Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games

Pierpaolo Battigalli
Department of Economics, Princeton University
Princeton NJ 08544-1021
and
I.G.I.E.R.
Via Salasco 5, Milano 20136 (Italy)

This draft: December, 1996
(preliminary, comments welcome)

Abstract

We show how to extend the construction of infinite hierachies of beliefs (Mertens and Zamir (1985), Brandenburger and Dekel (1993)) from the case of probability measures to the case of conditional probability systems (CPSs) defined with respect to a *fixed* collection of relevant hypotheses. The set of hierarchies of CPSs satisfying common certainty of coherency conditional on every relevant hypothesis corresponds to a universal type space. This construction provides a unified framework to analyze the epistemic foundations of solution concepts for dynamic games. As an illustration, we derive some results about conditional common certainty of rationality and rationalizability in multistage games with observed actions.

1. Introduction

Infinite hierachies of beliefs, that is, beliefs about beliefs about beliefs about ..., are used to model interactive epistemic systems and to characterize the epistemic assumptions underlying solution concepts in games (see e.g. Mertens and Zamir (1985), Brandenburger and Dekel (1993) and Tan and Werlang (1988)). In this

paper we extend the standard construction of infinite hierarchies of beliefs and universal beliefs spaces from the case of hierarchies of probability measures to the case of hierarchies of conditional probability systems. In our extension, we follow closely the construction used by Brandenburger and Dekel (1993) (henceforth BD) for the case of probability measures. We show that the space of hierarchies of beliefs satisfying common certainty of coherency is a universal type space, where a(n) (epistemic) type here does not (only) characterize the actual beliefs of an individual, but rather her disposition to have certain beliefs conditional on certain relevant hypotheses.

This framework can be used to clarify the foundations of the theory of dynamic games. One advantage of using this kind of type spaces is that one is able to distiguish between a player's *knowledge* at a decision node x of a game, which corresponds to the information set h containing x, and what she believes with *certainty* at x, that is, what she believes with probability one conditional on h. For example, a player cannot ever *know* that her opponent is rational, because rationality involves a relationship between strategic choices (only partially observable as the play unfolds) and beliefs (unobservable); but her beliefs about her opponent's strategy and beliefs can be such that she is *certain* of her opponent's rationality at some point of the game. Similarly, mutual or common certainty of rationality may obtain at some point of the game. As an illustration of our approach, we derive some some results about common certainty of rationality and rationalizability in multistage games with observed actions.

The rest of the paper is organized as follows. Section 2 defines hierarchies of conditional probability systems and extends known results about hierarchies of probability measures to these more general objects. Section 3 introduces an extended notion of type space and shows that the set of hierarchies of conditional systems satifying common certainty of coherency is a universal type space. Section 4 introduces conditional belief operators. Section 5 uses this framework to derive some results about common certainty of rationality conditional on collections of histories and to provide an epistemic foundation for extensive form rationalizability. Section 6 contains comments about closely related papers.

2. Infinite hierarchies of beliefs

2.1. Conditional Probability Systems and Higher Order Beliefs

For a given Polish (complete, separable, metrizable) space Z, let $\mathcal{A}$ be the Borel sigma-algebra on Z and $\mathcal{B} \subset \mathcal{A}$ a non-empty, countable collection such that $\emptyset \notin \mathcal{B}$.[1] The interpretation is that a certain individual i is uncertain about the "true" element $z \in Z$ and $\mathcal{B}$ represents the collection of "relevant conditions" or "relevant hypothesis." For example, Z may be the set of of sample paths in a repeated experiment or the set of complete histories in a dynamic game and $\mathcal{B}$ may be a set of cylinders. Other interpretations of Z are given below. In particular, Z may be derived from a more basic state space. When we say "relevant" we informally mean that, given the particular problem we are considering, it is interesting, or it makes sense, to ask "what would the beliefs of individual i be if he were given information B?" if and only if $B \in \mathcal{B}$.

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

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

satifying 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 $(Z, \mathcal{A})$.*

Axiom 3. *For all $A, B \in \mathcal{A}$, $C \in \mathcal{B}$ such that $B \cap C \in \mathcal{B}$, $\mu(A|B \cap C)\mu(B|C) = \mu(A \cap B|C)$.*[2]

It can be easily verified that, given Axioms 1 and 2, Axiom 3 can be replaced by the follwing weaker axiom:

Axiom 4. *For all $A \in \mathcal{A}$, $B, C \in \mathcal{B}$, $A \subset B \subset C \Rightarrow \mu(A|B)\mu(B|C) = \mu(A|C)$.*

[1] Alternatively, we may assume that Z is compact and $\mathcal{B}$ is an arbitrary non empty subcollection of the Borel sigma algebra (not containing $\emptyset$).

[2] The tuple $(Z, \mathcal{A}, \mathcal{B}, \mu)$ is called *conditional probability space* by Rènyi (1955). When Z is finite, $\mathcal{A} = 2^Z$, $\mathcal{B} = 2^Z \backslash \{\emptyset\}$, we obtain Myerson's (1986) conditional probability systems.

The set of probability measures on $(Z, \mathcal{A})$ is denoted by $\Delta(Z)$; the set of conditional probability systems on $(Z, \mathcal{A}, \mathcal{B})$ can be regarded as a subset of $[\Delta(Z)]^{\mathcal{B}}$ (the set of mappings from $\mathcal{B}$ to $\Delta(Z)$) and it is denoted by $\Delta^{\mathcal{B}}(Z)$. Accordingly, we often write $\mu = (\mu(\cdot|B))_{B\in\mathcal{B}} \in \Delta^{\mathcal{B}}(Z)$. The topology on Z and $\mathcal{A}$, the smallest sigma-algebra containing it, are always understood and need not be explicit in our notation. Thus we simply say "conditional probability system (or CPS) on $(Z, \mathcal{B})$." It is also understood that $\Delta(Z)$ is endowed with the weak topology and $[\Delta(Z)]^{\mathcal{B}}$ is endowed with the product topology. Thus $\Delta(Z)$ and $[\Delta(Z)]^{\mathcal{B}}$ (by countability of $\mathcal{B}$) are Polish spaces. Since $\Delta^{\mathcal{B}}(Z)$ is a closed subset of $[\Delta(Z)]^{\mathcal{B}}$, also $\Delta^{\mathcal{B}}(Z)$ is a Polish space (endowed with the relative topology inherited from $[\Delta(Z)]^{\mathcal{B}}$). The set $X = Z \times \Delta^{\mathcal{B}}(Z)$ endowed with the product topology is also a Polish space.

Let $\mathcal{C} : \mathcal{B} \to 2^X$ be defined by $\mathcal{C}(B) = B \times \Delta^{\mathcal{B}}(Z)$. Thus $\mathcal{C}(\mathcal{B}) = \{C \subset X : \exists B \in \mathcal{B}, C = B \times \Delta^{\mathcal{B}}(Z)\}$ is a set of "cylinders" generated by $\mathcal{B}$ and represents a copy of $\mathcal{B}$ in X. Then we can define the set of "second order" CPSs $\Delta^{\mathcal{C}(\mathcal{B})}(X)$. Since X is a Polish space, it follows that also $\Delta^{\mathcal{C}(\mathcal{B})}(X)$ (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 $z \in Z$ and $\mu_j \in \Delta^{\mathcal{B}}(Z)$ – individual j's conditional beliefs about $z \in Z$–, whereby the conditions or hypothesis are essentially the same as in $\mathcal{B}$. Recall that $\mathcal{B}$ is the collection of "relevant conditions or hypothesis," therefore it makes sense to condition only on (appropriate copies of) elements of $\mathcal{B}$ when we consider higher order conditional beliefs.

Note that $\Delta^{\mathcal{C}(\mathcal{B})}(X)$ can be regarded as a subset of $[\Delta(X)]^{\mathcal{B}}$, thus we can simplify our notation and write $\Delta^{\mathcal{B}}(X)$ even though $\mathcal{B}$ is not a collection of subsets of X. More formally, let $X = Z \times Q$, $\mathcal{B} \subset 2^X$, $\mathcal{B}_X = \{C \subset X : \exists B \in \mathcal{B}, C = B \times Q\}$; then the set of CPSs on $(X, \mathcal{B}_X)$ is denoted by $\Delta^{\mathcal{B}}(X)$. For any probability measure ν on the product space $X = Z \times Q$ let $mrg_Z\nu \in \Delta(Z)$ denote the marginal measure on X. In what follows it is useful to keep in mind that, if $\mu = (\mu(\cdot|B \times Q))_{B\in\mathcal{B}} \in \Delta^{\mathcal{B}}(X)$, then $(mrg_Z\mu(\cdot|B \times Q))_{B\in\mathcal{B}} \in \Delta^{\mathcal{B}}(Z)$.

2.2. Inductive Construction

We are now ready for the inductive construction of the space of infinite hierarchies of beliefs and the universal type space. For the sake of simplicity, we assume that there are only two individuals i and j with the same space of basic uncertainty Σ and collection of relevant conditions $\mathcal{B}$. They have conditional beliefs about Σ

and about each other for every hypothesis $B \in \mathcal{B}$. For the sake of simplicity, we omit to consider the beliefs of an individual about her own beliefs. As before we assume that Σ is a Polish space and $\mathcal{B}$ is a countable subcollection of its Borel sigma-algebra not containing the empty set. Define recursively X^n and $\mathcal{B}^n$ as follows:

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

for all $n \geq 0$,

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

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

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 hierachy represents an epistemic type and is therefore typically denoted by $t = (\mu^1, \mu^2, ..., \mu^n, ...)$. Note 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 $B \in \mathcal{B}$ and $n = 1, 2, ..., \infty$ let $\mathcal{C}^n(B)$ denote the element of $\mathcal{B}^n$ which is a copy of B in X^n, that is $\mathcal{C}^n(B)$ is the element $C \in \mathcal{B}^n$ such that the projection of C on Σ is B (note that, as the notation suggests, $\mathcal{C}(\mathcal{C}^{n-1}(B)) = \mathcal{C}^n(B)$). Recall that, for any probability measure ν on a product space $X \times Y$, $mrg_X \nu \in \Delta(X)$ denotes the marginal measure on X.

Definition 2.1. *An infinite hierarchy of CPS's $t = (\mu^1, \mu^2, ..., \mu^n, ...)$ is* coherent *if for all $B \in \mathcal{B}$, $n = 1, 2, ...$,*

$$mrg_{X^{n-1}} \mu^{n+1}(\cdot|\mathcal{C}^n(B)) = \mu^n(\cdot|\mathcal{C}^{n-1}(B)). \tag{2.1}$$

The set of coherent hierarchies is denoted by H_c.

Proposition 2.2. *(cf. BD, 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 $B \in \mathcal{B}$, $n = 1, 2, ...,$*

$$mrg_{X^{n-1}} \mu(\cdot | \mathcal{C}^{\infty}(B)) = \mu^n(\cdot | \mathcal{C}^{n-1}(B)). \tag{2.2}$$

We first prove the following lemma:

Lemma 2.3. *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 BD. ■

Proof of Proposition 2.2. For each $B \in \mathcal{B}$, let $\pi_B : H_c \to D$ be the following projection mapping:

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

π_B is clearly continuous. By Lemma 2.3 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. (2.2) 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. (2.2). 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. (2.2). 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 satifies Axiom 4 (coherency of t is satisfied by construction). For each n, let $A^n \subset X^n$ be measurable, $B, C \in \mathcal{B}$, $B^n = \mathcal{C}^n(B)$, $C^n = \mathcal{C}^n(C)$ and suppose that $A^n \subset B^n \subset C^n$ (thus $B \subset C$). Let $A^\infty = \mathcal{C}^\infty(A^n) = A^n \times \Delta^{\mathcal{B}}(X^n) \times \Delta^{\mathcal{B}}(X^{n+1}) \times ... \subset \Sigma \times H$. Similarly $B^\infty = \mathcal{C}^\infty(B)$ and $C^\infty = \mathcal{C}^\infty(C)$. Then A^∞ is measurable in $\Sigma \times H$, $B^\infty, C^\infty \in \mathcal{B}^\infty$ and $A^\infty \subset B^\infty \subset C^\infty$. Thus we can use Axiom 4 for μ and eq. (2.2) to show that $\mu^n(A^n|B^n)\mu^n(B^n|C^n) = \mu^n(A^n|C^n)$.

$\left(f(H_c) \subset \Delta^{\mathcal{B}}(\Sigma \times H)\right)$ Take $t = (\mu_1, ..., \mu_n, ...) \in H_c$ and let $\mu = f(t)$. We must verify that Axiom 4 holds for μ. Let $A^\infty \subset \Sigma \times H$ be measurable, $B, C \in \mathcal{B}$, $B^\infty = \mathcal{C}^\infty(C)$, $C^\infty = \mathcal{C}^\infty(C)$, $A^\infty \subset B^\infty \subset C^\infty$ (thus $B \subset C$). It is sufficient to consider the case where A^∞ is generated as the limit of a sequence of cylinders, that is, there is $\{A^n\}_{n\geq 0}$ such that for each n, $A^n \subset X^n$ is measurable and $A^n \subset \mathcal{C}(A^{n-1}) \subset \mathcal{C}^n(B)$, and $A^\infty = \bigcap_{n\geq 0} \mathcal{C}^\infty(A^n)$. Applying Axiom 4 to μ^n we obtain

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

Then eq. (2.2) yields

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

and by continuity we obtain

$$\mu(A^\infty|B^\infty)\mu(B^\infty|C^\infty) = \mu(A^\infty|C^\infty)$$

as desired. ■

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 $B \in \mathcal{B}$ (note that $\mathcal{B}$ does not contain epistemic events, thus there cannot be any inconsistency in assuming that there is common certainty of coherency conditional on any $B \in \mathcal{B}$).

Individual i endowed with coherent hierarchy of CPSs t_i is *certain* of some (measurable) event $E \subset \Sigma \times H$ (concerning the basic state and/or the other individual's beliefs) *given* $B \in \mathcal{B}$ if $f_B(t_i)(E) = 1$. Thus we can give the following inductive definition of common certainty of coherency given every $B \in \mathcal{B}$:

$H_c^1 = H_c$,
for all $k \geq 2$,
$H_c^k = \{t \in H_c : \forall B \in \mathcal{B}, f_B(t)(\Sigma \times H_c^{k-1}) = 1\}$,
$T = \bigcap_{k \geq 1} H_c^k$.

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

Proposition 2.4. *(cf. BD, 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 $g_B(t) = f_B(t)$ for all $B \in \mathcal{B}$, $t \in T$) such that, if $\mu = g(\mu^1, \mu^2, ...)$, then for all $n \geq 1$, $B \in \mathcal{B}$, $mrg_{X^{n-1}}\mu(\cdot|B \times T) = \mu^n(\cdot|\mathcal{C}^{n-1}(B))$.*

Proof. First verify that $T = \{t \in H_c : \forall B \in \mathcal{B}, f_B(t)(B \times T) = 1\}$. Thus $f(T) = \{\mu \in \Delta^{\mathcal{B}}(\Sigma \times H) : \forall B \in \mathcal{B}, \mu(B \times T | B \times H) = 1\}$, 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 4 and thus g is a homeomorphism between T and $\Delta^{\mathcal{B}}(\Sigma \times T)$ satisfying the marginalization property. ■

Proposition 2.4 shows that each element $t \in T$ corresponds to an epistemic type in the usual sense, except that here a type is associated to a conditional probability system on $(\Sigma \times T, \mathcal{B})$ instead of an ordinary probability measure on $\Sigma \times T$.

