EPICENTER Working Paper 2014–1

When Do Types Induce the Same Belief Hierarchy?

Andrés Perea*
Maastricht University

January 27, 2014

Abstract

Harsanyi (1967–1968) showed that belief hierarchies can be encoded by means of epistemic models with types. Indeed, for every type within an epistemic model we can derive the full belief hierarchy it induces. But for one particular belief hierarchy, there are in general many different ways of encoding it within an epistemic model. In this paper we give necessary and sufficient conditions such that two types, from two possibly different epistemic models, induce exactly the same belief hierarchy. The conditions are relatively easy to check, and seem relevant both for practical and theoretical purposes.

JEL Classification: C72

Keywords: Epistemic game theory, types, belief hierarchies.

1 Introduction

Belief hierarchies play a fundamental role in the modern analysis of games. This seems natural, as one cannot make a good decision in a game without first forming a belief about the opponents' choices, and in order to form a *reasonable* belief about the opponents' choices it seems necessary

* *Address*: EpiCenter and Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.
E-mail: a.perea@maastrichtuniversity.nl
Web: http://www.personeel.unimaas.nl/a.perea/

to first form some belief about the opponents' beliefs about their opponents' choices, and so on. This naturally leads to the concept of a *belief hierarchy,* which specifies what a player believes about the opponents' choices – his *first-order* belief – what the player believes about the opponents' first-order beliefs – his *second-order* belief – and so on.

To the best of my knowledge, Harsanyi (1962) was the first to formally define a belief hierarchy within the context of a game, although he did so for a very special setting. One important practical problem with belief hierarchies – and that may also have been a reason why belief hierarchies entered the game theory stage relatively late – is that these are *infinite* objects, with *infinitely* many layers of beliefs. It is thus impossible to explicitly write down a belief hierarchy, layer by layer, as there are infinitely many of these. But then, the question naturally arises: Is there a way to represent belief hierarchies in a compact and convenient way?

Harsanyi (1967–1968), some years after he introduced the notion of a belief hierarchy, gave a positive and elegant answer to this question. He focused on a setting in which the belief hierarchies concern the players' beliefs about the opponents' *utilities,* but his construction has later been extended to situations where players also hold beliefs about the opponents' *choices* – which is the relevant setting for our paper. The construction that Harsanyi proposed was very simple: For every player we define a set of types, and for every type we define a utility function, together with a probabilistic belief about the *opponents' types.* From this very simple construction we can then derive, for every type, a first-order belief about the opponents' utility functions, a second-order belief about the opponents' first-order beliefs, and so on. That is, for every type we can derive a *full belief hierarchy* on the players' possible utilities in the game. This construction by Harsanyi was a major step forward, as it allowed us to encode infinite belief hierarchies in a very short and convenient fashion.

Harsanyi's original idea can easily be adapted to a framework where players hold beliefs about the opponents' *choices* rather than the opponents' utilities. Indeed, consider for every player a set of types, and associate to every type a probabilistic belief about the opponents' *choices and types.* Then, similarly to Harsanyi's construction, we can then derive for every type a full belief hierarchy on the players' *choices* in the game. This construction, which we call an *epistemic model with types*, can thus be viewed as a possible way to encode belief hierarchies. During the last few decades, it has played a key role in the epistemic analysis of games.

The question we ask in this paper is actually very simple. Consider two epistemic models with finitely many types, and for a given player choose one type from each of the two models. When do these two types induce the same belief hierarchy?

Checking this directly, by explicitly comparing their induced first-order beliefs, second-order beliefs, and so on, may be quite cumbersome as one needs to check for infinitely many levels of belief. Instead, we present in Theorem 2 a (finite) set of necessary and sufficient conditions which are relatively easy to check. We think this theorem has important practical and theoretical implications. On a practical level, it may help to check whether for a given game, and for two given epistemic models, two types induce the same belief hierarchy or not. On a more theoretical level, it may be helpful for investigating the structure of type spaces or for designing proofs.

For instance, Theorem 2 can be used to characterize *non-redundant* epistemic models – which are models where no two different types induce the same belief hierarchy. Or, suppose one attempts to prove that two different game-theoretic concepts – which work on two different epistemic models – eventually generate the same set of belief hierarchies. Then, the necessary and sufficient conditions in Theorem 2 may be very helpful for designing such a proof, as they provide a way to "travel" from one epistemic model to the other without changing the induced belief hierarchy.

The outline of this paper is as follows. In Section 2 we formally introduce epistemic models, and show how to derive a belief hierarchy from a type within a given epistemic model. In Section 3 we state our main result – Theorem 2 – and illustrate it by means of an example. In Section 4 we state some preparatory results which are needed to prove the main result. In Section 5 we give a formal proof of the main result. In Section 6 we investigate what the main theorem implies for some interesting special cases. In Section 7 we discuss how the main theorem could be extended to other settings. In Section 8, finally, we give the proofs for the preparatory results in Section 4.

2 Belief Hierarchies and Types

In this section we show how belief hierarchies can be encoded by means of an epistemic model with types, and how every type can be "decoded" by deriving a full belief hierarchy from it.

