Reny [24] and the related comments in the next section.) But it can be shown that Σ^∞_ϕ is nonempty.[20] Given the fixed point property of $\Sigma^\infty_{\mathcal{F}}$ it is easy to verify that $\Sigma^\infty_{\mathcal{F}} \neq \emptyset$ if and only if, for all $h \in \mathcal{F}$, $\Sigma^\infty_{\mathcal{F}} \cap \Sigma(h) \neq \emptyset$.

The following results relate operator $\rho_{\mathcal{F}}$ and procedure $\{\Sigma^n_{\mathcal{F}}\}$ to (rationality and) common certainty of the opponent's rationality given $\mathcal{F}$.

---

[18] These results have been explicitly proved for normal-form games of complete information, but they can be extended to games of incomplete information.

[19] A set to set operator ρ is *monotone* if $A \subset B$ implies $\rho(A) \subset \rho(B)$ ($\subset$ denotes *weak* inclusion.)

[20] The proof goes as follows: Take any non-empty rectangular subset $\Sigma^* \subset \Sigma$. Then, for each player i and opponent j, there is a CPS $\mu_{ij} \in \Delta^{\mathcal{H}}(\Sigma_j)$ such that $\mu_{ij}(\Sigma^*_j|\Sigma_j) = 1$, and for each θ_i we can find a strategy $s_i \in S_i$ such that $(s_i,\theta_i) \in r_i(\mu_{ij})$. When we apply this construction to $\Sigma^* = \Sigma$, we obtain $\Sigma^1_\phi \neq \emptyset$. The construction can be applied inductively to $\Sigma^* = \Sigma^n_\phi \neq \emptyset$, thus obtaining $\Sigma^{n+1}_\phi \neq \emptyset$.

Lemma 5. *Let $\Sigma^* = \Sigma_1^* \times \Sigma_2^* \subset \Sigma$, $\emptyset \neq \mathcal{F} \subset \mathcal{H}$. If $\Sigma^* \subset \rho_{\mathcal{F}}(\Sigma^*)$, then there is a type space $\mathcal{T} = (\Sigma, \mathcal{H}, T_1, T_2, g_1, g_2)$ such that*
(a) $T_1 \times T_2$ has the same cardinality as Σ,
(b) for all $\sigma \in \Sigma^*$, there is a pair of epistemic types $(\tau_1, \tau_2) \in T_1 \times T_2$ such that

$$(\sigma, \tau_1, \tau_2) \in R \cap CCOR_{\mathcal{F}}.$$

Proof: See the Appendix. ■

Proposition 4. *Fix a non empty collection $\mathcal{F}$ of partial histories.*
(i) The set of σ consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$ (cf. Definition 8) is precisely $\Sigma_{\mathcal{F}}^{\infty}$.
(ii) There is a finite *type space $\mathcal{T}$ such that, for all $\sigma \in \Sigma$, σ is consistent with rationality and common certainty of the opponent's rationality if and only if*

$$(\sigma, \tau_1, \tau_2) \in R \cap CCOR_{\mathcal{F}}$$

for some (τ_1, τ_2) in $\mathcal{T}$ (events are defined in the finite type space $\mathcal{T}$.)
(iii) For all $\sigma \in \Sigma$, σ is consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$ if and only if there is some pair of hierarchies of CPSs $(t_1, t_2) \in T \times T$ such that

$$(\sigma, t_1, t_2) \in R \cap CCOR_{\mathcal{F}}$$

(events are defined in the universal type space $\mathcal{T}^u$.)

Proof: (i) Since $\Sigma_{\mathcal{F}}^{\infty} = \rho_{\mathcal{F}}(\Sigma_{\mathcal{F}}^{\infty})$, Lemma 5 implies that every σ in $\Sigma_{\mathcal{F}}^{\infty}$ is consistent with rationality and common certainty of the opponent's rationality. To prove the converse, fix a type space $\mathcal{T} = (\Sigma, \mathcal{H}, T_1, T_2, g_1, g_2)$ and, for $i = 1, 2$, consider the sequence of events $\{R_{i,\mathcal{F}}^n\}_{n \geq 1}$ defined in $\mathcal{T}$ as indicated in the preceding subsection. We show by induction that for each i and n the projection of $R_{i,\mathcal{F}}^n$ on Σ_i is (weakly) contained in $\Sigma_{i,\mathcal{F}}^n$. This implies the assertion, because $R_i \cap CCOR_{i,\mathcal{F}} = \bigcap_{n \geq 1} R_{i,\mathcal{F}}^n$. To simplify the notation, let $\lambda_{ij}(\tau_i) = (\lambda_{ij,h}(\tau_i))_{h \in \mathcal{H}}$ denote the system of first order beliefs for type τ_i concerning her opponent: that is, for all $h \in \mathcal{H}$,

$$\lambda_{ij,h}(\tau_i) = mrg_{\Sigma_j} g_{i,h}(\tau_i).$$

Base Step. Let $(\sigma, \tau_1, \tau_2) \in R_i = R_{i,\mathcal{F}}^1$. Then $\sigma_i \in r_i(\lambda_{ij}(\tau_i))$, which implies $\sigma_i \in \Sigma_{i,\mathcal{F}}^1$.

Induction Step. Suppose that for each player i and state $(\sigma', \tau_1', \tau_2')$, $(\sigma', \tau_1', \tau_2') \in R_i^n$ implies $\sigma_i' \in \Sigma_{i,\mathcal{F}}^n$. Let $(\sigma, \tau_1, \tau_2) \in R_{i,\mathcal{F}}^{n+1} = R_i \cap \beta_{i,\mathcal{F}}(R_{j,\mathcal{F}}^n)$. Since i is rational at (σ, τ_1, τ_2), $\sigma_i \in r_i(\lambda_{ij}(\tau_i))$. Furthermore, since type τ_i is certain of $R_{j,\mathcal{F}}^n$ at each $h \in \mathcal{F}$, the induction hypothesis (projection of $R_{j,\mathcal{F}}^n$ on Σ_j contained in $\Sigma_{j,\mathcal{F}}^n$) implies that for all $h \in \mathcal{F}$, $\lambda_{ij,h}(\tau_i)(\Sigma_{j,\mathcal{F}}^n) = 1$. Therefore $\sigma_i \in \Sigma_{i,\mathcal{F}}^{n+1}$.

(ii) Since $\Sigma_{\mathcal{F}}^{\infty} = \rho_{\mathcal{F}}(\Sigma_{\mathcal{F}}^{\infty})$, Lemma 5 and (i) imply that there is a type space $\mathcal{T}$ with the stated property and the same cardinality as the finite set Σ.