3. Type Spaces

Definition 3.1. *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 mapping*

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

where $i \neq j$.

Remark 1. *By Proposition 2.4 if we put $T_1 = T_2 = T$ and $g_1 = g_2 = g$ we obtain a (symmetric) type space which is denoted by $\mathcal{T}^u$.*

A type space is an "implicit representation" of an epistemic model because the sets of types are not derived from the more basic elements Σ and $\mathcal{B}$. On the other hand, hierarchies of beliefs are "explicit" representations of epistemic types.

A typical result about probabilistic epistemic models is that every type space can be regarded as a belief closed subset of the type space explicitly constructed using hierarchies of beliefs, which is therefore "universal".[3] Here we provide such a result for CPSs.

For any given measurable function $\varphi : \Sigma \times T \to \Sigma \times T'$, let $\widehat{\varphi} = (\widehat{\varphi}_B)_{B \in \mathcal{B}} : \Delta^{\mathcal{B}}(\Sigma \times T) \to \Delta^{\mathcal{B}}(\Sigma \times T')$ be the corresponding function associating to each CPS μ on $(\Sigma \times T, \mathcal{B})$ the induced CPS $\mu' = \widehat{\varphi}(\mu)$ on $(\Sigma \times T', \mathcal{B})$. That is, for all $\mu \in \Delta^{\mathcal{B}}(\Sigma \times T)$, $A' \subset \Sigma \times T'$ (measurable), $B \in \mathcal{B}$,

$$\widehat{\varphi}_B(\mu)(A') = \mu(\varphi^{-1}(A')|B \times T).$$

Definition 3.2. *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.

If there is a type-morphism from $\mathcal{T}$ and $\mathcal{T}'$, then $T_1 \times T_2$, up to renaming, corresponds to a belief-closed subset of $T_1' \times T_2'$ and thus $\mathcal{T}$ is essentially a subspace of $\mathcal{T}'$.

Remark 2. *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 $i \in \{1, 2\}$, $\tau_i \in T_i$, $B \in \mathcal{B}$,*

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

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

Definition 3.3. *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 $\mathcal{T}'$.*

[3] See Mertens and Zamir (1985) and Heifetz and Samet (1996a,b). Heifetz and Samet show that, if we drop the topological structure, the space of hierachies of beliefs (satisfying coherency and common certainty of coherency) is "larger" than the set of hierarchies generated by some type space.

Remark 3. *Any two universal type spaces are isomorphic.*

For any type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ there is a pair of canonical mappings (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 $\varphi_i = (\varphi_i^1, \varphi_i^2, ...) = \left[(\varphi_{i,B}^1)_{B\in\mathcal{B}}, (\varphi_{i,B}^2)_{B\in\mathcal{B}}, ...\right]$, $i = 1, 2$ are obtained with a canonical inductive construction:

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

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

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

$$\psi_{-i}^1(\sigma, \tau_j) = (\sigma, \varphi_j^1(\tau_j)),$$

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

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

$$\varphi_{i,B}^{n+1}(\tau_i)(A^n) = g_{i,B}(\tau_i)\left((\psi_{-i}^n)^{-1}(A^n)\right),$$

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

$$\psi_{-i}^{n+1}(\sigma, \tau_j) = \left(\psi_{-i}^n(\sigma, \tau_j), \varphi_j^{n+1}(\tau_j)\right),$$

that is, $\psi_{-i}^{n+1} = (\psi_{-i}^n, \varphi_j^{n+1})$. Thus we have $\varphi_i^{n+1} : T_i \to \Delta^{\mathcal{B}}(X^n)$ and $\psi_{-i}^{n+1} : \Sigma \times T_j \to X^{n+1}$.

Note that $\psi_{-i}^{n+1}(\sigma, \tau_j) = \left(\sigma, \varphi_j^1(\tau_j), ..., \varphi_j^n(\tau_j), \varphi_j^{n+1}(\tau_j)\right)$.

Proposition 3.4. *Let $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ be an arbitrary type space on $(\Sigma, \mathcal{B})$ and let φ_1 and φ_2 be the mappings defined above. 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. ($\varphi_i(T_i) \subset T$) We first verify that $\varphi_i(T_i) \subset H_c$, that is, for all $\tau_i \in T_i$, $n \geq 1$, $B \in \mathcal{B}$, $mrg_{X^{n-1}} \varphi_{i,B}^{n+1}(\tau_i) = \varphi_{i,B}^n(\tau_i)$. Take $A^{n-1} \subset X^{n-1}$ (measurable). Then

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

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

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

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

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

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

The equality $f_B(\varphi_i(\tau_i))(A) = g_{i,B}(\tau_i)\left((\varphi_{-i})^{-1}(A)\right)$ is extended by continuity to the sigma algebra generated by cylinders and 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

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

where $\phi_{-i,B}(g_i(\tau_i))$ 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 $n \geq 0$, $i = 1, 2$, $B \in \mathcal{B}$, $\tau_i \in T_i$, $mrg_{X^n}\phi_{-i,B}(g_i(\tau_i)) = \varphi_{i,B}^{n+1}(\tau_i)$. The statement is true for $n = 0$: take a measurable subset $A^0 \subset \Sigma := X^0$, then

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

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

$$mrg_\Sigma g_{i,B}(\tau_i)(A^0) = \varphi_{i,B}^1(\tau_i).$$

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

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

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

$$mrg_{X^k}\phi_{-i,B}(g_i(\tau_i))(A^k) = \phi_{-i,B}(g_i(\tau_i))(A) =$$

$$g_{i,B}(\tau_i)\left(\{(\sigma,\tau_j) : (\sigma,\phi_j(\tau_j)) \in A\}\right) =$$

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

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

This concludes the proof. ■

4. Conditional Belief Operators

Fix an arbitrary type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$. A point $(\sigma, \tau_1, \tau_2) \in \Sigma \times T_1 \times T_2$ is a *state of the world* and a measurable set $E \subset \Sigma \times T_1 \times T_2$ is an *event*. For each $\tau_i \in T_i$, $E_{\tau_i} \subset \Sigma \times T_j$ is the set of pairs (σ, τ_j) consistent with event E and 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 $\forall B \in \mathcal{F}$, $g_{i,B}(\tau_i)(E_{\tau_i}) \geq p$ (we are implicitly assuming that i is certain of her epistemic type). Thus, for every event E, probability $p \in [0,1]$, and collection of relevant hypotheses $\emptyset \notin \mathcal{F} \subset \mathcal{B}$, the event "$i$ would assign to E a probability of at least p conditional on every $B \in \mathcal{F}$" is

$$\beta^p_{i,\mathcal{F}}(E) := \{(\sigma, \tau_1, \tau_2) : \forall B \in \mathcal{F}, g_{i,B}(\tau_i)(E_{\tau_i}) \geq p\}$$

(note that $\beta^p_{i,\mathcal{F}}(E)$ is measurable for each (measurable) E).

Let $\mathcal{F} \subset \mathcal{B}$ be a collection of commonly observable events (technically, the commonly observable events are the subsets $E = F \times T_1 \times T_2$, $F \in \mathcal{F}$). Then it makes sense to define the event "it would be common p-belief given $\mathcal{F}$ that E"written $c\beta^p_{\mathcal{F}}(E)$. Let $\beta^p_{\mathcal{F}}(E) := \beta^p_{1,\mathcal{F}}(E) \cap \beta^p_{2,\mathcal{F}}(E)$ and $(\beta^p_{\mathcal{F}})^0(E) := E$. For each $n \geq 1$,

$$(\beta^p_{\mathcal{F}})^n(E) := \beta^p_{\mathcal{F}}\left((\beta^p_{\mathcal{F}})^{n-1}(E)\right),$$

then

$$c\beta^p_{\mathcal{F}}(E) := \bigcap_{n \geq 1} (\beta^p_{\mathcal{F}})^n(E).$$

Thus the event "E occurs and it would be common p-belief given $\mathcal{F}$ that E," is

$$E \cap c\beta^p_{\mathcal{F}}(E) = \bigcap_{n \geq 0} (\beta^p_{\mathcal{F}})^n(E).$$

$\beta^p_{i,\mathcal{F}}$, $\beta^p_{\mathcal{F}}$ and $c\beta^p_{\mathcal{F}}$ are examples of conditional belief operators. If $\mathcal{F}$ is a singleton, we replace it with its unique element as a subscript. If $p = 1$ we omit the superscript p. Thus $c\beta_{\mathcal{F}}(\cdot)$ is a conditional common certainty operator. If we have to emphasize the type space $\mathcal{T}$, we add $\mathcal{T}$ as a subscript to the belief operators, e.g. we write $\beta^p_{\mathcal{F},\mathcal{T}}(E)$ and $c\beta^p_{\mathcal{F},\mathcal{T}}(E)$.

Note the difference between the events $\bigcap_{B \in \mathcal{F}} c\beta^p_B(E)$ ("for all $B \in \mathcal{F}$, it would be common p-belief given B that E") and $c\beta^p_{\mathcal{F}}(E)$. In general, $c\beta^p_{\mathcal{F}}(E) \subset \bigcap_{B \in \mathcal{F}} c\beta^p_B(E)$.[4]

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

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

[4]Reny (1993) illustrates this difference in the context of games for the case where E is the event "every player is Bayesian rational." Although he does not explicitly use epistemic models, his analysis and examples can be reformulated within our framework.

Lemma 4.1. *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 for all $n \geq 0$, $\emptyset \notin \mathcal{F} \subset \mathcal{B}$, $p \in [0,1]$,*

$$\varphi\left((\beta^p_{\mathcal{F},\mathcal{T}})^n(E_{\mathcal{T}})\right) \subset (\beta^p_{\mathcal{F},\mathcal{T}'})^n(E_{\mathcal{T}'})$$

and

$$\varphi\left(c\beta^p_{\mathcal{F},\mathcal{T}}(E_{\mathcal{T}})\right) \subset c\beta^p_{\mathcal{F},\mathcal{T}'}(E_{\mathcal{T}'}).$$

Proof. The second statement follows from the first. Since φ is a type-morphism from $\mathcal{T}$ to $\mathcal{T}'$, for all i, τ_i, B, $mrg_{\Sigma} g_{i,B}(\tau_i) = mrg_{\Sigma} g'_{i,B}(\varphi_i(\tau_i))$. This implies $\varphi(E_{\mathcal{T}}) \subset E_{\mathcal{T}'}$. Thus the first stament is true for $n = 0$. Suppose that

$$\varphi\left((\beta^p_{\mathcal{F},\mathcal{T}})^n(E_{\mathcal{T}})\right) \subset (\beta^p_{\mathcal{F},\mathcal{T}'})^n(E_{\mathcal{T}'}).$$

Then Remark 2 implies

$$\varphi\left(\beta^p_{\mathcal{F},\mathcal{T}}\left((\beta^p_{\mathcal{F},\mathcal{T}})^n(E_{\mathcal{T}})\right)\right) \subset \beta^p_{\mathcal{F},\mathcal{T}'}\left((\beta^p_{\mathcal{F},\mathcal{T}'})^n(E_{\mathcal{T}'})\right)$$

and thus the statement is also true for $n+1$. ■

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. Fudenberg and Tirole (1991, Chapter 8)) without the probabilistic structure. Let $\mathcal{H}$, $\mathcal{Z}$ and Θ_i respectively denote the sets of non-terminal and terminal feasible histories, and 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 feasibility constraints and payoff-relevant aspects of the game and has to be distiguished from the epistemic type which specifies i's attitudes to have certain conditional beliefs given certain events. We will omit the adjective "payoff-relevant" whenever no confusion can arise. The set of feasible strategies (mappings

from $\mathcal{H}$ to feasible actions) for type θ_i is denoted $S_i(\theta_i)$. Player i preferences over lotteries are represented by a VNM utility function $u_i : \mathcal{Z} \times \Theta_1 \times \Theta_2 \rightarrow$. The game has *private values* if, for each i, u_i is independent of θ_j; it has *perfect information* if (a) for every history $h \in \mathcal{H}$ there is only one player, say $\iota(h)$, with more than one feasible action and (b) for each i, $S_i(\theta_i)$ is constant.[5]

The basic elements of our analysis are feasible strategy-type pairs: (s_i, θ_i) is a feasible pair if $s_i \in S_i(\theta_i)$. A generic feasible pair for player i is denoted σ_i and the set of such feasible pairs is

$$\Sigma_i := \{(s_i, \theta_i) : \theta_i \in \Theta_i, s_i \in S_i(\theta_i)\}$$

Thus here the basic uncertainty 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 Σ is simply the set of strategy pairs. For each history h, $\Sigma(h) = \Sigma_1(h) \times \Sigma_2(h)$ is the set of σ consistent with the occurrence of h. $\mathcal{H}(\sigma_i)$ is the set of non terminal histories consistent with (the strategy in) σ_i, that is, $\mathcal{H}(\sigma_i) := \{h \in \mathcal{H} : \sigma_i \in \Sigma_i(h)\}$. Considering non terminal histories $h \in \mathcal{H}$, $\Sigma_j(h)$ is a strategic form representation of i's information about j at h. Considering terminal histories $z \in \mathcal{Z}$, we can obtain a strategic form payoff function $U_i : \Sigma \rightarrow$ as follows: for all $z \in \mathcal{Z}$ and $(s_1, \theta_1, s_2, \theta_2) \in \Sigma(z)$, $U_i(s_1, \theta_1, s_2, \theta_2) = u_i(z, \theta_1, \theta_2)$.

We are interested in players' (mutual) conditional beliefs at each (commonly observable) non terminal history h. Thus the collection of relevant hypotheses is 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 $\phi \in \mathcal{H}$ 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 $(\theta_i, g_i(\tau_i)) \in \Theta_i \times \Delta^{\mathcal{B}}(\Sigma \times T_j)$.[6] 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 Ben Porath (1996) (for more on this see Section 6).

[5] The reason why we add (b) to the more familiar condition (a) is that we can represent a game with observed actions with an extensive form featuring perfect information (plus payoff functions $v_i : \Theta \times Z \rightarrow$, where Z is the set of terminal nodes) precisely when (a) *and* (b) are satisfied.

[6] In static games $\Theta_i \times T_i$ is the set of types in the sense of Harsanyi (1967-68). 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.

Since each element of $\mathcal{B}$ represents the event that some history h occurs, we simplify our notation on belief functions and belief operators replacing a strategic form event $B = \Sigma(h) \in \mathcal{B}$ with the corresponding history $h \in \mathcal{H}$ in the subscript denoting the conditioning event. We continue to identify singletons with their unique elemants. For example, given $h \in \mathcal{H}$ or $\mathcal{F} \subset \mathcal{H}$, we write $g_{i,h}(\tau_i) := g_{i,\Sigma(h)}(\tau_i)$, $c\beta_{\mathcal{F}}(E) := c\beta_{\{\Sigma(h):h\in\mathcal{F}\}}(E)$. In particular, the event "there is common certainty of E at the beginning of the game" is denoted $c\beta_{\phi}(E) := c\beta_{\{\Sigma\}}(E)$.

We are formally assuming that a player has beliefs about about her own strategy and payoff-relevant type, but not about her own epistemic type. However, we have already noticed that we implicitly assume that a player is certain of her epistemic type. The same will be assumed for her strategy and payoff-relevant type, that is, we will assume that a player is certain of her strategy and payoff-relevant type at every history consistent with that strategy.

For any $\sigma_i \in \Sigma_i$ let $\mathcal{B}_j(\sigma_i)$ denote the collection of "strategic information sets" concerning player j and corresponding to histories consistent with σ_i, that is

$$\mathcal{B}_j(\sigma_i) := \{B_j : \exists h \in \mathcal{H}(\sigma_i), B_j = \Sigma_j(h)\} .$$

For any pair $(\sigma_i, \mu) \in \Sigma_i \times \Delta^{\mathcal{B}}(\Sigma)$, define the vector $\mu_{\sigma_i} = (\mu_{\sigma_i}(\cdot|B_j))_{B_j\in\mathcal{B}_j(\sigma_i)}$ as follows: for all $B_j = \Sigma_j(h) \in \mathcal{B}_j(\sigma_i)$, all $A_j \subset \Sigma_j$,

$$\mu_{\sigma_i}(A_j|B_j) = \mu(\{\sigma_i\} \times A_j|\Sigma(h)).$$

Remark 4. *If $\mu \in \Delta^{\mathcal{B}}(\Sigma)$ is such that, for all $h \in \mathcal{H}(\sigma_i)$, $\mu(\{\sigma_i\} \times \Sigma_j|\Sigma(h)) = 1$, then μ_{σ_i} is a conditional probability system on $(\Sigma_j, \mathcal{B}_j(\sigma_i))$.*

The assumption that a player is certain of her strategy and type is embedded in the following definition of rationality:

Definition 5.1. *Let $(s_i, \theta_i) \in \Sigma_i$, $\mu \in \Delta^{\mathcal{B}}(\Sigma)$. Strategy s_i is* a sequential best response *to μ for type θ_i, written $(s_i, \theta_i) \in r_i(\mu)$, if for all $h \in \mathcal{H}(s_i, \theta_i)$, $s_i' \in S_i(\theta_i)$*
(a) $\mu(\{(s_i, \theta_i)\} \times \Sigma_j|\Sigma(h)) = 1$,
(b) if $(s_i', \theta_i) \in \Sigma_i(h)$ then

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

Two features of this definition are worth noting. First, we regard strategies as plans of actions, because rationality is imposed only at histories consistent

with the given strategy. Second, according to this particular definition, the set $r_i(\mu)$ of best responses to μ is either empty or a singleton. In fact, suppose $\sigma_i \in r_i(\mu)$. Then, by condition (a), the probability of σ_i conditional on the empty history must be one: $\mu(\{\sigma_i\} \times \Sigma_j | \Sigma(\phi)) = 1$ (recall that $\Sigma(\phi) = \Sigma$). Thus no other σ_i' can satisfy condition (a). We find (a) quite compelling as a rationality condition. In particular, it implies that if $\sigma_i \in r_i(\mu)$, where μ represents the first order conditional beliefs of player i, then player i is certain that she is rational conditional on every history consistent with σ_i. Either feature could be changed (asking for maximization at every history and/or ignoring the beliefs of a player about her strategy) without affecting the following results in any essential way.

5.2. Common Certainty of Rationality

Definition 5.2. *Fix a type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$. Player i is* rational *at state (σ, τ_1, τ_2) in $\mathcal{T}$ if $\sigma_i \in r_i\left((mrg_\Sigma g_{i,h}(\tau_i))_{\Sigma(h) \in \mathcal{B}}\right)$. The set of states in $\mathcal{T}$ where player i is rational is denoted $R_{i,\mathcal{T}}$ and $R_\mathcal{T} := R_{1,\mathcal{T}} \cap R_{2,\mathcal{T}}$. Let $\emptyset \notin \mathcal{F} \subset \mathcal{H}$, $\sigma \in \Sigma$ is* consistent with rationality and common certainty of rationality given $\mathcal{F}$ *if there is an epistemic type space $\mathcal{T}$ and a pair of epistemic types (τ_1, τ_2) such that $(\sigma, \tau_1, \tau_2) \in R_\mathcal{T} \cap c\beta_{\mathcal{F},\mathcal{T}}(R_\mathcal{T})$.*[7]

A few remarks about this definition are worth mentioning.

[7] This is an appropriate point to comment on the assumption that there are observed actions. The analysis can be extended to arbitrary extensive form games. However, the interpretation of the conditional belief operators would change. In the observed actions case, a relevant hypothesis $B = \Sigma(h)$ represents an event that becomes common knowledge when history h occurs. Thus, for example, it is legitimate to interpret the formula $(\sigma, \tau_1, \tau_2) \in c\beta_{h^\ell}(E) \cap \Sigma(h^\ell) \times T_1 \times T_2$, where h^ℓ is a history of length ℓ, as saying that at state (σ, τ_1, τ_2) there is common certainty of E after ℓ periods, that is, when it becomes common knowledge that history h^ℓ (induced by σ) has occurred. This interpretation is incorrect if there is imperfect, asymmetric information about past moves. For some purposes this is irrelevant. For example, we may be interested in what is commonly certain at the beginning of the game (as in Ben Porath (1996)). But if we want to describe the dynamics of interactive beliefs as the play unfolds, we have to augment the standard definition of an extensive form game with a specification of each player's information about past moves at each node of the game, including the decision nodes of other players. This yields information partitions for each stage of the game (see Battigalli and Bonanno (1995)). Let $H_i^\ell(h^\ell)$ denote the information set for player i at stage ℓ after history h^ℓ. Then $\mathcal{B}$ can be chosen as the closure under union of the following collection: $\left\{B : \exists i, \exists h^\ell \in \mathcal{H}, B = \Sigma\left(H_i^\ell(h^\ell)\right)\right\}$. Conditional common belief operators can be meaningfully defined with respect to the elements of the finest common coarsening of the information partitions.

- $R_{\mathcal{T}}$ is the event in $\mathcal{T}$ corresponding to the (measurable) set of strategy-type pairs and first order beliefs for each player satisfying sequential rationality.

- Condition (a) of our definition of rationality implies that if at state $(\sigma_1, \sigma_2, \tau_1, \tau_2)$ player i is rational and is certain of her rationality conditional on history h, then $\sigma_i \in \Sigma_i(h)$. Therefore, if σ is consistent with rationality and common certainty of rationality given h, then $\sigma \in \Sigma(h)$. For the same reason it is impossible to have certainty of one's own rationality given $\mathcal{F}$ whenever $\mathcal{F}$ contains incompatible histories, that is, if $h', h'' \in \mathcal{F}$ and $\Sigma(h') \cap \Sigma(h'') = \emptyset$, there is a player – say i – such that $\Sigma_i(h') \cap \Sigma_i(h'') = \emptyset$ and $\beta_{i,\mathcal{F}}(R_i) = \emptyset$.[8]

- But the following example shows that there is a more interesting reason why, typically, player i cannot be certain of her rationality after every history in $\mathcal{F}$: the fact that i knows her epistemic type.

Example. Consider the following two stage game with complete information. The payoffs of the second stage are independent of the outcome of the first stage:

1^{st} stage	l	r
U	6,0	0,0
M	5,0	5,0
D	0,0	6,0

2^{nd} stage	c	d
a	1,1	0,0
b	0,0	1,1

[8] Let $\rho_i(\tau_i)$ (an empty set or a singleton) denote the set of (s_i, θ_i) such that s_i is a sequential best response to the CPS on $(\Sigma, \mathcal{B})$ induced by $g_i(\tau_i)$. Then

$$R_i \cap \beta_{i,h}(R_i) =$$

$$\{(\sigma, \tau_1, \tau_2) : \{\sigma_i\} = \rho(\tau_i),\ g_{i,h}(\tau_i)(\rho(t_i) \times \Sigma_j \times T_j) = 1\} =$$

$$\{(\sigma, \tau_1, \tau_2) : g_{i,h}(\tau_i)(\{\sigma_i\} \times \Sigma_j \times T_j) = 1\} \subset \Sigma_i(h) \times \Sigma_j \times T_j\,,$$

where the latter inclusion follows from the fact that τ_i assigns probability one to $\Sigma(h)$ conditional on h.

Suppose that $\mathcal{F}$ contains incompatible histories. Then, for at least one player, say i, $\Sigma_i(h') \cap \Sigma_i(h'') = \emptyset$, where $h', h'' \in \mathcal{F}$. We have

$$\beta_{i,\mathcal{F}}(R_i) = \{(\sigma, \tau_1, \tau_2) : \forall h \in \mathcal{F}, g_{i,h}(\tau_i)\,[(R_i)_{\tau_i}] = 1\} =$$

$$\{(\sigma, \tau_1, \tau_2) : \forall h \in \mathcal{F}, g_{i,h}(\tau_i)\,(\rho_i(\tau_i) \times \Sigma_j \times T_j) = 1\} \subset$$

$$\{(\sigma, \tau_1, \tau_2) : \forall h \in \mathcal{F}, \rho_i(\tau_i) \cap \Sigma_i(h) \neq \emptyset\}\,.$$

Since $\bigcap_{h \in \mathcal{F}} \Sigma_i(h) = \emptyset$ and $\rho_i(\tau_i)$ is (at most) a singleton, $\beta_{i,\mathcal{F}}(R_i) = \emptyset$.

Suppose that $\mathcal{F} = \mathcal{H}$ and the epistemic type τ_1 of player 1 (the Row player) is such that she is certain of her rationality after U. Then no strategy choosing D in the first stage can be a best response to the belief of τ_1 about player 2. But after history (D, l) or (D, r) type τ_1 must assign positive probability only to such strategies. Thus no type τ_1 can be certain that she is best responding after U *and* after D.□

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

(i) When we consider the set of strategy-type pairs consistent with common certainty of rationality given $\mathcal{F}$ can we restrict our attention to *finite* type spaces (more generally, type spaces with the same cardinality of Σ)?

(ii) 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?

(iii) How can we characterize the set of type-strategy pairs consistent with common certainty of rationality given $\mathcal{F}$ without any reference to epistemic types?

We start from the last question. The answer should rely on an inductive construction. For any $\Lambda \subset \Delta^{\mathcal{B}}(\Sigma)$, let $r(\Lambda) := r_1(\Lambda) \times r_2(\Lambda)$. For any $\widehat{\Sigma} \subset \Sigma$, $\mathcal{F} \subset \mathcal{H}$, let

$$\Lambda_{\mathcal{F}}(\widehat{\Sigma}) := \left\{\mu \in \Delta^{\mathcal{B}}(\Sigma) : \forall h \in \mathcal{F}, \mu(\widehat{\Sigma}|\Sigma(h)) = 1\right\}.$$

(Note that, if $\mathcal{F}$ is "large" and $\widehat{\Sigma}$ is "small," $\Lambda_{\mathcal{F}}(\widehat{\Sigma})$ is typically empty.) The inductive construction is as follows:

- $\Sigma^0_{\mathcal{F}} := \Sigma$,
- for all $n \geq 0$, $\Sigma^{n+1}_{\mathcal{F}} := r\left[\Lambda_{\mathcal{F}}(\Sigma^n_{\mathcal{F}})\right]$.

Note that $\Sigma^1_{\mathcal{F}}$ is independent of $\mathcal{F}$ and, for every type space $\mathcal{T}$ on $(\Sigma, \mathcal{B})$,

$$\left(\left\{(\sigma, \mu_1, \mu_2) \in \Sigma^1_{\mathcal{F}} \times \Delta^{\mathcal{B}}(\Sigma) \times \Delta^{\mathcal{B}}(\Sigma) : \sigma_i = r_i(\mu_i), i = 1, 2\right\}\right)_{\mathcal{T}} = R_{\mathcal{T}}.$$

A perhaps obvious conjecture is that the set of σ consistent with rationality and common certainty of rationality given $\mathcal{F}$ is $\Sigma^\infty_{\mathcal{F}} := \bigcap_{n\geq 1} \Sigma^n_{\mathcal{F}}$. But the example above shows that this conjecture is not correct in general: one can check that in that example $\Sigma^\infty_{\mathcal{B}} = \Sigma$ even if $c\beta_{\mathcal{B}}(R) = \emptyset$. We will see that, in general, $\Sigma^\infty_{\mathcal{F}}$ is a superset of of the set of σ consistent with common certainty of rationality given $\mathcal{F}$ and that the conjecture is essentially correct when $\mathcal{F}$ is a singleton.

Since the composite mapping $r \circ \Lambda_{\mathcal{F}}$ is a monotone set to set operator, the sequence $\{\Sigma^n_{\mathcal{F}}\}^\infty_{n=0}$ is (weakly) decreasing. Thus, by finiteness of Σ, there is some N such that for all $n \geq N$, $\Sigma^\infty_{\mathcal{F}} := \Sigma^n_{\mathcal{F}}$. This means that $\Sigma^\infty_{\mathcal{F}}$ has the familiar

fixed point property: $\Sigma_{\mathcal{F}}^{\infty} = r\left[\Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}}^{\infty})\right]$. It is easy to prove (using monotonicity of $r \circ \Lambda_{\mathcal{F}}$) that every rectangular subset $\Sigma_{\mathcal{F}}$ such that $\Sigma_{\mathcal{F}} \subset r\left[\Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}})\right]$ is a subset of $\Sigma_{\mathcal{F}}^{\infty}$. In general, $\Sigma_{\mathcal{F}}^{\infty}$ may well be empty (cf. Reny (1993) and the related comments in the next section). But it can be shown that Σ_{ϕ}^{∞} is nonempty.[9] Given the fixed point property of $\Sigma_{\mathcal{F}}^{\infty}$ it is easy to verify that $\Sigma_{\mathcal{F}}^{\infty} \neq \emptyset$ if and only if $\Sigma_{\mathcal{F}} \cap (\bigcup_{h \in \mathcal{F}} \Sigma(h)) \neq \emptyset$.