(iii) The "if" part is true by definition. To prove the "only if" part, fix σ and suppose that for some type space $\mathcal{T}$ and some pair of types (τ_1, τ_2),

$$(\sigma, \tau_1, \tau_2) \in R_{\mathcal{T}} \cap CCOR_{\mathcal{F},\mathcal{T}},$$

where we use subscript $\mathcal{T}$ to denote that events and belief operators appearing in the construction of (R and) $CCOR_{\mathcal{F}}$ are defined in the space $\mathcal{T}$. By Proposition 3 there is a type morphism $\varphi = (Id_\Sigma, \varphi_1, \varphi_2)$ from $\mathcal{T}$ to the universal space $\mathcal{T}^u$. We prove that

$$(\sigma, \varphi_1(\tau_1), \varphi_2(\tau_2)) \in R \cap CCOR_{\mathcal{F}}.$$

where the absence of the subscript $\mathcal{T}$ indicates that events and belief operators are defined in $\mathcal{T}^u$. We will adhere to this convention throughout the proof.

The claim follows from Lemma 4. To see this, for each $i = 1, 2$, let $E_i \subset \Sigma \times I\Delta^{\mathcal{H}}(\Sigma_i, \Sigma_j)) \times I\Delta^{\mathcal{H}}(\Sigma_i, \Sigma_j)$ be set of (σ, μ_1, μ_2) such that

- μ_1 and μ_2 are independent CPSs; denote by μ_{ii} and μ_{ij} the marginal CPS such that $\mu_i(.|\Sigma(h)) = \mu_{ii}(.|\Sigma_i(h)) \otimes \mu_{ij}(.|\Sigma_j(h))$ for all $h \in \mathcal{H}$;
- for $\sigma_i = (s_i, \theta_i)$, $\mu_{ii}(\sigma_i|\Sigma_i(h)) = 1$ for all $h \in \mathcal{H}(s_i)$, and $\mu_{ii}(S_i \times \{\theta_i\}|\Sigma_i(h)) = 1$ for all $h \in \mathcal{H}$;
- $\sigma_i \in r_i(\mu_{ij})$.

(compare this with Definition 5.3.) Also define

$$I_{i,\mathcal{T}'} = \{(\sigma, \tau_1', \tau_2') \in \Sigma \times T_1' \times T_2' \ : \ g_i'(\tau_i') \in I\Delta^{\mathcal{H}}(\Sigma_i, \Sigma_j \times T_j')\}$$

for any type space $\mathcal{T}' = (\Sigma, \mathcal{H}, T_1', T_2', g_1', g_2')$. Lemma 3 in the Appendix immediately implies that

$$\varphi(I_{i,\mathcal{T}}) \subset I_{i,\mathcal{T}^u}$$

Also observe that $R_{i,\mathcal{T}} = E_{i,\mathcal{T}} \cap I_{i,\mathcal{T}}$ and $R_i = E_{i,\mathcal{T}^u} \cap I_{i,\mathcal{T}^u}$. Then Lemma 4 implies that for all $i = 1, 2$, $n \geq 1$,

$$\varphi(R_{i,\mathcal{T}}) \subset \varphi(E_{i,\mathcal{T}}) \cap \varphi(I_{i,\mathcal{T}}) \subset E_{i,\mathcal{T}^u} \cap I_{i,\mathcal{T}^u} = R_i$$

and similarly

$$\varphi\left((\beta_{i_1,\mathcal{F},\mathcal{T}} \circ ... \circ \beta_{i_n,\mathcal{F},\mathcal{T}})(R_{i_{n+1},\mathcal{T}})\right) \subset (\beta_{i_1,\mathcal{F}} \circ ... \circ \beta_{i_n,\mathcal{F}})(R_{i_{n+1}}),\ i_1 = i,\ i_{k+1} \neq i_k.$$

■

5.2.3 Examples (Reprise)

We can finally go back to the examples in Figures 1 and 2 and provide the details of the arguments sketched in Subsection 5.2.1.

Consider first the signalling game of Figure 1. We claimed above that, if $x < 0$, then common certainty of rationality given $\mathcal{F} = \{\phi, (R)\}$ is impossible.

To see this, note that $\Sigma^1_{1,\mathcal{F}} = \{(L,\theta'),(L,\theta''),(R,\theta'')\}$ and $\Sigma^2_{2,\mathcal{F}} = \{u,d\}$; then $\Sigma^2_{1,\mathcal{F}} = \Sigma^1_{1,\mathcal{F}}$, while $\Sigma^2_{2,\mathcal{F}} = \{d\}$, because now Player 2 must assign probability 1 to $\Sigma_1((R)) \cap \Sigma^1_{1,\mathcal{F}} = (R,\theta'')$ after observing R, and we are assuming that $x < 0$. But then $\Sigma^3_{1,\mathcal{F}} = \{(L,\theta'),(L,\theta'')\}$, $\Sigma^3_{2,\mathcal{F}}$ is unchanged, and finally $\Sigma^4_{2,\mathcal{F}} = \emptyset$, because $\Sigma_1((R)) \cap \Sigma^3_{1,\mathcal{F}} = \emptyset$ (which implies $\Lambda_{2,\mathcal{F}}(\Sigma^3_{1,\mathcal{F}}) = \emptyset$.) The characterization result (Proposition 4) now implies that $R \cap CCOR_{\mathcal{F}} = \emptyset$ in any type space.[21]

Similarly, for the game in Figure 2 we obtain $\Sigma^4_{\{\phi,(L),(R)\}} = \emptyset$ and this implies that common certainty of the opponent's rationality given the collection of histories $\mathcal{F} = \{\phi,(L),(R)\}$ is impossible.

5.2.4 CCOR in games with perfect information

In light of this conclusion, it is natural to ask whether one can find conditions which ensure the possibility of rationality and common certainty of the opponent's rationality given a collection of "interesting" histories.

Using results from Reny [24], we are able to provide an answer to this question for generic games with complete and perfect information.[22]

First, following Reny, we deem a history $h \in \mathcal{H}$ *relevant* (*i.e.,* "interesting") if (i) h is consistent with rationality: $\Sigma(h) \cap \Sigma^1_\phi \neq \emptyset$[23]; and (ii) at least one player i has a payoff-type θ_i which does *not* have a *dominant choice*[24].