Lemma 5.3. *Let $\Sigma^* = \Sigma_1^* \times \Sigma_2^* \subset \Sigma$, $h^* \in \mathcal{H}$. If $\Sigma^* \subset r\left[\Lambda_{h^*}(\Sigma^*)\right]$, then there is a* finite *type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ such that*

$$\forall \sigma \in \Sigma^* \cap \Sigma(h^*), \exists (\tau_1, \tau_2) \in T_1 \times T_2,\ (\sigma, \tau_1, \tau_2) \in R_{\mathcal{T}} \cap c\beta_{h^*,\mathcal{T}}(R_{\mathcal{T}}).$$

Proof. The statement is trivially true if $\Sigma^* \cap \Sigma(h^*) = \emptyset$. Suppose $\emptyset \neq \Sigma^* \cap \Sigma(h^*) \subset r\left[\Lambda_{h^*}(\Sigma^*)\right]$. Construct $\mathcal{T}$ as follows. Let $T_1 \times T_2 = \Sigma^* \cap \Sigma(h^*)$. Then, for each $\tau_i \in T_i = \Sigma_i^* \cap \Sigma_i(h^*)$, we can choose a CPS $\lambda_i(\tau_i) = (\lambda_{i,h}(\tau_i))_{\Sigma(h) \in \mathcal{B}} \in \Lambda_{h^*}(\Sigma^*)$ such that $\tau_i \in r_i(\lambda_i(\tau_i))$. For all $\tau_i \in T_i \subset \Sigma_i(h^*)$,

$$\lambda_i(\tau_i)\left(\{\tau_i\} \times T_j\right) = 1, \tag{5.1}$$

because $\lambda_i(\tau_i) \in \Lambda_{h^*}(\Sigma^*)$ and τ_i and $\lambda_i(\tau_i)$ satisfy part (a) of Definition 5.1. Mapping $g_i(\cdot)$ is derived from $\lambda_i(\cdot)$ as follows:

Let $\ell(h^*, \tau_i)$ be the first history h such that $\lambda_{i,h}(\tau_i)(\Sigma(h^*)) > 0$, that is

$$\ell(h^*, \tau_i) := \min_{\preceq} \left\{h \in \mathcal{H} : \lambda_{i,h}(\tau_i)(\Sigma(h^*)) > 0\right\},$$

where $\preceq$ is the (reflexive) precedence relation over $\mathcal{H}$. Note that, by definition, $\ell(h^*, \tau_i) \preceq h^*$; thus $T_1 \times T_2 \subset \Sigma(h^*) \subset \Sigma(\ell(h^*, \tau_i))$. For all $\tau_i \in T_i$, $\sigma \in \Sigma$, $\tau_j \in T_j$, let

$$g_{i,\ell(h^*,\tau_i)}(\tau_i)(\sigma, \tau_j) = \begin{cases} \lambda_{i,\ell(h^*,\tau_i)}(\tau_i)(\sigma)/\left|T_j\right| & \text{if } \sigma \notin \Sigma(h^*) \\ \lambda_{i,\ell(h^*,\tau_i)}(\tau_i)(\sigma), & \text{if } (\tau_1, \tau_2) = \sigma \\ 0, & \text{if } (\tau_1, \tau_2) \neq \sigma \in \Sigma(h^*) \end{cases}.$$

[9] 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 $\mu_{ij} \in \Delta^{\mathcal{B}_j}(\Sigma_j)$ such that $\mu_{ij}(\Sigma_j^* | \Sigma_j) = 1$, and for each θ_i we can find a strategy $s_i \in S_i(\theta_i)$ such that condition (b) in the definition of sequential best response is satisfied. Then we can construct a CPS $\mu_i \in \Delta^{\mathcal{B}}(\Sigma)$ such that $\mu_i(\{(s_i, \theta_i)\} \times \Sigma_j^* | \Sigma) = 1$. Thus $(s_i, \theta_i) \in r_i(\mu)$. When we apply this construction to $\Sigma^* = \Sigma$, we obtain that $\Sigma_{\phi}^1 \neq \emptyset$. When we apply the construction to $\Sigma^* = \Sigma_{\phi}^1$, we necessarily have $(s_i, \theta_i) \in \Sigma_{i,\phi}^1$. Thus $\mu_i(\Sigma_{\phi}^1 | \Sigma) = 1$, which implies $(s_i, \theta_i) \in \Sigma_{i,\phi}^2$. An induction argument yields the result.

where $|T_j|$ is the cardinality of T_j. (This is possible by equation 5.1.) Taking into account that $\lambda_i(\tau_i)$ is a CPS and $\lambda_{i,h^*}(\tau_i)(\Sigma^*) = 1$, it can be checked that

$$\forall \sigma \in \Sigma, \sum_{\tau_j \in T_j} g_{i,\ell(h^*,\tau_i)}(\tau_i)(\sigma, \tau_j) = \lambda_i(\tau_i)(\sigma)$$

and

$$\sum_{\sigma,\tau_j} g_{i,\ell(h^*,\tau_i)}(\tau_i)(\sigma, \tau_j) = 1.$$

The second equation follows from the first, which we verify only for the less obvious case. Let $\sigma \in \Sigma(h^*)$, $\sigma_j \notin T_j$, $\lambda_i(\tau_i) = \mu(\cdot|\cdot)$. Then $\sum_{\tau_j \in T_j} g_{i,\ell(h^*,\tau_i)}(\tau_i)(\sigma, \tau_j) = 0$. We must show that $\mu(\sigma|\Sigma(\ell(h^*, \tau_i))) = 0$. Since $\sigma \in \Sigma(h^*) \subset \Sigma(\ell(h^*, \tau_i))$,

$$\mu(\sigma|\Sigma(\ell(h^*, \tau_i))) = \mu(\sigma|\Sigma(h^*))\mu(\Sigma(h^*)|\Sigma(\ell(h^*, \tau_i)))$$

But $\mu(\sigma|\Sigma(h^*)) = 0$ because $\sigma_j \notin T_j$ and $\mu(\{\tau_i\} \times T_j) = 1$.

For $h \neq \ell(h^*, \tau_i)$, if $\lambda_{i,\ell(h^*,t_i)}(\tau_i)(\Sigma(h)) > 0$ (which implies $\ell(h^*) \preceq h$), then $g_{i,h}(\tau_i)$ can be derived from $g_{i,\ell(h^*)}(\tau_i)$ via Bayes rule. Otherwise,

$$g_{i,h}(\tau_i)(\sigma, \tau_j) = \frac{\lambda_{i,h}(\tau_i)(\sigma)}{|T_j|}.$$

It can be checked that $g_i(\tau_i) \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$ and for all h, $mrg_{\Sigma} g_{i,h}(\tau_i) = \lambda_{i,h}(\tau_i)$. Furthermore,

$$g_{i,h^*}(\tau_i)(\sigma, \tau_j) = \begin{cases} \lambda_{i,h^*}(\tau_i)(\sigma), & \text{if } (\tau_1, \tau_2) = \sigma \\ 0, & \text{if } (\tau_1, \tau_2) \neq \sigma \in \Sigma(h^*) \end{cases}$$

Thus, for every $(\tau_1, \tau_2) \in T_1 \times T_2 = \Sigma^* \cap \Sigma(h^*)$, $(\tau_1, \tau_2, \tau_1, \tau_2) \in R_{\mathcal{T}}$ and, for each i,

$$1 = g_{i,h^*}(\tau_i)\left(\left\{(\sigma', \tau_j') : \sigma_i' = \tau_i, \sigma_j' = \tau_j')\right\}\right) \leq$$
$$\leq g_{i,h^*}(\tau_i)\left((R_{\mathcal{T}})_{\tau_i}\right).$$

An easy induction argument shows that for every $\sigma = (\sigma_1, \sigma_2) \in \Sigma^* \cap \Sigma(h^*) = T_1 \times T_2$, $(\sigma, \sigma_1, \sigma_2) \in R_{\mathcal{T}} \cap c\beta_{h^*,\mathcal{T}}(R_{\mathcal{T}})$. ■

Proposition 5.4. *Fix $\emptyset \neq \mathcal{F} \subset \mathcal{H}$. (a) The set of $\sigma \in \Sigma$ consistent with rationality and common certainty of rationality given $\mathcal{F}$ is contained in $\Sigma^{\infty}_{\mathcal{F}}$.*
(b) For all $h \in \mathcal{H}$, there is a finite *type space $\mathcal{T}$ such that, for all $\sigma \in \Sigma(h)$, $\sigma \in \Sigma^{\infty}_h$ if and only if there is some pair of types (τ_1, τ_2) such that $(\sigma, \tau_1, \tau_2) \in$*

$R_{\mathcal{T}} \cap c\beta_{h,\mathcal{T}}(R_{\mathcal{T}})$. Since $\Sigma(\phi) = \Sigma$, this implies that Σ_{ϕ}^{∞} is the set of σ consistent with rationality and common certainty of rationality at the beginning of the game.
(c) For all $\sigma \in \Sigma$, σ is consistent with rationality and common certainty of 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 $(\sigma, t_1, t_2) \in R_{\mathcal{T}^u} \cap c\beta_{\phi,\mathcal{T}^u}(R_{\mathcal{T}^u})$.
(d) For all $h \in \mathcal{H}$ and $\sigma \in \Sigma(h)$, there is a pair of hierarchies of CPSs $(t_1, t_2) \in T \times T$ such that $(\sigma, t_1, t_2) \in R_{\mathcal{T}^u} \cap c\beta_{h,\mathcal{T}^u}(R_{\mathcal{T}^u})$ if and only if $\sigma \in \Sigma_h^{\infty}$.

Proof. (a) Fix a type space $\mathcal{T}$ on $(\Sigma, \mathcal{B})$. We show that for all $n \geq 0$, $(\sigma, \tau_1, \tau_2) \in \bigcap_{k=0}^{k=n} (\beta_{\mathcal{F}})^k(R_{\mathcal{T}})$ implies $\sigma \in \Sigma_{\mathcal{F}}^n$. This implies the thesis. The statement is true by definition for $n = 0$. Assume it is true for n. Let

$$(\sigma, \tau_1, \tau_2) \in \bigcap_{k=0}^{k=n+1} (\beta_{\mathcal{F}})^k(R_{\mathcal{T}}) := R_{\mathcal{T}} \cap \left[\bigcap_{k=0}^{k=n} \beta_{\mathcal{F}} \left((\beta_{\mathcal{F}})^k(R_{\mathcal{T}}) \right) \right].$$

Let $\mu = (mrg_{\Sigma} g_{i,h}(\tau_i))_{\Sigma(h) \in \mathcal{B}}$. Then $\sigma_i \in r_i(\mu)$ and

$$\forall h \in \mathcal{F}, \forall k \in \{0, ..., n\}, \; g_{i,h}(\tau_i) \left(\left((\beta_{\mathcal{F}})^k(R_{\mathcal{T}}) \right)_{\tau_i} \right) = 1,$$

that is,

$$\forall h \in \mathcal{F}, \; g_{i,h}(\tau_i) \left(\left\{ (\sigma', \tau_j') : (\sigma', \tau_i, \tau_j') \in \bigcap_{k=0}^{k=n} (\beta_{\mathcal{F}})^k(R_{\mathcal{T}}) \right\} \right) = 1.$$

Therefore the induction hypothesis implies

$$\forall h \in \mathcal{F}, \; g_{i,h}(\tau_i)(\Sigma_{\mathcal{F}}^n \times T_j) = \mu(\Sigma_{\mathcal{F}}^n | \Sigma(h)) = 1$$

and $\sigma_i \in \Sigma_{i,\mathcal{F}}^{n+1}$.

(b) The "if" part follows from (a). The "only if" part follows from Lemma 5.3, because $\Sigma_h^{\infty} = r\left[\Lambda_h(\Sigma_h^{\infty})\right]$.

(c) The "if" part is true by definition. The "only if" part is a consequence of Proposition 3.4 and Lemma 4.1.

(d) This is a consequence of (b) and (c). ■

5.3. Common Certainty of the Opponent's Rationality

We have seen that $\Sigma_{\mathcal{F}}^{\infty}$ may be larger than the set of σ consistent with rationality and common certainty of rationality given $\mathcal{F}$, if $\mathcal{F}$ is not a singleton. The reason

is that a player of a given type cannot be certain of her own rationality after incompatible histories. But in game theory we are interested in the following question (among others): "What might player i do if she is rational and (1) she believes that her opponent is rational, (2) she believes that her opponent believes that she is rational, (3) she believes that her opponent believes that she believes that her opponent is rational, (4) ...?" In other words we ask for the consequences of rationality and common certainty of the *opponent*'s rationality.

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:

$$c\beta_{ij,\mathcal{F}}(R_1, R_2) := \beta_{i,\mathcal{F}}(R_j) \cap \beta_{i,\mathcal{F}}(\beta_{j,\mathcal{F}}(R_i)) \cap \beta_{i,\mathcal{F}}(\beta_{j,\mathcal{F}}(\beta_{i,\mathcal{F}}(R_j))) \cap ... =$$

$$\left[\bigcap_{n\geq 0} \beta_{i,\mathcal{F}}\left((\beta_{j,\mathcal{F}} \circ \beta_{i,\mathcal{F}})^n(R_j)\right)\right] \cap \left[\bigcap_{n\geq 1} \left((\beta_{i,\mathcal{F}} \circ \beta_{j,\mathcal{F}})^n(R_i)\right)\right].$$

Definition 5.5. *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*

$$(\sigma, \tau_1, \tau_2) \in [R_1 \cap c\beta_{12,\mathcal{F}}(R_1, R_2)] \cap [R_2 \cap c\beta_{21,\mathcal{F}}(R_1, R_2)].$$

Proposition 5.6. *Let $\emptyset \neq \mathcal{F} \subset \mathcal{H}$. The set of σ consistent with rationality and common certainty of the opponent's rationality given $\mathcal{F}$ is precisely $\Sigma_{\mathcal{F}}^{\infty}$.*

One can find examples where some σ is consistent with rationality and common knowledge of the opponent's rationality given h even if σ and h are incompatible ($\sigma \notin \Sigma(h)$). By Proposition 5.6 the same examples show that $\Sigma_{\mathcal{F}}$ need not be contained in $\bigcup_{h\in\mathcal{F}} \Sigma(h)$.

The proof of Proposition 5.6 relies on two lemmata. The first characterizes $\Sigma_{\mathcal{F}}^{\infty}$ in a way which is intuitively related to common certainty of the opponent's rationality. Define $\widehat{\Sigma}_{i,\mathcal{F}}^{n}$ inductively as follows:

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

That is, $\widehat{\Sigma}_{i,\mathcal{F}}^{n+1}$ is the set of (s_i, θ_i) such that s_i is a sequential best response for θ_i to some CPS μ satisfying $\mu(\Sigma_i \times \widehat{\Sigma}_{j,\mathcal{F}}^n | \Sigma(h)) = 1$ for all $h \in \mathcal{F}$. Note that each sequence of subsets $\left\{\widehat{\Sigma}_{i,\mathcal{F}}^n\right\}_{n\geq 0}$ is (weakly) decreasing, because $r_i \circ \Lambda_{\mathcal{F}}$ is a monotone set to set operator.

Lemma 5.7. *For all* $n \geq 0$, $\widehat{\Sigma}_{1,\mathcal{F}}^n \times \widehat{\Sigma}_{2,\mathcal{F}}^n = \Sigma_{1,\mathcal{F}}^n \times \Sigma_{2,\mathcal{F}}^n$.