In the game of Figure 2, both (L) and (R) are relevant (as is ϕ.) Thus, in that game, rationality and common certainty of the opponent's rationality given all relevant nodes is *not* possible. Indeed, the class of games for which rationality and common certainty of the opponent's rationality given all relevant nodes are possible is very small. On the other hand, when these conditions hold, the backward induction outcome obtains:

Proposition 5. *(cf. Reny [24]) Consider a game with perfect and complete information and no ties between payoffs at terminal histories. Let $\mathcal{R}$ be the set of its relevant histories. Then*

[21] On the other hand, Player 2, upon observing R, may conclude that Player 1 is rational, but not not very sophisticated (*i.e.,* she does not realize that Player 2 will interpret her choice of R as a signal that her type is θ''.) This reinforces his inducement to play d, which again leads Player 1 to choose L. Battigalli and Siniscalchi [5] develop these ideas and show that they lead to an epistemic characterization of extensive-form rationalizability (Pearce [22].) A similar set of assumptions yields the backward induction solution in the game of Figure 2.

[22] Reny attempts to capture the intuitive notion of rationality and common certainty of rationality by defining particular subsets of (Bayesian rational) strategy profiles satisfying a belief closure condition with respect to a collection of nodes. His results concern the (non)emptiness of such sets of strategy profiles.

[23] The formal definition is motivated by the observation that, for *any* collection of histories $\mathcal{F} \subset \mathcal{H}$, $\Sigma^1_{\mathcal{F}} = \Sigma^1_\phi$.

[24] Type θ_i of Player i has a dominant choice at $h \in \mathcal{H}$ if and only if there exists a strategy $s_i \in S_i(h)$ such that, for all $s'_i \in S_i(h)$ such that $s'_i(h) \neq s_i(h)$, $U_i(s_i,\theta_i,s_j,\theta_j) > U_i(s'_i,\theta_i,s_j,\theta_j)$ for all $(s_j,\theta_j) \in \Sigma_j$.

(a) there is a type space $\mathcal{T}$ for the game such that $R_{\mathcal{T}} \cap CCOR_{\mathcal{R},\mathcal{T}} \neq \emptyset$ if and only if every history $h \in \mathcal{R}$ is on the backward induction path;
(b) for all states $(\sigma, \tau_1, \tau_2) \in R_{\mathcal{T}} \cap CCOR_{\mathcal{R},\mathcal{T}}$, σ induces the backward induction path.

Proof: (a) ($\Rightarrow$) Suppose $R_{\mathcal{T}} \cap CCOR_{\mathcal{R},\mathcal{T}} \neq \emptyset$. Then, by Proposition 4, $\Sigma^{\infty}_{\mathcal{R}} \neq \emptyset$.

Observe that, for $i = 1, 2$ and $j \neq i$, for any collection $\mathcal{F} \subset \mathcal{H}$ and for any $K_j \subset \Sigma_j$,

$$\Lambda_{i,\mathcal{F}}(K_j) = \Lambda_{i,\mathcal{F}}(K_j \cap \bigcup_{h \in \mathcal{F}} \Sigma_j(h)).$$

This follows directly from the definition of $\Lambda_{i,\mathcal{F}}$. Since for each $i = 1, 2$ and $j \neq i$, $\Sigma^{\infty}_{i,\mathcal{R}} = r_i \left[\Lambda_{i,\mathcal{R}}(\Sigma^{\infty}_{j,\mathcal{R}})\right]$, we conclude that

$$\Sigma^{\infty}_{i,\mathcal{R}} = r_i \left[\Lambda_{i,\mathcal{R}}(\Sigma^{\infty}_{j,\mathcal{R}} \cap \bigcup_{h \in \mathcal{R}} \Sigma_j(h))\right]$$

But then $\Sigma^{\infty}_{\mathcal{R}} \neq \emptyset$ implies that $\Sigma^{\infty}_{j,\mathcal{R}} \cap \bigcup_{h \in \mathcal{R}} \Sigma_j(h) \neq \emptyset$, for $i = 1, 2$ and $j \neq i$. Therefore we conclude that

$$\emptyset \neq \Sigma^{\infty}_{i,\mathcal{R}} \cap \bigcup_{h \in \mathcal{R}} \Sigma_i(h) = r_i \left[\Lambda_{i,\mathcal{R}}(\Sigma^{\infty}_{j,\mathcal{R}} \cap \bigcup_{h \in \mathcal{R}} \Sigma_j(h))\right] \cap \bigcup_{h \in \mathcal{R}} \Sigma_i(h)$$

That is, letting $\Sigma^{*}_{i,\mathcal{R}} \equiv \Sigma^{\infty}_{i,\mathcal{R}} \cap \bigcup_{h \in \mathcal{R}} \Sigma_i(h)$ for $i = 1, 2$ and $j \neq i$, each set $\Sigma^{*}_{i,\mathcal{R}}$ satisfies

$$\emptyset \neq \Sigma^{*}_{i,\mathcal{R}} = r_i \left[\Lambda_{i,\mathcal{R}}(\Sigma^{*}_{j,\mathcal{R}}\right] \cap \bigcup_{h \in \mathcal{R}} \Sigma_i(h)$$

so that the pair $(\Sigma^{*}_{1,\mathcal{R}}, \Sigma^{*}_{2,\mathcal{R}})$ constitutes a nonempty *jointly rational beliefs system* for $\mathcal{R}$ as defined in Reny [24], p. 269. Hence, by the main Theorem in that paper, all relevant histories are on the backward induction path.

($\Leftarrow$) Let $s^B \in S$ be the backward induction strategy profile. Suppose that relevant histories are all on the backward induction path: $h \in \mathcal{R} \Rightarrow s^B \in S(h)$. For $i = 1, 2$, let $\Sigma^{*}_i = \{s^B_i\}$ and consider the CPS $\mu^B_i \in \Delta^{\mathcal{H}}(S_i)$ defined as follows: (i) if $s^B_i \in S_i(h)$, then $\mu^B_i(\{s^B_i\}|S_i(h)) = 1$; (ii) otherwise, let $[s^B_i]_h = \{s_i \in S_i(h) \ : \ s_i(h') = s^B_i(h') \text{ for all } h' \text{ weakly following } h\}$, and let $\mu^B_i(\{s_i\}|S_i(h)) = 1/\#[s^B_i]_h$ for all $s_i \in [s^B_i]_h$. Then, for $i = 1, 2$ and $j \neq i$, $\mu^B_j \in \Lambda_{i,\mathcal{R}}(\Sigma^{*}_j)$, and $s^B_i \in r_i(\mu^B_j)$. That is, $\Sigma^{*}_i \subset r_i \left[\Lambda_{i,\mathcal{R}}(\Sigma^{*}_j)\right]$ for $i = 1, 2$. Now Lemma 5 implies that $R_{\mathcal{T}} \cap CCOR_{\mathcal{R},\mathcal{T}} \neq \emptyset$ for some (finite) type space $\mathcal{T}$.

We omit the proof of part (b). ■

5.2.5 Common Certainty of *Both* Players' Rationality

Several papers on the epistemic analysis of games focus on common certainty of *both* players' rationality at a given history, rather than common certainty of the *opponent's* rationality conditional on a collection of histories (see in particular

Stalnaker [28] and Ben Porath [7].) In static games, there is no relevant difference between these two sets of assumptions. Since a rational player knows her strategy and beliefs, she is certain of her own rationality. Therefore, rationality and mutual certainty of the opponent's rationality is equivalent to rationality and mutual certainty of both players' rationality. But since players' beliefs satisfy positive introspections, this also implies that rationality and common certainty of the opponent's rationality is indeed equivalent to rationality and common certainty of both players' rationality.

In dynamic games the result can be extended as follows.

Fix a type space $\mathcal{T}$ and a history h. The event that there is (would be) common certainty of rationality at h is

$$CCR_h = \beta_h(R) \cap \beta_h(\beta_h(R)) \cap ... = \bigcap_{n \geq 1} \beta_h^n(R),$$

where $\beta_h(E) = \beta_{1,h}(E) \cap \beta_{2,h}(E)$ and $\beta_h^{n+1}(E) = \beta_h(\beta_h^n(E))$. Let $[h] := \Sigma(h) \cap T_1 \cap T_2$ denote the event that history h occurs.

Proposition 6. *For every type space $\mathcal{T}$ and history $h \in \mathcal{H}$,*

$$[h] \cap R \cap CCOR_h = [h] \cap R \cap CCR_h.$$

Proof: Omitted.

However, we may have

$$\text{(i) } R \cap \beta_{1,h}(R_2) \cap \beta_{2,h}(R_1) \neq R \cap \beta_h(R)$$

and

$$\text{(ii) } R \cap \beta_{\mathcal{F}}(R) = \emptyset \neq R \cap \beta_{1,\mathcal{F}}(R_2) \cap \beta_{2,\mathcal{F}}(R_1).$$

To see (i) consider a state $(\sigma, \tau_1, \tau_2) \in R \cap \beta_{1,h}(R_2) \cap \beta_{2,h}(R_1)$ where history h is counterfactual, for example, because $\sigma_1 \notin \Sigma_1(h)$. Suppose that all best responses to type τ_1's (first-order) beliefs prevent h from being reached. If h were reached, player 1 could not believe that she is rational, because she would know that she has deviated from her best response. To see (ii) suppose that $\mathcal{F}$ contains two mutually exclusive histories h' and h'' encoding different actions for player 1 at a common predecessor h. Then it may be impossible to find a single (first-order) belief for player 1 justifying both actions even if both h' and h'' are consistent with player 1's rationality.

6 Concluding Remarks

In this paper we provide the main tools for the epistemic analysis of multi-agent dynamic models and we consider some applications to multistage games with observed actions. Taking as given a collection of conditioning events — or "relevant hypotheses" — concerning external (*i.e.,* non-epistemic) states, we construct a belief-closed space T of (coherent) infinite hierarchies of conditional

probability systems (CPSs.) An infinite hierarchy of CPSs encodes an individual's dispositions to believe conditional on every relevant hypothesis — that is, an individual's epistemic type.

The set $\Omega = \Sigma \times T \times T$ of profiles of external states and infinite hierarchies of CPSs can be interpreted as a universal (semantic) model providing truth conditions, at every state $\omega \in \Omega$, for subjunctive conditionals of the form "if B occurred, player i would believe E," where B is a relevant hypothesis, E is an event concerning the external state and/or the agents' interactive conditional beliefs, and the truth value is assigned even if B is counterfactual at ω ($\omega \notin B$.) Of course, subjunctive conditionals are crucial for the analysis of counterfactual reasoning in extensive form (dynamic) games.

Other authors, including Ben Porath [7] and Stalnaker [28], put forward "extensive form" epistemic models, but — to the best of our knowledge — we are the first to provide the explicit construction of a *universal* type space of this sort, thus extending classical results of Mertens and Zamir [19] and Brandenburger and Dekel [10] to a dynamic framework. In particular, to facilitate the comparison with this literature, we mimic as closely as possible the elegant and relatively simple construction of [10].

The space of infinite hierarchies of CPSs is an important analytical tool because it does not exclude any "conceivable" epistemic type; thus, it provides an "epistemically neutral" representation of interactive conditional beliefs. This allows us to state characterization results in a clean "if-and-only-if" form (*i.e.,* "for all $\sigma \in \Sigma$, σ belongs to the solution set Σ^* *if and only if* there is a profile of epistemic types such that ...")

Universal type spaces are particularly important for the epistemic analysis of solution concepts featuring forward induction. According to forward induction reasoning, a player always seeks a "rational" explanation of her opponent's observed behavior. When the extensive form game is embellished with an epistemic model, this amounts to looking for an opponent's epistemic type (equivalently, a hierarchy of conditional beliefs) that "rationalizes" the opponents' actions. Thus, adopting a non-universal model effectively *restricts* the alternative explanations available to a player. While constraining players' inferences may be desirable in certain applications, the restrictions implicit in a non-universal model prevent a neutral analysis of forward induction reasoning. We pursue this topic in Battigalli and Siniscalchi [5, 6] (see also Stalnaker [29].)

On the other hand, for many purposes — in particular, for the analysis of specific examples and in the proofs of some results — it is more convenient to work with "small" (*e.g.,* finite) non-universal type spaces. Therefore it is important to be able to relate extensive form type spaces to each other and to the space of infinite hierarchies of CPSs. We extend Mertens and Zamir's (1985) notion of "belief-preserving" mappings between type spaces (type-morphisms) and their fundamental result showing that every type space is equivalent to a belief-closed subset of the space of infinite hierarchies of beliefs.

The main result of our game theoretic analysis is the characterization of (rationality and) common certainty of the opponent's rationality given an *arbitrary* collection $\mathcal{F}$ of histories (*i.e.,* $R \cap CCOR_{\mathcal{F}}$, in the notation of Section 5.)

We build on and extend previous work by Ben Porath [7] and Reny [24]. Ben Porath restricts his analysis to finite type spaces for perfect information games and characterizes the strategies consistent with initial common certainty of rationality.[25] He conjectures that his characterization is also valid for infinite type spaces. Our results confirm his conjecture, show that his type spaces can be embedded into an explicitly constructed universal type space, and generalize his characterization.