Proof. The statement is true for $n = 0$. Suppose that

$$\widehat{\Sigma}_{i,\mathcal{F}}^n = \Sigma_{i,\mathcal{F}}^n,\ i = 1, 2.$$

Then $\Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}}^n) \subset \Lambda_{\mathcal{F}}(\Sigma_i \times \widehat{\Sigma}_{j,\mathcal{F}}^n)$ and

$$\Sigma_{i,\mathcal{F}}^{n+1} = r_i\left[\Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}})\right] \subset r_i\left[\Lambda_{\mathcal{F}}(\Sigma_i \times \widehat{\Sigma}_{j,\mathcal{F}}^n)\right] = \widehat{\Sigma}_{i,\mathcal{F}}^{n+1}.$$

We have to show that $\widehat{\Sigma}_{1,\mathcal{F}}^{n+1} \times \widehat{\Sigma}_{2,\mathcal{F}}^{n+1} \subset \Sigma_{1,\mathcal{F}}^{n+1} \times \Sigma_{2,\mathcal{F}}^{n+1}$. If $\widehat{\Sigma}_{1,\mathcal{F}}^{n+1} \times \widehat{\Sigma}_{2,\mathcal{F}}^{n+1} = \emptyset$, there is nothing to prove: by the inclusion above, both sets are empty. Thus suppose that $\widehat{\Sigma}_{1,\mathcal{F}}^{n+1} \times \widehat{\Sigma}_{2,\mathcal{F}}^{n+1} \neq \emptyset$ and let $\sigma_i \in \widehat{\Sigma}_{i,\mathcal{F}}^{n+1}$. Then there is some $\mu \in \Lambda_{\mathcal{F}}(\Sigma_i \times \widehat{\Sigma}_{j,\mathcal{F}}^n)$ such that $\sigma_i \in r_i(\mu)$. Since $\left\{\widehat{\Sigma}_{i,\mathcal{F}}^k\right\}_{k\geq 0}$ is a decreasing sequence, the inductive hypothesis implies $\sigma_i \in \widehat{\Sigma}_{i,\mathcal{F}}^{n+1} \subset \Sigma_{i,\mathcal{F}}^n$. Using again the inductive hypothesis (for j) and part (a) of Definition 5.1, it follows that

$$\forall h \in \mathcal{H}(\sigma_i),\ 1 = \mu(\{\sigma_i\} \times \Sigma_{j,\mathcal{F}}^n | \Sigma(h)) \leq \mu(\Sigma_{\mathcal{F}}^n | \Sigma(h)).$$

Now we construct $\nu \in \Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}}^n)$ such that $\sigma_i \in r_i(\nu)$, this shows that $\sigma_i \in \Sigma_{i,\mathcal{F}}^{n+1}$.

We first note that the inductive hypothesis implies that, if $\widehat{\Sigma}_{1,\mathcal{F}}^{n+1} \times \widehat{\Sigma}_{2,\mathcal{F}}^{n+1} \neq \emptyset$, then $\Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}}^n) \neq \emptyset$. To prove the contrapositive of this statement, suppose that $\Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}}^n) = \emptyset$. Then there is $h \in \mathcal{F}$ such that $\Sigma(h) \cap \Sigma_{\mathcal{F}}^n = \emptyset$. By the inductive hypothesis

$$\Sigma_1(h) \times \Sigma_2(h) \cap (\widehat{\Sigma}_{1,\mathcal{F}}^n \times \widehat{\Sigma}_{2,\mathcal{F}}^n) = \emptyset,$$

that is

$$\exists i \in \{1, 2\},\ \Sigma_i(h) \cap \widehat{\Sigma}_{i,\mathcal{F}}^n = \emptyset.$$

But this implies that $\Lambda_{\mathcal{F}}(\widehat{\Sigma}_{i,\mathcal{F}} \times \Sigma_j) = \emptyset$ and thus $\widehat{\Sigma}_{j,\mathcal{F}}^{n+1} = r_j\left[\Lambda_{\mathcal{F}}(\widehat{\Sigma}_{i,\mathcal{F}} \times \Sigma_j)\right] = \emptyset$. Therefore we can assume that there is some $\mu' \in \Lambda_{\mathcal{F}}(\Sigma_{\mathcal{F}}^n)$. We derive ν from μ and μ':

$$\forall h \in \mathcal{H},\ \nu(\cdot|h) = \begin{cases} \mu(\cdot|\Sigma(h)), & \text{if } h \in \mathcal{H}(\sigma_i) \\ \mu'(\cdot|\Sigma(h)), & \text{if } h \notin \mathcal{H}(\sigma_i) \end{cases}.$$

Since ν and μ coincide on $\mathcal{H}(\sigma_i)$, $\sigma_i \in r_i(\nu)$. Taking into account the properties of μ, it can be checked that ν is a CPS and, for all $h \in \mathcal{H}$, $\nu(\Sigma^n_{\mathcal{F}}|\Sigma(h)) = 1$, i.e $\nu \in \Lambda_{\mathcal{F}}(\Sigma^n_{\mathcal{F}})$. ■

Lemma 5.8. *Let $\Sigma^* = \Sigma^*_1 \times \Sigma^*_2 \subset \Sigma$, $\emptyset \neq \mathcal{F} \subset \mathcal{H}$. If $\Sigma^* \subset r\left[\Lambda_{\mathcal{F}}(\Sigma^*)\right]$, then there is a* finite *type space $\mathcal{T} = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ such that*

$$\forall \sigma \in \Sigma^*, \exists (\tau_1, \tau_2) \in T_1 \times T_2,$$

$$(\sigma, \tau_1, \tau_2) \in \left[R_1 \cap c\beta_{12,\mathcal{F}}(R_1, R_2)\right] \cap \left[R_2 \cap c\beta_{21,\mathcal{F}}(R_1, R_2)\right].$$

Proof. This proof is similar to the proof of Lemma 5.3. The statement is trivially true if $\Sigma^* = \emptyset$. Suppose $\emptyset \neq \Sigma^* \subset r\left[\Lambda_{\mathcal{F}}(\Sigma^*)\right]$. Construct $\mathcal{T}$ as follows. Let $T_1 \times T_2 = \Sigma^*$. Then, for each i we can construct a mapping $\lambda_i : T_i \to \Delta^{\mathcal{B}}(\Sigma)$ such that for all $\tau_i \in T_i = \Sigma^*_i$, $h \in \mathcal{F}$,

$$\tau_i \in r_i(\lambda_i(\tau_i)),\ \lambda_{i,h}(\tau_i)(T_1 \times T_2) = 1.$$

$g_i(\cdot)$ is derived from $\lambda_i(\cdot)$ as follows:

Recall that for any history h^* and type τ_i, $\ell(h^*, \tau_i)$ denotes the least element $h \preceq h^*$ such that $\lambda_{i,h}(\tau_i)(\Sigma(h^*)) > 0$. Let $\ell(\mathcal{F}, \tau_i)$ be the image of $\mathcal{F}$ through mapping $\ell(\cdot, \tau_i)$, that is

$$\ell(\mathcal{F}, \tau_i) := \{h \in \mathcal{H} : \exists h^* \in \mathcal{F}, \lambda_{i,h}(\tau_i)(\Sigma(h^*)) > 0\}.$$

We first define $g_{i,h}(\tau_i)(\cdot, \cdot)$ for histories $h \in \ell(\mathcal{F}, \tau_i)$. For any such history h let $\mathcal{U}(h, \tau_i)$ be the union of the "strategic information sets" corresponding to histories in $\ell(\mathcal{F}, \tau_i)$ following h, i.e.

$$\mathcal{U}(h, \tau_i) := \bigcup_{h^* \in \mathcal{F}, h = \ell(h^*, \tau_i)} \Sigma(h^*).$$

Then $\forall \tau_i \in T_i, \forall h \in \ell(\mathcal{F}, \tau_i), \forall \sigma \in \Sigma, \forall \tau_j \in T_j$,

$$g_{i,h}(\tau_i)(\sigma, \tau_j) = \begin{cases} \lambda_{i,h}(\tau_i)(\sigma)/\left|T_j\right| & \text{if } \sigma \notin \mathcal{U}(h, \tau_i) \\ \lambda_{i,h}(\tau_i)(\sigma), & \text{if } \sigma \in \mathcal{U}(h, \tau_i), \tau_j = \sigma_j \\ 0, & \text{if } \sigma \in \mathcal{U}(h, \tau_i), \tau_j \neq \sigma_j \end{cases}.$$

This definition of $g_{i,h}(\tau_i)$ and the properties of $\lambda_i(\tau_i)$ imply that

$$\forall \sigma \in \Sigma, \sum_{\tau_j \in T_j} g_{i,h}(\tau_i)(\sigma, \tau_j) = \lambda_{i,h}(\tau_i)(\sigma)$$

and thus

$$\sum_{\sigma,\tau_j} g_{i,h}(\tau_i)(\sigma,\tau_j) = 1.$$

We verify the first of these equations only for the less obvious case. Suppose that $\sigma \in \mathcal{U}(h,\tau_i)$, $\sigma_j \notin T_j$. In this case $\sum_{\tau_j \in T_j} g_{i,h}(\tau_i)(\sigma,\tau_j) = 0$. Since $\sigma \in \mathcal{U}(h,\tau_i)$, there is some $h^* \in \mathcal{F}$ such that $\sigma \in \Sigma(h^*) \subset \Sigma(h)$. Let $\lambda_i(\tau_i) = \mu(\cdot|\cdot)$. Then

$$\lambda_{i,h}(\tau_i)(\sigma,\tau_i) = \mu(\sigma|\Sigma(h)) = \mu(\sigma|\Sigma(h^*))\mu(\Sigma(h^*)|\Sigma(h)) = 0,$$

because $\sigma_j \notin T_j$ and $\mu(T_1 \times T_2|\Sigma(h^*)) = 1$.

Fix τ_i and $h' \in \mathcal{H}$. Either there is a *unique* $h \in \ell(\mathcal{F},\tau_i)$ such that $h \preceq h'$ and $\lambda_{i,h}(\tau_i)(\Sigma(h)) > 0$, or there no such h at all. We have just proved that for all h' and all $h \in \ell(\mathcal{F},\tau_i)$, $g_{i,h}(\tau_i)(\Sigma(h') \times T_j) = \lambda_{i,h}(\Sigma(h'))$. Therefore, in the first case we can derive $g_{i,h'}(\tau_i)$ from $g_{i,h}(\tau_i)$; otherwise, let

$$\forall \sigma \in \Sigma, \forall \tau_j \in T_j,\ g_{i,h'}(\tau_i)(\sigma,\tau_j) = \frac{\lambda_{i,h'}(\tau_i)(\sigma)}{|T_j|}. \tag{5.2}$$

It can be checked that for every i and τ_i, $g_i(\tau_i) \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$ (thus we have a well defined type space) and for all h, $mrg_{\Sigma} g_{i,h}(\tau_i) = \lambda_{i,h}(\tau_i)$. Furthermore, for all $h^* \in \mathcal{F}$,

$$\forall \sigma \in \Sigma, \forall \tau_j \in T_j,\ g_{i,h^*}(\tau_i)(\sigma,\tau_j) = \begin{cases} \lambda_{i,h^*}(\tau_i)(\sigma), & \text{if } (\tau_1,\tau_2) = \sigma \\ 0, & \text{if } (\tau_1,\tau_2) \neq \sigma \end{cases}.$$

It follows from this construction that in this type space

$$\forall \sigma \in \Sigma, \forall (\tau_1,\tau_2) \in T_1 \times T_2,$$

$$(\tau_1,\sigma_2,\tau_1,\tau_2) \in R_1, (\sigma_1,\tau_2,\tau_1,\tau_2) \in R_2 \tag{5.3}$$

and

$$\forall (\tau_1,\tau_2) \in T_1 \times T_2, \forall i,j \in \{1,2\}, i \neq j, \forall h \in \mathcal{F},$$

$$g_{i,h}(\tau_i)\left(\left\{(\sigma',\tau_j') : \sigma_j' = \tau_j'\right\}\right) = 1. \tag{5.4}$$

Equations (5.3) and (5.4) imply that

$$\beta_{1,\mathcal{F}}(R_2) = \Sigma \times T_1 \times T_2 = \beta_{2,\mathcal{F}}(R_1),$$

which in turn implies that common certainty of the opponent's rationality given $\mathcal{F}$ always holds:

$$\Sigma \times T_1 \times T_2 = c\beta_{12,\mathcal{F}}(R_1, R_2) \cap c\beta_{21,\mathcal{F}}(R_1, R_2). \tag{5.5}$$

Therefore, if for any $\sigma \in \Sigma^* = T_1 \times T_2$ we choose $(\tau_1, \tau_2) = \sigma$, from (5.3) and (5.5) we get

$$(\sigma, \tau_1, \tau_2) \in [R_1 \cap c\beta_{12,\mathcal{F}}(R_1, R_2)] \cap [R_1 \cap c\beta_{12,\mathcal{F}}(R_1, R_2)]$$

as desired. ■

Proof of Proposition 5.6. Since $\Sigma^\infty_{\mathcal{F}} = r\left[\Lambda_{\mathcal{F}}(\Sigma^\infty_{\mathcal{F}})\right]$, Lemma 5.8 implies that every σ in $\Sigma^\infty_{\mathcal{F}}$ is consistent with rationality and common certainty of the opponent rationality. The opposite inclusion is proved in a way similar to the proof of Proposition 5.4 (a) making use of Lemma 5.7. ■

5.4. Strong Beliefs and Extensive Form Rationalizability

Pearce (1984) has defined a notion of "extensive form rationalizability," which can be reformulated as follows:[10]

- $\Sigma^0 := \Sigma$, $\Lambda^0 = \Delta^{\mathcal{B}}(\Sigma)$
- for $n = 0, 1, , ...,$ $\Sigma^{n+1} := r(\Lambda^n)$,

  $$\Lambda^{n+1} := \left\{\mu \in \Delta^{\mathcal{B}}(\Sigma) : \forall k = 0, ..., n, \forall h \in \mathcal{H}, \Sigma(h) \cap \Sigma^k \neq \emptyset \Rightarrow \mu(\Sigma^k | \Sigma(h)) = 1\right\}.$$

Write $\Sigma^n = \Sigma^n_1 \times \Sigma^n_2$. Σ^1_i is the set of sequentially rational strategy-types σ_i. Σ^2_i is meant to represent the set of σ_i that are sequentially rational given that player i continues to believe that (R1) everybody is rational, as long as (R1) does not contradict the evidence. Σ^3_i is meant to represent the subset of such σ_i that are sequentially rational given that player i continues to believe that (R2) everybody is rational and everybody continues to believe that everybody is rational as long as this does not contradict the evidence, as long as (R2) does not

[10] The analysis can be extended to cover all games with (incomplete information and) perfect recall. As before Σ denotes the set of feasible profiles of types and strategies. $\mathcal{H}$ is the collection of information sets. $\mathcal{B}$ is the closure under union of the collection $\{\Sigma(h) : h \in \mathcal{H}\}$ of "strategic form information sets" or any larger collection which is closed under union and does not contain the empty set, e.g. $2^\Sigma \setminus \{\emptyset\}$. See also Reny (1992) and Battigalli (1996a,b) for similar solution procedures.

contradict the evidence. And so on. Clearly, $\Sigma_i^{n+1} \subset \Sigma_i^n$. It is possible to show by standard arguments that for all n, Σ^n and Λ^n are non empty. Finiteness of Σ implies that there is an integer N such that $\Sigma^N = \Sigma^n$ for all $n \geq N$. Σ^N is the set of *extensive form rationalizable* strategies. Extensive form rationalizability provides a (non-equilibrium) formalization of the foward induction principle and is outcome-equivalent to subgame perfection in generic games of perfect and complete information (Battigalli (1996b)).