Reny [24] studies the possibility of common certainty of the opponent's rationality conditional on certain collections of nodes in a perfect information game. His analysis does not employ a formal extensive-form epistemic model, but rather verifies whether one can find non-empty subsets of strategy profiles satisfying an intuitive fixed point property. Our results provide an "epistemic validation" of Reny's analysis.

Our work is also related to Stalnaker's [28, 29] analysis of counterfactual reasoning in games. Stalnaker's approach draws on the philosophical work discussing the axioms that belief revision should satisfy independently of any particular information structure (see, *e.g.,* Gärdenfors [14] and references therein.) A *belief revision function* specifies which events an individual would believe if she came to be certain of any particular — epistemic and/or external — event B. The probabilistic version of a belief revision function is a *complete* conditional probability system, specifying conditional beliefs for *every* nonempty subset of the relevant set of states (Myerson [20].) To use our terminology and notation, let $g_i(\tau_i) \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$ be the CPS corresponding to type τ_i. While in our notion of type space $\mathcal{B}$ is a collection of non empty subsets of Σ typically given by some kind of information structure, in Stalnaker [28, 29] $\mathcal{B}$ is the collection of *all* nonempty subsets of $\Sigma \times T_j$. Clearly, this is not a trivial difference. Since we are given an information structure, we are only interested in the beliefs an individual would have conditional on observable events. Hence we can afford to be more parsimonious in representing epistemic types and we are able to construct a universal type space. While Lemma 4 "justifies" using type spaces in our sense, we doubt that an analogous result holds for Stalnaker's epistemic spaces. However, it is easily shown that every epistemic model *à la* Stalnaker generates a type space in our sense (*i.e.,* a type space *à la* Ben Porath) and — more interestingly — every finite type space in our sense can be "enriched" so as to become a type space *à la* Stalnaker.

Finally, we find it useful to compare our epistemic analysis with Aumann [1] and related papers, such as Aumann [2], Samet [27] and Balkenborg and Winter [3]. There are two main differences between Aumann's approach and ours. First, Aumann and the other authors just mentioned assume that the players' initial epistemic state can be described by means of knowledge partitions on the set of states of the world. This can be expressed within our framework as a property which holds "locally" (*i.e.,* an event): players' initial beliefs (in a finite type space) assign positive probability to the true state and this is (initially) common certainty.

[25] He also provides sufficient epistemic conditions for Nash equilibrium outcomes.

The second difference is more radical and makes it difficult to compare this set of papers with those discussed above:[26] in Aumann's epistemic model, a state of the world describes the players' strategies (dispositions to act) and their initial epistemic state, but it does *not* describe how a player would revise her beliefs, should she *learn* that a particular history h has occurred. However, a belief revision theory of a sort is implicit in his definition of "rationality" (and made explicit in Aumann [2]): Suppose that player i is initially certain that his opponent's strategy prescribes action a at history h', which (weakly) follows history h, then she is certain of this also at h, whatever "at h" means. Note that this is completely unrelated to Bayesian updating. There is no notion that, upon learning that h has occurred, player i discards all the states of the world inconsistent with h. Indeed, it may well be the case that, in a model *à la* Aumann, no state of the world is consistent with h and yet each player has well-defined beliefs at h.[27]

Also, Samet's [27] theory of "hypothetical knowledge," — although interesting in its own right — is unrelated to Bayesian updating. In that paper, a state of the world does not only describe players' strategies and initial epistemic state (knowledge), but also what each player imagines she would know if any hypothetical event H (possibly inconsistent with her initial knowledge) were the case. This is *different* from this player imagining what she would know (or believe) if she *learned* that H has occurred. In fact, Samet does *not* assume that player i imagines that if H were the case she would know it. (For example, we know that the Earth is not flat, but we can imagine worlds were the Earth is flat and we hotly debate the competing theories about its shape without really knowing which is true.)

7 Appendix

7.1 Proof of Lemma 1

Remark 5. *Given Axioms 1 and 2, Axiom 3 is equivalent to the following:* ***Axiom 3'.*** *For all $B, C \in \mathcal{B}$ such that $B \subset C$ and all measurable functions $f : X \to [0,1]$ such that $f(X \backslash B) = \{0\}$*

$$\int f d\mu(\cdot|C) = \mu(B|C) \int f d\mu(\cdot|B).$$

Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of CPSs weakly converging to $\mu \in [\Delta(X)]^{\mathcal{B}}$. We must show that μ satisfies Axioms 1 and 3.

[26] For more on this comparison see also Stalnaker [29].

[27] On the other hand, suppose that player i is initially certain that her opponent's strategy prescribes either action a or action b at history h' which weakly follows h, but she is also initially certain that the only opponent's strategy consistent with h being reached prescribes a at h'. According to Bayesian updating, player i should be certain at h that the opponent would choose a at h'. But in Aumann's model this inference is incorrect.

(**Axiom 1 holds**) For all $B, C \in \mathcal{B}$, since B is *clopen* (*cl*osed and *open*), its boundary is empty. Therefore B must be a $\mu(\cdot|C)$-continuity set and $\lim_{n\to\infty} \mu_n(B|C) = \mu(B|C)$ (see, *e.g.,* Dudley [12], Theorem 11.1.1.) In particular,

$$\mu(B|B) = \lim_{n\to\infty} \mu_n(B|C) = 1.$$

(**Axiom 3 holds)** Fix $A \in \mathcal{A}$, $B, C \in \mathcal{B}$ such that $A \subset B \subset C$. Since any finite Borel measure on X is (closed) regular (Dudley [12], Theorem 7.1.3), for all $\varepsilon > 0$, we can find a closed set A' and an open set A^* such that $A' \subset A \subset A^*$ and

$$\max\left\{\left(\mu(A^*|C) - \mu(A'|C)\right), \left(\mu(A^*|B) - \mu(A'|B)\right)\right\} \leq \varepsilon,$$

Recall that B is (closed and) open. Therefore, the set $A'' := B \cap A^*$ is open. Furthermore, $A' \subset A \subset A'' \subset B$ and

$$\max\left\{\left(\mu(A''|C) - \mu(A'|C)\right), \left(\mu(A''|B) - \mu(A'|B)\right)\right\} \leq \varepsilon.$$

Since A' and $X\backslash A''$ are disjoint closed subsets of the normal topological space X, by Urysohn's lemma we can find a continuous function $f : X \to [0, 1]$ such that $f(A') = \{1\}$ and $f(X\backslash A'') = \{0\}$. In particular, $f(X\backslash B) = \{0\}$. Thus, by Remark 5, for all n

$$\int f d\mu_n(\cdot|C) = \mu_n(B|C) \int f d\mu_n(\cdot|B).$$

Since $\mu_n(\cdot|C)$ and $\mu_n(\cdot|B)$ weakly converge to $\mu(\cdot|C)$ and $\mu(\cdot|B)$, B is clopen, and f is bounded and continuous, by taking limits we obtain

$$\int f d\mu(\cdot|C) = \mu(B|C) \int f d\mu(\cdot|B).$$

Collecting all these equalities and inequalities and taking into account the properties of f we obtain

$$\mu(A|C) \leq \mu(A'|C) + \varepsilon \leq \mu(B|C) \int f d\mu(\cdot|B) + \varepsilon \leq \mu(B|C)(\mu(A|B) + \varepsilon) + \varepsilon$$

and

$$\mu(A|C) \geq \mu(A''|C) - \varepsilon \geq \mu(B|C) \int f d\mu(\cdot|B) - \varepsilon \geq \mu(B|C)\left(\mu(A|B) - \varepsilon\right) - \varepsilon.$$

Since ε is arbitrary, $\mu(A|C) = \mu(B|C)\mu(A|B)$.

7.2 Proof of Proposition 3

($\varphi_i(T_i) \subset T$) We first verify that $\varphi_i(T_i) \subset H_c$, that is, for all $\tau_i \in T_i$, $n \geq 1$, $B \in \mathcal{B}$, $mrg_{X^{n-1}} \varphi^{n+1}_{i,B}(\tau_i) = \varphi^n_{i,B}(\tau_i)$. Take $A^{n-1} \subset X^{n-1}$ (measurable.) Then

$$\varphi^{n+1}_{i,B}(\tau_i)(A^{n-1} \times \Delta^{\mathcal{B}}(X^{n-1})) = g_{i,B}(\tau_i)\left((\psi^n_{-i})^{-1}(A^{n-1} \times \Delta^{\mathcal{B}}(X^{n-1})\right) =$$

$$g_{i,B}(\tau_i)\left(\left\{(\sigma,\tau_j):\psi_{-i}^{n-1}(\sigma,\tau_j)\in A^{n-1}\right\}\right)=\varphi_{i,B}^{n}(\tau_i)(A^{n-1}).$$

Claim. $f\circ\varphi_i=\widehat{\varphi_{-i}}\circ g_i$, where $\varphi_{-i}=(Id_\Sigma,\varphi_j)$.

Proof of the claim. Take $A^n\subset X^n$ (measurable), $B\in\mathcal{B}$, and let $A=\mathcal{C}^\infty(A^n)$. Then

$$f_B(\varphi_i(\tau_i))(A)=\varphi_{i,B}^{n+1}(\tau_i)(A^n)=$$

$$g_{i,B}(\tau_i)\left((\psi_{-i}^{n})^{-1}((A^n))\right)=g_{i,B}(\tau_i)\left(\left\{(\sigma,\tau_j):(\sigma,\varphi_j^1(\tau_j),...,\varphi_j^n(\tau_j))\in A^n\right\}\right)=$$

$$g_{i,B}(\tau_i)\left(\{(\sigma,\tau_j):(\sigma,\varphi_j(\tau_j))\in A\}\right)=g_{i,B}(\tau_i)\left((\varphi_{-i})^{-1}(A)\right).$$

We now invoke the extension argument used in the proof of Proposition 1. Since the equality $f_B(\varphi_i(\tau_i))(A)=g_{i,B}(\tau_i)\left((\varphi_{-i})^{-1}(A)\right)$ holds on the algebra of cylinders, it extends to the sigma-algebra generated by the latter, which coincides with the Borel sigma-algebra generated by the product topology by second-countability. Thus, the claim is proved.

Next we show by induction that for each i, $\varphi_i(T_i)\subset T:=\bigcap_{n\geq 1}H_c^n$. Recall that $\varphi_i(\tau_i)\in H_c^n$, $n\geq 2$, if for all $B\in\mathcal{B}$, $f_B(\varphi_i(\tau_i))(\Sigma\times H_c^{n-1})=1$. We have just shown that $\varphi_i(T_i)\subset H_c^1$ for each i (by definition, $H_c^1=H_c$.) Now suppose that $\varphi_j(T_j)\subset H_c^{n-1}$. Then for all $\tau_i\in T_i$, $B\in\mathcal{B}$,

$$f_B(\varphi_i(\tau_i))(\Sigma\times H_c^{n-1})=g_{i,B}(\tau_i)\left(\left\{(\sigma,\tau_j):\varphi_j(\tau_j)\in H_c^{n-1}\right\}\right)=$$

$$g_{i,B}(\tau_i)(\Sigma\times T_j)=1,$$

where the first equality follows from the claim above and the second from the induction hypothesis.

(Continuity) Continuity of φ_i is also proved by induction. Since g_i is continuous and $\varphi_{i,B}^1(\tau_i)=mrg_\Sigma g_{i,B}(\tau_i)$, φ_i^1 is also continuous. Suppose that for $i=1,2$, $k=1,...,n$, φ_i^k is continuous. Then $\psi_{-i}^n(\sigma,\tau_j)=(\sigma,\varphi_j^1(\tau_j),...,\varphi_j^n(\tau_j))$ is continuous in (σ,τ_j). Thus, also $\widehat{\psi_{-i}^n}$ is continuous. Continuity of $\widehat{\psi_{-i}^n}$ and g_i implies that $\varphi_i^{n+1}=\widehat{\psi_{-i}^n}\circ g_i$ is continuous. Thus far we have proved that each φ_i is a continuous mapping from T_i to T and that $g\circ\varphi_i=\widehat{\varphi_{-i}}\circ g_i$. Therefore $(Id_\Sigma,\varphi_1,\varphi_2)$ is a type-morphism from $\mathcal{T}$ to $\mathcal{T}^u$.