Does the intended interpretation of this solution concept correspond to an appropriate formulation based on type spaces?

Fix a type space $\mathcal{T}$ based on $(\Sigma, \mathcal{B})$ and consider the following notion of "strong belief":

Definition 5.9. *For any event E and type τ_i in type space $\mathcal{T}$, we say that type τ_i* strongly believes *E if for all histories $h \in \mathcal{H}$,*

$$E_{\tau_i} \cap (\Sigma(h) \times T_j) \neq \emptyset \Rightarrow g_{i,h}(\tau_i)(E_{\tau_i}) = 1.$$

Let $\beta_i^*(E)$ denote the event that player i strongly believes E and let $\beta^*(E)$ denote the event that everybody strongly believes E, that is:

- $\beta_i^*(E) := \{(\sigma, \tau_i, \tau_j) : \forall h \in \mathcal{H}, E_{\tau_i} \cap (\Sigma(h) \times T_j) \neq \emptyset \Rightarrow g_{i,h}(\tau_i)(E_{\tau_i}) = 1\}$,
- $\beta^*(E) := \beta_1^*(E) \cap \beta_2^*(E)$.

Note that, unlike standard epistemic operators, the strong belief operator β_i^* is not monotone.

The event that everybody is rational and strongly believes in rationality is $R \cap \beta^*(E)$. According to its intended interpretation Σ^2 should be the projection on Σ of this event. Similarly, Σ^3 should be the projection on Σ of the conjunction of $R \cap \beta^*(E)$ with the event that everybody strongly believes $R \cap \beta^*(E)$, that is, Σ^3 should be the projection of event $R \cap \beta^*(E) \cap \beta^*[R \cap \beta^*(E)]$. For any event E, let

$$\gamma(E) := E \cap \beta^*(E).$$

Iterations of operator γ are defined in the usual way. In particular we obtain the following identities:

$$\gamma^0(R) = R,$$
$$\gamma^1(R) = R \cap \beta^*(R),$$

$$\gamma^2(R) = \gamma\left[R \cap \beta^*(R)\right] = R \cap \beta^*(R) \cap \beta^*\left[R \cap \beta^*(R)\right],$$

$$\ldots .$$

We continue to specify (when necessary) the given type space as a subscript of events and operators writing, for example, $R_{\mathcal{T}}$, $\beta^*_{\mathcal{T}}(R_{\mathcal{T}})$, $\gamma^n_{\mathcal{T}}(R_{\mathcal{T}})$.

Remark 5. *By inspection of the definitions above*

$$\gamma^n(R) = \left[R_1 \cap \left(\bigcap_{k=0}^{n-1} \beta^*_1(\gamma^k(R))\right)\right] \cap \left[R_2 \cap \left(\bigcap_{k=0}^{n-1} \beta^*_2(\gamma^k(R))\right)\right].$$

Therefore $(\sigma, \tau_1, \tau_2) \in \gamma^n(R)$ *if and only if, for each player* i, $\sigma_i \in r_i\left((mrg_{\Sigma} g_{i,h}(\tau_i))_{\Sigma(h)\in\mathcal{B}}\right)$ *and* τ_i *strongly believes* $\gamma^k(R)$ *for all* $k = 0, ..., n-1$.

We would like to show that Σ^{n+1} is the projection on Σ of $\gamma^{n+1}(R)$. But it easy to find examples of games and type spaces $\mathcal{T}$ where this is not the case. The reason is simple: $\mathcal{T}$ may have "too few types." If, for example, each T_i is a singleton, then event $R_{\mathcal{T}}$ and its projection on Σ are also singletons.[11] If Σ^1 contains more than one element, it cannot be the projection of $R_{\mathcal{T}}$. We stipulate that the intended interpretation of Σ^{n+1} is appropriate if, for every "sufficiently rich" type space $\mathcal{T}$ (including the universal type space $\mathcal{T}^u$), Σ^{n+1} is the projection of $\gamma^n_{\mathcal{T}}(R_{\mathcal{T}})$ on Σ.

We first construct a "smallest" type space $\mathcal{T}^* = (\Sigma, \mathcal{B}, T_1, T_2, g_1, g_2)$ where this is indeed the case. For each player i, let

$$T_i = \Sigma^1_i.$$

By definition there is a mapping $\lambda_i : \Sigma^1_i \to \Delta^{\mathcal{B}}(\Sigma)$ such that for all $n = 0, 1, ...$, $\tau_i \in T_i = \Sigma^1_i$,

$$\tau_i \in \Sigma^{n+1}_i \Rightarrow \left[\lambda_i(\tau_i) \in \Lambda^n \text{ and } \tau_i \in r_i(\lambda_i(\tau_i))\right].$$

Note that if there is some h such that $0 < \lambda_{i,h}(\tau_i)(\Sigma^1) < 1$, then $\lambda_{i,h}(\tau_i) \notin \Lambda^1$, and it must be the case that $\tau_i \in \Sigma^1_i \backslash \Sigma^2_i$. If $\tau_i \in \Sigma^2_i$, then $\lambda_{i,h}(\tau_i) \in \Lambda^1$ and $\lambda_{i,h}(\tau_i)(\Sigma^1) \in \{0, 1\}$ for all h. Construct $g_i : T_i \to \Delta^{\mathcal{B}}(\Sigma \times T_j)$ as follows: for all $\tau_i \in T_i$, $(\sigma_1, \sigma_2) \in \Sigma$, $\tau_j \in T_j$, $h \in \mathcal{H}$,

[11] While this is *necessarily* true for our definition of rationality, it is only *typically* true for other plausible definitions (up to equivalences between strategies). See our comments to the definition of rationality.

- (a) if $\tau_i \in \Sigma_i^2$ and $\lambda_{i,h}(\tau_i)(\Sigma^1) = 1$, then

$$g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \tau_j) = \begin{cases} \lambda_{i,h}(\tau_i)(\sigma_1, \sigma_2), & \text{if } \sigma_j = \tau_j \\ 0, & \text{if } \sigma_j \neq \tau_j \end{cases},$$

- (b) if either $\tau_i \in \Sigma_i^1 \backslash \Sigma_i^2$ or $\lambda_{i,h}(\tau_i)(\Sigma^1) < 1$, then

$$g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \tau_j) = \frac{1}{|T_j|} \lambda_{i,h}(\tau_i)(\sigma_1, \sigma_2).$$

To show that, for each τ_i, $g_i(\tau_i) \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$, we first prove that $mrg_{\Sigma} g_{i,h}^*(\tau_i) = \lambda_{i,h}(\tau_i)$. This is obvious in case (b). In case (a), if $\sigma_j \in T_j$, then

$$\sum_{\tau_j \in T_j} g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \tau_j) =$$

$$g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \sigma_j) + \sum_{\tau_j \in T_j \backslash \{\sigma_j\}} g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \tau_j) =$$

$$g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \sigma_j) = \lambda_{i,h}(\tau_i)(\sigma_1, \sigma_2).$$

If $\sigma_j \notin T_j$, then both $\lambda_{i,h}(\tau_i)(\sigma_1, \sigma_2)$ and $g_{i,h}(\tau_i)(\sigma_1, \sigma_2, \tau_j)$ are equal to zero.

We only have to show that $g_i(\tau_i)$ satisfies Axiom 4, that is, whenever $\sigma \in \Sigma(h') \subset \Sigma(h)$ $(h \preceq h')$,

$$g_{i,h}(\tau_i)(\sigma, \tau_j) = g_{i,h'}(\tau_i)(\sigma, \tau_j) g_{i,h}(\tau_i)(\Sigma(h') \times T_j), \tag{5.6}$$

for all $\tau_j \in T_j$. Recall that $\lambda_i(\tau_i)$ is a CPS on $(\Sigma, \mathcal{B})$ and each $\lambda_{i,h}(\tau_i)$ is the marginal of $g_{i,h}(\tau_i)$ on Σ $(\Sigma(h) \in \mathcal{B})$. Thus Eq. (5.6) is obvious if $\tau_i \notin \Sigma_i^2$, because we are always in case (b). Thus suppose that $\tau_i \in \Sigma_i^2$ and $\sigma \in \Sigma(h') \subset \Sigma(h)$. Consider the following cases:

- $\Sigma^1 \cap \Sigma(h') \neq \emptyset$. Then $\lambda_{i,h'}(\tau_i)(\Sigma^1) = \lambda_{i,h}(\tau_i)(\Sigma^1) = 1$ and case (a) applies for h and h'. Either $\sigma_j \neq \tau_j$ and both sides of Eq. (5.6) are equal to zero, or $\sigma_j = \tau_j$ and

$$g_{i,h}(\tau_i)(\sigma, \tau_j) = \lambda_{i,h}(\tau_i)(\sigma) =$$

$$\lambda_{i,h'}(\tau_i)(\sigma) \lambda_{i,h}(\tau_i)(\Sigma(h')) = g_{i,h'}(\tau_i)(\sigma, \tau_j) g_{i,h}(\tau_i)(\Sigma(h') \times T_j).$$

- $\Sigma^1 \cap \Sigma(h') = \emptyset$ and $\Sigma^1 \cap \Sigma(h) \neq \emptyset$. Then $\lambda_{i,h}(\tau_i)(\Sigma^1) = 1$ and $\lambda_{i,h}(\tau_i)(\sigma) = \lambda_{i,h}(\tau_i)(\Sigma(h')) = 0$. This implies that both sides of Eq. (5.6) are equal to zero.

- $\Sigma^1 \cap \Sigma(h') = \Sigma^1 \cap \Sigma(h) = \emptyset$. Then case (b) applies to h and h' and the result is obvious.

Thus $\mathcal{T}^*$ is indeed a type space.

Lemma 5.10. *Fix the type space $\mathcal{T}^*$. For all $n = 0, 1, ...$ and $\sigma = (\sigma_1, \sigma_2) \in \Sigma$, $(\sigma, \tau_1, \tau_2) \in \gamma^n(R)$ if and only if $(\sigma_1, \sigma_2) = (\tau_1, \tau_2)$ and $\sigma \in \Sigma^{n+1}$.*

Proof. The statement is obviously true for $n = 0$. Suppose by way of induction that the statement is true for $n = k-1$. We must show that $(\sigma_1, \sigma_2, \sigma_1, \sigma_2) \in \gamma^k(R)$ if and only if $(\sigma_1, \sigma_2) \in \Sigma^{k+1}$.

Let $(\sigma_1, \sigma_2, \sigma_1, \sigma_2) \in \gamma^k(R)$. Recall that $\gamma^k(R) = \gamma^{k-1}(R) \cap \beta^*(\gamma^{k-1}(R))$. Therefore $(\sigma_1, \sigma_2, \sigma_1, \sigma_2) \in \gamma^{k-1}(R)$ and by the induction hypothesis $(\sigma_1, \sigma_2) \in \Sigma^k$. Thus, for each i, $\lambda_i(\sigma_i) \in \Lambda^{k-1}$ and $\sigma_i \in r_i(\lambda_i(\sigma_i))$. We show that, for all $h \in \mathcal{H}$, if $\Sigma(h) \cap \Sigma^k \neq \emptyset$ and $\sigma_i \in \Sigma_i(h)$, then $\lambda_{i,h}(\sigma_i)(\Sigma^k) = 1$. This implies that we can find some CPS $\mu \in \Lambda^k$ which coincides with $\lambda_i(\sigma_i)$ at all such histories (take any two histories $h \prec h'$ such that $\sigma_i \in \Sigma_i(h)$, $\Sigma(h) \cap \Sigma^k \neq \emptyset$ and either $\sigma_i \notin \Sigma_i(h')$ or $\Sigma(h') \cap \Sigma^k = \emptyset$, then $\lambda_{i,h}(\sigma_i)(\Sigma(h')) = 0$ and this allows to modify $\lambda_i(\sigma_i)$ at histories like h' obtaining $\mu \in \Lambda^k$). It follows that $\sigma_i \in r_i(\mu)$ and $\sigma_i \in \Sigma_i^{k+1}$.

Thus assume that $\Sigma(h) \cap \Sigma^k \neq \emptyset$ and $\sigma_i \in \Sigma_i(h)$ (hence $\sigma_i \in \Sigma_i(h) \cap \Sigma_i^k$). By the inductive hypothesis $\gamma^{k-1}(R)$ is the diagonal of $\Sigma^k \times \Sigma^k$. Therefore

$$(\Sigma(h) \times T_j) \cap (\gamma^{k-1}(R))_{\sigma_i} =$$

$$\left\{(\sigma_1', \sigma_2', \tau_j) : \sigma_i' \in \Sigma_i(h) \cap \Sigma_i^k \cap \{\sigma_i\}, \tau_j = \sigma_j' \in \Sigma_j(h) \cap \Sigma_j^k\right\} \neq \emptyset.$$

Since $(\sigma_1, \sigma_2, \sigma_1, \sigma_2) \in \beta_i^*(\gamma^{k-1}(R))$,

$$1 = g_{i,h}(\sigma_i)\left((\gamma^{k-1}(R))_{\sigma_i}\right) \leq g_{i,h}(\sigma_i)(\Sigma^k \times T_j) = \lambda_{i,h}(\sigma_i)(\Sigma^k).$$

Let $(\sigma_1, \sigma_2) \in \Sigma^{k+1}$. Then $(\sigma_1, \sigma_2) \in \Sigma^k$ and by the induction hypothesis $(\sigma_1, \sigma_2, \sigma_1, \sigma_2) \in \gamma^{k-1}(R)$. We only have to show that, for each player i, $(\sigma_1, \sigma_2, \sigma_1, \sigma_2) \in \beta_i^*(\gamma^{k-1}(R))$, that is, for each $h \in \mathcal{H}$, if $(\Sigma(h) \times T_j) \cap (\gamma^{k-1}(R))_{\sigma_i} \neq \emptyset$, then $g_{i,h}(\sigma_i)\left((\gamma^{k-1}(R))_{\sigma_i}\right) = 1$. Suppose that $(\Sigma(h) \times T_j) \cap (\gamma^{k-1}(R))_{\sigma_i} \neq \emptyset$. Since $\gamma^{k-1}(R)$ is the diagonal of $\Sigma^k \times \Sigma^k$, $\sigma_i \in \Sigma_i(h)$ and $\Sigma(h) \cap \Sigma^k \neq \emptyset$. Taking into account (1) that $\lambda_i(\sigma_i) \in \Lambda^k$, (2) that $\lambda_{i,h}(\sigma_i)(\{\sigma_i\} \times \Sigma_j) = 1$ (because $\sigma_i \in \Sigma_i(h)$) and (3) the definition of $g_{i,h}(\sigma_i)$ in case (a), we get

$$1 = \lambda_{i,h}(\sigma_i)(\Sigma^k) = \sum_{\sigma_j' \in \Sigma_j^k} \lambda_{i,h}(\sigma_i)(\sigma_i, \sigma_j') =$$

$$\sum_{\sigma'_j \in \Sigma_j^k} g_{i,h}(\sigma_i)(\sigma_i, \sigma'_j, \sigma'_j) = g_{i,h}(\sigma_i)\left((\gamma^{k-1}(R))_{\sigma_i}\right).$$

This concludes the proof. ■

Lemma 5.11. *Suppose that* $\varphi = (Id_\Sigma, \varphi_1, \varphi_2)$ *is a type-morphism from* $\mathcal{T}^*$ *to* $\mathcal{T}' = (\Sigma, \mathcal{B}, T'_1, T'_2, g'_1, g'_2)$. *Then for all* $n = 0, 1, ...$
(a.n) $\varphi(\gamma^n_{\mathcal{T}^*}(R_{\mathcal{T}^*})) \subset \gamma^n_{\mathcal{T}'}(R_{\mathcal{T}'})$ *and*
(b.n) $\forall \sigma \in \Sigma$, $\sigma \in \Sigma^{n+1}$ *if and only if* $\exists (\tau_1, \tau_2) \in T'_1 \times T'_2$, $(\sigma, \tau_1, \tau_2) \in \gamma^n_{\mathcal{T}'}(R_{\mathcal{T}'})$.