(Uniqueness) Suppose that $\phi=(Id_\Sigma,\phi_1,\phi_2)$ is a type-morphism from $\mathcal{T}$ to $\mathcal{T}^u$. We must prove that $\phi=\varphi$. Since $g\circ\phi_i=\widehat{\phi_{-i}}\circ g_i$ and g is invertible, $\phi_i=g^{-1}\circ\widehat{\phi_{-i}}\circ g_i$. Thus we can write the $(n+1)^{th}$ element of $\phi_i(\tau_i)$ as

$$\phi_i^{n+1}(\tau_i)=\left(mrg_{X^n}\widehat{\phi_{-i,B}}(g_i(\tau_i))\right)_{B\in\mathcal{B}},$$

where $\widehat{\phi_{-i,B}}(g_i(\tau_i))$ is the probability measure conditional on $B\times T$ of the CPS $\widehat{\phi_{-i}}(g_i(\tau_i))\in\Delta^{\mathcal{B}}(\Sigma\times T)$. Thus it is sufficient to show that for all $n\geq 0$, $i=1,2$, $B\in\mathcal{B}$, $\tau_i\in T_i$, $mrg_{X^n}\widehat{\phi_{-i,B}}(g_i(\tau_i))=\varphi_{i,B}^{n+1}(\tau_i)$. The statement is true for $n=0$: take a measurable subset $A^0\subset\Sigma:=X^0$, then

$$mrg_{X^0}\widehat{\phi_{-i,B}}(g_i(\tau_i))(A^0)=\widehat{\phi_{-i,B}}(g_i(\tau_i))(A^0\times T)=$$

$$g_{i,B}(\tau_i)\left(\left\{(\sigma,\tau_j):(\sigma,\phi_j(\tau_j))\in A^0\times T\right\}\right)=g_{i,B}(\tau_i)(A^0\times T_j)=$$
$$mrg_\Sigma g_{i,B}(\tau_i)(A^0)=\varphi^1_{i,B}(\tau_i).$$

Suppose that the statement is true for $n=0,...,k-1$. Then

$$\left(\sigma,\left(mrg_{X^0}\widehat{\phi_{-i,B}}(g_i(\tau_i))\right)_{B\in\mathcal{B}},...,\left(mrg_{X^{k-1}}\widehat{\phi_{-i,B}}(g_i(\tau_i))\right)_{B\in\mathcal{B}}\right)=\psi^k_{-i}(\sigma,\tau_j).$$

Take $A^k\subset X^k$ (measurable) and let $A=\mathcal{C}^\infty(A^k)$, then

$$mrg_{X^k}\widehat{\phi_{-i,B}}(g_i(\tau_i))(A^k)=\widehat{\phi_{-i,B}}(g_i(\tau_i))(A)=$$
$$g_{i,B}(\tau_i)\left(\{(\sigma,\tau_j):(\sigma,\phi_j(\tau_j))\in A\}\right)=$$
$$g_{i,B}(\tau_i)\left(\left\{(\sigma,\tau_j):\left(\sigma,\left(mrg_{X^0}\widehat{\phi_{-i,B}}(g_i(\tau_i))\right)_{B\in\mathcal{B}},...,\left(mrg_{X^{k-1}}\widehat{\phi_{-i,B}}(g_i(\tau_i))\right)_{B\in\mathcal{B}}\right)\in A^k\right\}\right)=$$
$$g_{i,B}(\tau_i)\left(\left\{(\sigma,\tau_j):\psi^k_{-i}(\sigma,\tau_j)\in A^k\right\}\right)=\varphi^{k+1}_{i,B}(\tau_i)(A^k).$$

This concludes the proof.

7.3 Proof of Lemma 5

Proof. The statement is trivially true if $\Sigma^*=\emptyset$. Suppose $\emptyset\neq\Sigma^*\subset\rho_{\mathcal{F}}(\Sigma^*)$. Construct $\mathcal{T}$ as follows. Let $T_1\times T_2=\Sigma$. Then, for each i we can construct a function $\lambda_{ij}:T_i\to\Delta^{\mathcal{H}}(\Sigma_j\times T_j)$ such that for all $\tau_i\in\Sigma^*_i$,

$$\tau_i\in r_i(\lambda_{ij}(\tau_i))\quad\text{and}\quad\lambda_{ij,h}(\tau_i)(\Sigma^*_j)=1,\ \forall h\in\mathcal{F}.$$

To complete the definition, fix $\mu_j\in\Delta^{\mathcal{H}}(\Sigma_j)$ and, for $\tau_i\notin\Sigma^*_i$, let $\lambda_{ij}(\tau_i)=\mu_j$.

Also, for any $\tau_i=(s_i,\theta_i)\in T_i=\Sigma_i$, it is always possible to construct a CPS $\lambda_{ii}(\tau_i)$ such that $\lambda_{ii,h}(\tau_i)(S_i\times\{\theta_i\})=1$ for all $h\in\mathcal{H}$, and $\lambda_{ii,h}(\tau_i)(\{\tau_i\})=1$ for all $h\in\mathcal{H}(s_i)$.

$g_i(\cdot)$ is derived from $\lambda_{ii}(\cdot)$ and $\lambda_{ij}(\cdot)$ as follows. First, for $i=1,2$ and for all $\tau_i\in T_i=\Sigma_i$, define a new function $\lambda^d_{ij}:T_i\to[\Delta(\Sigma_j\times T_j)]^{\mathcal{H}}$ by letting

$$\lambda^d_{ij,h}(\tau_i)(\{\sigma_j,\sigma_j\})=\lambda_{ij,h}(\tau_i)(\sigma_j)\quad\forall\sigma_j\in\Sigma_j=T_j,\ h\in\mathcal{H}$$

It is easy to verify that each $\lambda^d_{ij}(\tau_i)$ is indeed a CPS.[28]

[28] Let $D_j=\{(\sigma_j,\tau_j):\tau_j=\sigma_j\in\Sigma_j\}$. Every $\lambda^d_{ij,h}(\tau_i)$ is indeed a probability distribution over $\Sigma_j\times T_j$ (in particular, probabilities add up to one along D_j) which is concentrated on $\Sigma_j(h)\times T_j$ (in particular, on the set $\{(\sigma_j,\sigma_j):\sigma_j\in\Sigma_j(h)\}$.) As for Bayes' rule, for $A_j\subset\Sigma_j(h)\times T_j\subset\Sigma_j(h')\times T_j$, we have

$$\begin{aligned}\lambda^d_{ij,h'}(A_j)&=\lambda_{ij,h'}(\text{proj}_{\Sigma_j}(A_j\cap D_j))=\\&=\lambda_{ij,h}(\text{proj}_{\Sigma_j}(A_j\cap D_j))\cdot\lambda_{ij,h'}(\text{proj}_{\Sigma_j}((\Sigma_j(h)\times T_j)\cap D_j))=\\&=\lambda^d_{ij,h}(A_j)\cdot\lambda^d_{ij,h'}(\Sigma_j(h)\times T_j).\end{aligned}$$

Next, for all $\tau_i \in T_i$, let

$$g_i(\tau_i) = \lambda_{ii}(\tau_i) \otimes \lambda^d_{ij}(\tau_i)$$

Therefore, for every i and τ_i, $g_i(\tau_i) \in I\Delta^{\mathcal{H}}(\Sigma_i, \Sigma_j \times T_j)$. Thus we have a well defined type space, and in each state (i) beliefs are independent, (ii) at any history, players are certain of their payoff-type and (if possible) of their strategy; finally, (iii) properties (i) and (ii) are common certainty at any point in the game.