Proof. The statement is obviously true for $n = 0$. Suppose by way of induction that the statement is true for $n = k - 1$. We first prove that (a.k) $\varphi\left(\gamma^k_{\mathcal{T}^*}(R_{\mathcal{T}^*})\right) \subset \gamma^k_{\mathcal{T}'}(R_{\mathcal{T}'})$.

Let $(\sigma, \tau_1, \tau_2) \in \gamma^k_{\mathcal{T}^*}(R_{\mathcal{T}^*})$. By Lemma 5.10, it must be the case that $(\tau_1, \tau_2) = \sigma$ (recall that in $\mathcal{T}^*$ $\sigma_i \in r_i\left((mrg_\Sigma g_{i,h}(\tau_i))_{\Sigma(h)\in\mathcal{B}}\right)$ iff $(mrg_\Sigma g_{i,h}(\tau_i))_{\Sigma(h)\in\mathcal{B}} = \lambda_i(\sigma_i)$ and $\sigma_i = \tau_i$) and $\sigma \in \Sigma^{k+1}$. We must show that $(\sigma, \varphi_1(\sigma_1), \varphi_2(\sigma_2)) \in \gamma^k_{\mathcal{T}'}(R_{\mathcal{T}'})$. Since $\gamma^k_{\mathcal{T}'}(R_{\mathcal{T}'}) = \gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}) \cap \beta^*(\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}))$ and by the inductive hypothesis $(\sigma, \varphi_1(\sigma_1), \varphi_2(\sigma_2)) \in \gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'})$, we only have to show that, for each i, $(\sigma, \varphi_1(\sigma_1), \varphi_2(\sigma_2)) \in \beta^*_i(\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}))$. Thus suppose that $(\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}))_{\varphi_i(\sigma_i)} \cap (\Sigma(h) \times T'_j) \neq \emptyset$. We first show that this implies $\sigma_i \in \Sigma_i(h)$. In fact, by the inductive hypothesis the projection of $\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'})$ on Σ is Σ^k. Furthermore, by definition $\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}) \subset R_{\mathcal{T}'}$ and $\left(mrg_\Sigma g'_{i,h}(\varphi_i(\sigma_i))\right)_{\Sigma(h)\in\mathcal{B}} = \lambda_i(\sigma_i)$. This implies:

$$(\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}))_{\varphi_i(\sigma_i)} \cap (\Sigma(h) \times T'_j) :=$$

$$\left\{(\sigma', \tau'_j) : (\sigma'_i, \sigma'_j, \varphi_i(\sigma_i), \tau'_j) \in \gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}) \cap (\Sigma_i(h) \times \Sigma_j(h) \times T'_i \times T'_j)\right\} \subset$$

$$\subset \left\{(\sigma', \tau'_j) : \sigma'_i \in r_i(\lambda_i(\sigma_i)), \sigma' \in \Sigma(h) \cap \Sigma^k\right\} =$$

$$\left\{(\sigma', \tau'_j) : \sigma'_i = \sigma_i, \sigma' \in \Sigma(h) \cap \Sigma^k\right\}.$$

Since $(\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}))_{\varphi_i(\sigma_i)} \cap (\Sigma(h) \times T'_j) \neq \emptyset$, it must be the case that $\sigma_i \in \Sigma_i(h)$. Using the inductive hypothesis and Lemma 5.10, it follows that $(\gamma^{k-1}_{\mathcal{T}^*}(R_{\mathcal{T}^*}))_{\sigma_i} \cap (\Sigma(h) \times T_j) \neq \emptyset$. Therefore

$$g_{i,h}(\sigma_i)\left((\gamma^{k-1}_{\mathcal{T}^*}(R_{\mathcal{T}^*}))_{\sigma_i}\right) = 1.$$

By the inductive hypothesis $\varphi(\gamma^{k-1}_{\mathcal{T}^*}(R_{\mathcal{T}^*})) \subset \gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'})$. By Remark 2, this inclusion and the equality above imply $g_{i,h}(\varphi_i(\sigma_i))\left((\gamma^{k-1}_{\mathcal{T}'}(R_{\mathcal{T}'}))_{\varphi_i(\sigma_i)}\right) = 1$ as desired. This concludes the proof that $\varphi\left(\gamma^k_{\mathcal{T}^*}(R_{\mathcal{T}^*})\right) \subset \gamma^k_{\mathcal{T}'}(R_{\mathcal{T}'})$.

We now prove that (bk) $\forall \sigma \in \Sigma$, $\sigma \in \Sigma^{k+1}$ if and only if $\exists (\tau_1, \tau_2) \in T_1' \times T_2'$, $(\sigma, \tau_1, \tau_2) \in \gamma^k_{T'}(R_{T'})$. Suppose that $\sigma \in \Sigma^{k+1}$. Then Lemma 5.10 and (a.k) imply that $(\sigma, \varphi_1(\sigma_1), \varphi_2(\sigma_2)) \in \gamma^k_{T'}(R_{T'})$. Suppose that $(\sigma, \tau_1, \tau_2) \in \gamma^k_{T'}(R_{T'})$. We must show that $\sigma \in \Sigma^{k+1}$, that is, for each player i, there is some $\mu \in \Lambda^k$ such that $\sigma_i \in r_i(\mu)$. Take an arbitrary CPS μ' in the (non empty) set Λ^k. Construct μ as follows. For all $h \in \mathcal{H}$,

$$\mu(\cdot \mid \Sigma(h)) = \begin{cases} marg_{\Sigma} g'_{i,h}(\tau_i), & \text{if } \sigma_i \in \Sigma_i(h) \text{ and } \Sigma(h) \cap \Sigma^k \neq \emptyset \\ \mu'(\cdot \mid \Sigma(h)), & \text{if } \sigma_i \notin \Sigma_i(h) \text{ and } \Sigma(h) \cap \Sigma^k \neq \emptyset \\ \lambda_{i,h}(\sigma_i), & \text{if } \Sigma(h) \cap \Sigma^k = \emptyset \end{cases} .$$

One can verify that, by construction, $\mu \in \Delta^{\mathcal{B}}(\Sigma)$, $\sigma_i \in r_i(\mu)$ and for all $h \in \mathcal{H}$ and $\ell = 1, ..., k-1$, if $\Sigma(h) \cap \Sigma^k = \emptyset$ and $\Sigma(h) \cap \Sigma^\ell \neq \emptyset$, then $\mu(\Sigma^\ell \mid \Sigma(h)) = 1$. We only have to show that if $\Sigma(h) \cap \Sigma^k \neq \emptyset$ then $\mu(\Sigma^k \mid \Sigma(h)) = 1$.

Suppose that $\Sigma(h) \cap \Sigma^k \neq \emptyset$ and $\sigma_i \notin \Sigma_i(h)$. Then $\mu(\Sigma^k \mid \Sigma(h)) = \mu'(\Sigma^k \mid \Sigma(h)) = 1$.

Suppose that $\Sigma(h) \cap \Sigma^k \neq \emptyset$ and $\sigma_i \in \Sigma_i(h)$. By the inductive hypothesis Σ^k is the projection of $\gamma^{k-1}_{T'}(R_{T'})$ on Σ. Therefore there is some $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\tau}_1, \hat{\tau}_2)$ such that

$$(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\tau}_1, \hat{\tau}_2) \in (\Sigma(h) \times T_1 \times T_2) \cap \gamma^{k-1}_{T'}(R_{T'}) \neq \emptyset.$$

By assumption $\sigma_i \in r_i \left((marg_{\Sigma} g'_{i,h}(\tau_i))_{\Sigma(h) \in \mathcal{B}} \right)$, $\sigma_i \in \Sigma_i(h)$ and τ_i strongly believes $\gamma^\ell_{T'}(R_{T'})$ for all $\ell = 0, 1, ..., k-2$. The same holds for $\hat{\sigma}_j$ and $\hat{\tau}_j$. Given the Cartesian structure of event $\gamma^{k-1}_{T'}(R_{T'})$ (see Remark 5), it follows that

$$(\sigma_i, \hat{\sigma}_j, \tau_i, \hat{\tau}_j) \in (\Sigma_i(h) \times \Sigma_j(h) \times T_i' \times T_j') \cap \gamma^{k-1}_{T'}(R_{T'}).$$

Therefore

$$(\Sigma(h) \times T_j') \cap (\gamma^{k-1}_{T'}(R_{T'}))_{\tau_i} \neq \emptyset.$$

Since by assumption τ_i strongly believes $\gamma^{k-1}_{T'}(R_{T'})$, we obtain

$$g'_{i,h}(\tau_i) \left((\gamma^{k-1}_{T'}(R_{T'}))_{\tau_i} \right) = 1.$$

By the inductive hypothesis Σ^k contains the projection on Σ of $(\gamma^{k-1}_{T'}(R_{T'}))_{\tau_i}$. Thus the equality above and the definition of μ yield

$$\mu(\Sigma^k | \Sigma(h)) = g'_{i,h}(\tau_i)(\Sigma^k \times T_j') = 1$$

as desired. ■

As an immediate consequence of Lemma 5.11 and Proposition 3.4 we obtain the following result:

Proposition 5.12. *Consider the universal type space* $\mathcal{T}^u = (\Sigma, \mathcal{B}, T, T, g, g)$. *For all* $n = 0, 1, ...,$ *for all* $\sigma \in \Sigma$, $\sigma \in \Sigma^{n+1}$ *if and only if there is a pair of hierarchies of CPSs* $(t_1, t_2) \in T \times T$ *such that* $(\sigma, t_1, t_2) \in \gamma^n(R)$.

5.5. Common Certainty of Rationality and Iterated Dominance

The set of σ consistent with rationality and common certainty of (the opponent's) rationality given $\mathcal{F}$ can be further characterized for generic games in terms of dominance relations. We say that a game has *no relevant tie* if the following holds: for each player i, all $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ and all terminal histories $z', z'' \in \mathcal{Z}$, if there is a history $h \in \mathcal{H}$ and actions a_1', a_2', a_1'' , a_2'' feasible after h for θ_1 and θ_2 (respectively) such that $a_i' \neq a_i''$, z' follows $(h, (a_1', a_2'))$ and z'' follows $(h, (a_1''$, $a_2''))$, then $u_i(z', \theta_1, \theta_2) \neq u_i(z'', \theta_1, \theta_2)$.

We say that strategy $s_i \in S_i(\theta_i)$ is *weakly dominated* by mixed strategy $m_i \in \Delta(S_i(\theta_i))$ *for type* θ_i *on* $\widehat{\Sigma_j}$ if

$$\forall \sigma_j \in \widehat{\Sigma_j},\ U_i(s_i, \theta_i, \sigma_j) \leq \sum_{s_i'} m_i(s_i') U_i(s_i', \theta_i, \sigma_j)$$

and

$$\exists \sigma_j' \in \widehat{\Sigma_j},\ U_i(s_i, \theta_i, \sigma_j) < \sum_{s_i'} m_i(s_i') U_i(s_i', \theta_i, \sigma_j).$$

The definition of strict dominance is analogous (all weak inequalities are replaced by strict inequalities). For any given rectangular subset $\hat{\Sigma} = \widehat{\Sigma_1} \times \widehat{\Sigma_2} \subset \Sigma$ let $\mathcal{W}(\hat{\Sigma})$ $(\mathcal{S}(\hat{\Sigma}))$ denote the set of $(s_1, \theta_1, s_2, \theta_2)$ such that s_i is not weakly (strictly) dominated for θ_i on $\widehat{\Sigma_j}$ and let $\mathcal{SW}(\hat{\Sigma}) = \mathcal{S}(\hat{\Sigma}) \cap \mathcal{W}(\Sigma)$. The iterated operator $\mathcal{SW}^n$ is defined in the usual way: $\mathcal{SW}^n(\hat{\Sigma}) := \mathcal{SW}\left(\mathcal{SW}^{n-1}(\hat{\Sigma})\right)$, where $\mathcal{SW}^0(\hat{\Sigma}) := \hat{\Sigma}$. A subscript p denotes that we only consider *weak* domination by *pure* strategies. Thus $\mathcal{SW}_p(\hat{\Sigma}) = \mathcal{S}(\hat{\Sigma}) \cap \mathcal{W}_p(\Sigma)$. Note that $\mathcal{S}$ is a monotone operator. Therefore, also $\mathcal{SW}$ and $\mathcal{SW}_p$ are monotone operators.

Proposition 5.13. *(a) In every game with no relevant ties, if* $\phi \in \mathcal{F}$, $\Sigma_{\mathcal{F}}^{\infty} \subset \mathcal{SW}_p^{\infty}(\Sigma)$.
(b) In every game with perfect information, private values, and no relevant ties, if $\phi \in \mathcal{F}$, $\Sigma_{\mathcal{F}}^{\infty} \subset \mathcal{SW}^{\infty}(\Sigma)$ *and* $\Sigma_{\phi}^{\infty} = \mathcal{SW}^{\infty}(\Sigma)$.

Propositions 5.4 and 5.13 imply that, in every generic game with perfect information and private values, common certainty of rationality at the beginning of the game is characterized by the following procedure: first eliminate all (s_i, θ_i) such that s_i is weakly dominated for θ_i, then iteratively eliminate the (s_i, θ_i) such that s_i is strictly dominated for θ_i in the residual strategic form (cf. Ben Porath (1996, Theorem 1) and Dekel and Gul (1996), Proposition 7)). Note that the set of outcomes (terminal histories) induced by strategies in $\mathcal{SW}^\infty(\Sigma)$ is superset of the set of outcomes induced by strategies in $\mathcal{W}^\infty(\Sigma)$ and typically the inclusion is sctrict (take, for example a three-legged version of Rosenthal's (1981) Centipede). Therefore common certainty of rationality at the beginning of the game is not related to *iterated weak* dominance. However, Battigalli (1996b) shows that extensive form rationalizability is outcome-equivalent to iterated weak dominance in all (finite) games of complete and perfect information with no relevant ties. Therefore subsection 5.4 implicitly provides an epistemic characterization of this controversial solution procedure.