Moreover, for all h,

$$mrg_{\Sigma_j} g_{i,h}(\tau_i) = \lambda_{ij}(\tau_j).$$

which immediately implies that, for $i = 1, 2$, $\sigma^*_i \in \Sigma^*_i$, $\sigma_j \in \Sigma_j$ and $\tau_j \in T_j$,

$$(\sigma^*_i, \sigma_j, \sigma^*_i, \tau_j) \in R_i = R^1_{i,\mathcal{F}}.$$

That is, $\Sigma^*_i \times \Sigma_j \times \Sigma^*_i \times T_j \subset R^1_{i,\mathcal{F}}$. Assume now that $\Sigma^*_i \times \Sigma_j \times \Sigma^*_i \times T_j \subset R^n_{i,\mathcal{F}}$ for $n \geq 1$ and $i = 1, 2$. Then, since $\beta_{i,\mathcal{F}}$ is monotonic, and for any $\sigma^*_i \in \Sigma^*_i$, $mrg_{\Sigma_j \times T_j} g_i(\sigma^*_i) = \lambda^d_{ij,h}(\sigma^*_i)$ is concentrated on $\{(\sigma^*_j, \sigma^*_j) : \sigma^*_j \in \Sigma^*_j\}$ at all $h \in \mathcal{F}$,

$$\Sigma^*_i \times \Sigma_j \times \Sigma^*_i \times T_j \subset \beta_{i,\mathcal{F}}(\Sigma_i \times \Sigma^*_j \times T_i \times \Sigma^*_j) \subset \beta_{i,\mathcal{F}}(R^n_{j,\mathcal{F}})$$

which implies $\Sigma^*_i \times \Sigma_j \times \Sigma^*_i \times T_j \subset R^{n+1}_{i,\mathcal{F}}$. Hence, $\emptyset \neq \Sigma^*_i \times \Sigma_j \times \Sigma^*_i \times T_j \subset \bigcap_{n \geq 1} R^n_{i,\mathcal{F}} = R_i \cap CCOR_{i,\mathcal{F}}$ for $i = 1, 2$, as required.

References

[1] R. J. Aumann, Backward induction and common knowledge of rationality, *Games Econ. Behav.* **8** (1995), 6-19.

[2] R. J. Aumann, Reply to Binmore, *Games Econ. Behav.* **17** (1996), 138-146.

[3] D. Balkenborg, and E. Winter, A necessary and sufficient epistemic condition for playing backward induction, *J. Math. Econ.* **27** (1997), 325-345.

[4] P. Battigalli and G. Bonanno, Synchronic information and common knowledge in extensive games, *in* "Epistemic Logic and the Theory of Games and Decisions" (M. Bacharach, L.A. Gerard-Varet, P. Mongin, and H. Shin, Eds.), Kluwer, Dordrecht, 1997.

[5] P. Battigalli and M. Siniscalchi, An epistemic characterization of extensive-form rationalizability, Social Science Working Paper 1009, California Institute of Technology, 1997.

[6] P. Battigalli and M. Siniscalchi, Interactive beliefs, epistemic independence and strong rationalizability, *Res. Econ.* **82** (1999), No. 3, forthcoming.

[7] E. Ben Porath, Rationality, Nash equilibrium and backwards induction in perfect information games, *Rev. Econ. Stud.* **64** (1997), 23-46.

[8] D. Bernheim, Rationalizable strategic behavior, *Econometrica* **52** (1984), 1002-1028.

[9] K. Binmore, A note on backward induction, *Games Econ. Behav.* **17** (1996), 135-137.

[10] A. Brandenburger, A. and E. Dekel, Hierarchies of beliefs and common knowledge, *J. Econ. Theory* **59** (1993), 189-198.

[11] E. Dekel and F. Gul, Rationality and knowledge in game theory, *in* "Advances in Economics and Econometrics" (D. Kreps and K. Wallis, Eds.), Cambridge University Press, Cambridge (UK), 1997.

[12] R. M. Dudley, "Real Analysis and Probability," Wadsworth & Brooks, Cole, CA, 1989.

[13] D. Fudenberg and J. Tirole, "Game Theory," MIT Press, Cambridge, MA 1991.

[14] P. Gärdenfors, "Knowledge in Flux," MIT Press, Cambridge, MA, 1988.

[15] J. Harsanyi, Games of incomplete information played by Bayesian players. Parts I, II, III, *Manage. Sci.* **14** (1967-68), 159-182, 320-334, 486-502.

[16] A. Heifetz, A. and D. Samet, Coherent beliefs are not always types, mimeo, Tel Aviv University, Tel Aviv, 1996.

[17] A. Heifetz, A. and D. Samet, Topology-free typology of beliefs, *J. Econ. Theory* **82**, (1998), 324-341.

[18] A. Kechris, "Classical Descriptive Set Theory," Springer Verlag, Berlin, 1995.

[19] J. F. Mertens and S. Zamir, Formulation of Bayesian analysis for games with incomplete information, *Int. J. Game Theory* **14** (1985), 1-29.

[20] R. Myerson, Multistage games with communication, *Econometrica* **54** (1986), 323-358.

[21] M. Osborne A. Rubinstein, "A Course in Game Theory," MIT Press, Cambridge, MA, 1994.

[22] D. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52** (1984), 1029-1050.

[23] P. Reny, Backward Induction, normal form perfection and explicable equilibria, *Econometrica* **60** (1992), 626-649.

[24] P. Reny, Common belief and the theory of games with perfect information, *J. Econ. Theory* **59** (1993), 257-274.

[25] P. Reny, Rational behaviour in extensive form games, *Can. J. Econ.* **28** (1995), 1-16.

[26] A. Rênyi, On a new axiomatic theory of probability, *Acta Mathematica Academiae Scientiarum Hungaricae* **6** (1955), 285-335.

[27] D. Samet, Hypothetical knowledge and games with perfect information, *Games Econ. Behav.* **17** (1996), 230-251.

[28] R. Stalnaker, Knowledge, belief and counterfactual reasoning in games, *Econ. Philos.* **12** (1996), 133-163.

[29] R. Stalnaker, Belief revision in games: forward and backward induction, *Math. Soc. Sci.* **36** (1998), 31-56.

[30] T. Tan and S. Werlang, The Bayesian foundation of solution concepts of games, *J. Econ. Theory* **45** (1988), 370-391.