Proof of Proposition 5.13. If $(s_i, \theta_i) \in r_i(\mu)$, then s_i is a best reply to the (prior) belief $\mu(\cdot|\Sigma)$ for type θ_i. This implies that s_i cannot be strictly dominated for type θ_i (see Pearce (1984, Lemma 3)). Thus $\Sigma^1_{\mathcal{F}} = r(\Delta^{\mathcal{B}}(\Sigma)) \subset \mathcal{S}(\Sigma)$. If we assume that the game has no relvant tie, then $\Sigma^1_{\mathcal{F}} \subset \mathcal{W}_p(\Sigma)$ (see Battigalli (1996b, Lemma 3). Thus $\Sigma^1_{\mathcal{F}} \subset \mathcal{S}(\Sigma) \cap \mathcal{W}_p(\Sigma) = \mathcal{SW}_p(\Sigma)$. Suppose that

$$\Sigma^n_{\mathcal{F}} \subset \mathcal{SW}^n_p(\Sigma).$$

Since we assume that $\mathcal{F}$ contains the empty history, it follows that

$$\Sigma^{n+1}_{\mathcal{F}} \subset r\left(\left\{\mu \in \Delta^{\mathcal{B}}(\Sigma) : \mu(\mathcal{SW}^n_p(\Sigma)|\Sigma) = 1\right\}\right) \subset$$

$$\mathcal{S}(\mathcal{SW}^n_p(\Sigma)) \cap \mathcal{W}_p(\Sigma) = \mathcal{SW}^{n+1}_p(\Sigma).$$

This proves statement (a).

In every perfect information game with private values, $\mathcal{W}_p(\Sigma) = \mathcal{W}(\Sigma)$ (Battigalli (1996, Lemma 4) shows this result for games with perfect and complete information, the proof can be easily adapted to cover the present more general case). Thus, if the game has no relevant tie, $\Sigma^1_{1,\mathcal{F}} \times \Sigma^1_{2,\mathcal{F}} \subset \mathcal{W}(\Sigma)$.[12] The same argument as above then proves the first part of statement (b). Now let $\mathcal{F}$ contain only the empty history ϕ. For all k,

$$\Sigma^k_\phi = r\left(\left\{\mu \in \Delta^{\mathcal{B}}(\Sigma) : \mu(\Sigma^{k-1}_\phi|\Sigma) = 1\right\}\right).$$

[12] This is also proved by Ben Porath (1996, Lemma 2.1).

Suppose that

$$\Sigma_{\phi}^{n} = \mathcal{SW}^{n}(\Sigma)$$

and let $(s_1, \theta_1, s_2, \theta_2) \in \mathcal{SW}^{n+1}(\Sigma)$. By the induction hypothesis and the definition of operator $\mathcal{SW}$, $(s_1, \theta_1, s_2, \theta_2) \in \mathcal{S}(\Sigma_{\phi}^{n}) \cap \mathcal{W}(\Sigma) \subset \Sigma_{\phi}^{n}$. Thus for each i, there are $\nu', \nu'' \in \Delta(\Sigma_j)$ such that $\nu'(\Sigma_{j,\phi}^{n}) = 1$, ν'' is strictly positive and s_i is a best response to ν' and ν'' for type θ_i (Pearce (1984, Lemmata 3 and 4)). Costruct $\mu \in [\Delta(\Sigma)]^{\mathcal{B}}$ as follows: for all $h \in \mathcal{H}(s_i, \theta_i)$, $A_j \subset \Sigma_j(h)$,

$$\mu(\{(s_i, \theta_i)\} \times A_j | \Sigma(h)) = \frac{\nu(A_j)}{\nu(\Sigma_j(h))},$$

where $\nu = \nu'$, if $\nu'(\Sigma_j(h)) > 0$, and $\nu = \nu''$ otherwise; for all $h \notin \mathcal{H}(s_i, \theta_i)$, $\sigma_i \in \Sigma_i(h)$, $A_j \subset \Sigma_j(h)$,

$$\mu(\{\sigma_i\} \times A_j | \Sigma(h)) = \frac{\nu(A_j)}{|\Sigma_i(h)| \cdot \nu(\Sigma_j(h))},$$

where $|A|$ denotes the cardinality of A and ν is chosen as before. It can be checked that $\mu \in \Delta^{\mathcal{B}}(\Sigma)$, $\mu(\Sigma_{\phi}^{n} | \Sigma) = 1$ and $(s_i, \theta_i) \in r_i(\mu)$. Thus $(s_i, \theta_i) \in \Sigma_{i,\phi}^{n+1}$. ■

6. Related Literature

In this section we offer some comments relating the present work to a few papers about interactive epistemology in dynamic games.[13]

Ben Porath (1996) considers finite games of perfect and complete information. His notion of type space ("world" in his terminology) is essentially the same as in this paper, but his analysis is limited to an *implicit* representation of interactive conditional beliefs.[14] Lemma 4.1 shows that the epistemic models considered by

[13] For more on hierarchies of beliefs and type spaces see Heifetz and Samet (1996a). For more on interactive epistemology in games see Dekel and Gul (1996).

[14] There are two related (minor) differences. First, Ben Porath implicitly requires that a stochastic independence condition is satisfied. In our notation, for every τ_i, $g_i(\tau_i) \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$ is such that a "marginal" CPS μ_{ij} about j can be derived from $g_i(\tau_i)$. Formally, there is some $\mu_{ij} \in \Delta^{\mathcal{B}_j}(\Sigma_j \times T_j)$ satisfying: $mrg_{\Sigma_j \times T_j} g_{i,h}(\tau_i) = \mu_{ij}(\cdot | \Sigma_j(h))$ for all $h \in \mathcal{H}$. This is a plausible restriction, but given our definition of sequential rationality for plans of action, it has no consequence at all. However – and this is the second difference – Ben Porath's notion of sequential rationality requires expected utility maximization at *every* history (not only those consistent with the given strategy). The foregoing assumption is relevant if we consider the strategies consistent with this notion of rationality. But the set of plans of action (equivalence classes of strategies) and hence the set of histories consistent with (common certainty of) rationality in his sense and ours are the same.

Ben Porath are isomorphic to belief-closed subsets of the universal type space containing all the hierarchies of CPSs satisfying common certainty of coherency. Ben Porath (1996, Theorem 1) characterizes the set of outcomes consistent common certainty of rationality at the beginning of the game in *finite* type spaces. Propositions 5.4 (c) and 5.13 generalize and extend Ben Porath's result also proving his claim that considering only finite type spaces is without loss of generality.

Reny (1993, 1995) also considers finite games of perfect and complete information, but without a formal reference to type spaces. He analyzes the possibility that there is common certainty of rationality at particular decision nodes (non terminal histories) as well as conditional on a given set of nodes. His main result characterizes the set of (generic) "belief consistent" games. His analysis can be reformulated in our framework.[15] A *jointly rational belief system (JRBS)* for the set of (non terminal) histories $\mathcal{F}$ is a pair of subsets (Σ_1^*, Σ_2^*) such that, for $i = 1, 2$,

$$\emptyset \neq \Sigma_i^* = r_i \left[\Lambda_{\mathcal{F}}(\Sigma_1^* \times \Sigma_2^*)\right] \cap \left[\bigcup_{h \in \mathcal{F}} \Sigma_i(h)\right].$$

A history $h \in \mathcal{H}$ is *relevant* if it is reacheable by a pair of rational strategies ($\Sigma_\phi^1 \cap \Sigma(h) \neq \emptyset$) and no player has a strictly dominant choice given h. A game is *belief consistent* if there is a JRBS for the set of relevant histories. Reny (1993) shows that a generic game of perfect and complete information is belief consistent if and only if every history off the backward induction path is irrelevant. Reny (1995) shows that in games like the Centipede there can be common certainty of rationality only at the outset.[16] Proposition 5.4 implies that there is a JRBS for a single history h if and only if there is a σ consistent with rationality and common certainty of rationality conditional on h. Proposition 5.6 implies that there is a JRBS for a set of histories $\mathcal{F}$ if and only if there is a σ consistent with rationality and common certainty of the *opponent*'s rationality given $\mathcal{F}$.

Stalnaker (1996) analyzes counterfactual reasoning in games of complete information. In his epistemic models each type corresponds to a *complete* conditional probability system on the set of strategy pairs and types of the opponent. To use

[15] More details on this reformulation are available by request.

[16] This formalization is faithful to the spirit of Reny's (1993) analysis, but there are some differences. First, Reny allows for the choice of mixed strategies. Second, players do not have beliefs about themselves. Third, beliefs satisfy only a very weak form of bayesian updating. The first two points are inessential, while weakening bayesian updating expands the collection of $\mathcal{F}$ for which there exists a jointly rational belief system. But this has the only effect to make Reny's "impossibility result" stronger. The analysis of Reny (1995) is fully consistent with ours.

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 emtpy subsets of Σ, in Stalnaker (1996) $\mathcal{B}$ is the collection of all non empty subsets of $\Sigma \times T_j$.[17] Thus there is an obvious (projection) mapping between Stalnaker's epistemic spaces and our type spaces, but this mapping is not invertible.[18] To be more explicit, let $\mathcal{P}(X)$ denote the collection of all non empty subsets of a given set X. Then every complete CPS $\mu \in \Delta^{\mathcal{P}(\Sigma \times T_j)}(\Sigma \times T_j)$ corresponds to a CPS $\pi_{\mathcal{B}}(\mu) \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$, where $\mathcal{B} \subset \mathcal{P}(\Sigma)$ and the (projection) mapping $\pi_{\mathcal{B}}$ satifies the following obvious condition:

$$\forall B \in \mathcal{B},\ \pi_{\mathcal{B}}(\mu)(\cdot|B \times T_j) = \mu(\cdot|B \times T_j).$$

The projection mapping $\pi_{\mathcal{B}}$ is not one to one (unless $\mathcal{B} = \mathcal{P}(\Sigma)$ and T_j is a singleton). But from the point of view of game theoretic analysis it is interesting to know whether for every CPS $\nu \in \Delta^{\mathcal{B}}(\Sigma \times T_j)$ one can find a corresponding complete CPS $\mu \in \Delta^{\mathcal{P}(\Sigma \times T_j)}(\Sigma \times T_j)$ such that $\nu = \pi_{\mathcal{B}}(\mu)$, i.e. whether $\pi_{\mathcal{B}}$ is onto. One can show that if the collection of relevant hypotheses $\mathcal{B}$ corresponds to the set $\mathcal{H}$ of histories of the game, as assumed in Section 5, then $\pi_{\mathcal{B}}$ is onto (cf. Battigalli (1994, Theorem 1)).[19] Given our notion of rationality in terms of conditional expected utility maximization, this implies that results about consistency of outcomes with rationality and conditional mutual certainty of rationality do not change when we use epistemic models *à la* Stalnaker instead of type spaces as defined in this paper. Actually, Stalnaker uses a notion "perfect rationality" (relying on lexicographic utility maximization) which cannot be defined in our framework. But for generic payoffs over terminal histories the two notions of rationality are equivalent.

References

[1] BATTIGALLI, P. (1994): "Structural Consistency and Strategic Independence in Extensive Games," *Ricerche Economiche,* **48**, 357-376.

[17] The mapping relating each type τ_i to a complete CPS on $\Sigma \times T_j$ is derived from (a strictly positive prior on the set of states of the world and) a belief revision function. Also Samet (1993) models counterfactual reasoning in dynamic games using belief revision functions. We plan to relate Samet's analysis to ours in the next version of this paper.

[18] This is not a trivial difference. While Lemma 4.1 "justifies" using type spaces in our sense, we are not aware of analogous results for Stalnaker's epistemic spaces.

[19] As a consequence of Theorem 1 in Rényi (1956)), the same is true if $\mathcal{B}$ is closed under (finite) union.

[2] BATTIGALLI, P. (1996a): "Strategic Rationality Orderings and the Best Rationalization Principle," *Games and Economic Behavior,* **13**, 178-200.

[3] BATTIGALLI, P. (1996b): "On Rationalizability in Extensive Games," *Journal of Economic Theory*, forthcoming.

[4] BATTIGALLI, P. and G. BONANNO (1995): "Synchronic Information and Common Knowledge in Extensive Games," forthcoming in (M. Bacharach, L.A. Gerard-Varet, P. Mongin, and H. Shin editors) *Epistemic Logic and the Theory of Games and Decisions*. Dordrecht: Kluwer.

[5] BEN PORATH, E. (1996): "Rationality, Nash Equilibrium and Backwards Induction in Perfect Information Games," *Review of Economic Studies,* forthcoming.

[6] BRANDENBURGER, A. and E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, **59**, 189-198.

[7] DEKEL, E. and F. GUL (1996): "Rationality and Knowledge in Game Theory," forthcoming in *Advances in Economics and Econometrics* (D. Kreps and K. Wallis, Eds.). Cambridge UK: Cambridge University Press.

[8] FUDENBERG D. and J. TIROLE (1991): *Game Theory.* Cambridge MA: MIT Press.

[9] HARSANYI, J. (1967-68): "Games of Incomplete Information Played by Bayesian Players. Parts I, II, III," *Management Science,* **14**, 159-182, 320-334, 486-502.

[10] HEIFETZ, A. and D. SAMET (1996a): "Topology-Free Typology of Beliefs," mimeo, Tel Aviv University.

[11] HEIFETZ, A. and D. SAMET (1996b): "Mutual Beliefs Are Not AlwaysTypes," mimeo, Tel Aviv University.

[12] MERTENS J.F. and S. ZAMIR (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, **14**, 1-29.

[13] MYERSON, R. (1986): "Multistage Games with Communication," *Econometrica,* **54**, 323-358.

[14] PEARCE, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica,* **52**, 1029-1050.

[15] RENY, P. (1992): "Backward Induction, Normal Form Perfection and Explicable Equilibria," *Econometrica,* **60**, 626-649.

[16] RENY, P. (1993): "Common Belief and the Theory of Games with Perfect Information," *Journal of Economic Theory*, **59**, 257-274.

[17] RENY, P. (1995): "Rational Behaviour in Extensive Form Games," *Canadian Journal of Economics*, **28**, 1-16.

[18] RÊNYI, A. (1955): "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungaricae,* **6**, 285-335.

[19] RÊNYI, A. (1956): "On Conditional Probability Spaces Generated by a Conditionally Ordered Set of Measures," *Theory of Probability and Its Applications,* **1**, 61-71.

[20] ROSENTHAL, R. (1981): "Games of Perfect Information, Predatory Pricing and the Chain-Store paradox," *Journal of Economic Theory*, **25**, 92-100.

[21] SAMET, D. (1993): "Hypothetical Knowledge and Games with Perfect Information," *Games and Economic Behavior*, forthcoming.

[22] STALNAKER, R. (1996): "Knowledge, Belief and Counterfactual Reasoning in Games," *Economics and Philosophy*, **12**, ..-...

[23] TAN, T. and S. WERLANG (1988): "The Bayesian Foundation of Solution Concepts of Games," *Journal of Economic Theory,* **45**, 370-391.