2.1 From Belief Hierarchies to Types

Consider a finite static game $\Gamma = (C_i, u_i)_{i \in I}$ where I is the finite set of players, C_i is the finite set of choices for player i, and $u_i : \times_{j \in I} C_j \to \mathbb{R}$ is player i's utility function. A *belief hierarchy* for player i specifies a probabilistic belief about the opponents' choices – the *first-order* belief – a probabilistic belief about the opponents' possible first-order beliefs – the *second-order* belief – and so on. Following Harsanyi's (1967–1968) approach, we will encode such infinite belief hierarchies by means of an epistemic model with *types.* In this paper we focus on epistemic models with *finitely* many types, which of course imposes restrictions on the possible belief hierarchies we can encode. Indeed, there are belief hierarchies which can simply not be encoded by an epistemic model with finitely many types. See Section 7 for a discussion of this restriction to finite type spaces, and how the main result could possibly be extended to a setting with infinitely many types.

Definition 1 (Epistemic Model) *A finite epistemic model for* Γ *is a tuple* $M = (T_i, b_i)_{i \in I}$ *where, for every player* i,

(a) T_i *is the finite set of types for player* i, *and*

(b) $b_i : T_i \to \Delta(C_{-i} \times T_{-i})$ *is a mapping that assigns to every type* t_i *a probabilistic belief* $b_i(t_i) \in \Delta(C_{-i} \times T_{-i})$ *on the opponents' choice-type combinations.*

Here, $\Delta(X)$ denotes the set of probability distributions on X, for every finite set X. In our main result – Theorem 2 – an important role is played by the sets $T^*(t_i)$, representing the set of types that "enter type t_i's belief hierarchy". We will now define these sets $T^*(t_i)$ formally. To do so, we first recursively define sets of types $T^1(t_i), T^2(t_i), ...$ for all types t_i in M. To start, let

$$T^1(t_i) := \{t_i\} \cup \left[\bigcup_{j \neq i} \{t_j \in T_j \mid b_i(t_i)(C_{-i} \times \{t_j\} \times T_{-ij}) > 0\} \right],$$

where $T_{-ij} = \times_{k \in I \backslash \{i,j\}} T_k$. Hence, $T^1(t_i)$ contains t_i itself and all opponents' types to which t_i assigns positive probability.

Now, suppose that $n \geq 2$, and that $T^{n-1}(t_i)$ has been defined for all types t_i in M. Then, for every player i, and every type $t_i \in T_i$, we define

$$T^n(t_i) := \bigcup_{t \in T^{n-1}(t_i)} T^1(t).$$

Finally, we define the set

$$T^*(t_i) := \bigcup_{n \in \mathbb{N}} T^n(t_i),$$

representing the set of types that "enter t_i's belief hierarchy".

One can easily visualize the definition of $T^*(t_i)$ by a directed graph, where the nodes are the types in M, and where there is an edge from a type t to another type t' whenever t assigns positive probability to t'. Then, $T^n(t_i)$ contains precisely those types that can be reached within at most n steps from t_i, and $T^*(t_i)$ contains those types that can be reached within finitely many steps when starting at t_i.

2.2 From Types to Belief Hierarchies

In the previous subsection we have introduced an epistemic model as a way to encode belief hierarchies. We will now show how to "decode" a type within an epistemic model, by deriving the full belief hierarchy it induces.

Consider a finite epistemic model $M = (T_i, b_i)_{i \in I}$ for Γ. Then, every type t_i within M induces an infinite belief hierarchy

$$h_i(t_i) = (h_i^1(t_i), h_i^2(t_i), ...),$$

where $h_i^1(t_i)$ is the induced first-order belief, $h_i^2(t_i)$ is the induced second-order belief, and so on. We will inductively define, for every n, the n-th order beliefs induced by types t_i in M, building upon the $(n-1)$-th order beliefs that have been defined in the preceding step. We start by defining the first-order beliefs.

For every player i, and every type $t_i \in T_i$, define the first-order belief $h_i^1(t_i) \in \Delta(C_{-i})$ by

$$h_i^1(t_i)(c_{-i}) := b_i(t_i)(\{c_{-i}\} \times T_{-i}).$$

Let

$$h_i^1(T_i) := \{h_i^1(t_i) \mid t_i \in T_i\} \subseteq \Delta(C_{-i})$$

be the finite set of first-order beliefs for player i induced by types in T_i.

Now, suppose that $n \geq 2$, and assume that the beliefs $h_i^{n-1}(t_i)$ and the sets $h_i^{n-1}(T_i)$ have been defined for all players i, and every type $t_i \in T_i$. For every $h_i^{n-1} \in h_i^{n-1}(T_i)$, let

$$T_i[h_i^{n-1}] := \{t_i \in T_i \mid h_i^{n-1}(t_i) = h_i^{n-1}\}.$$

We define the n-th order beliefs $h_i^n(t_i)$ and the sets $h_i^n(T_i)$ as follows. For every type $t_i \in T_i$, let $h_i^n(t_i) \in \Delta(C_{-i} \times h_{-i}^{n-1}(T_{-i}))$ be given by

$$h_i^n(t_i)(c_{-i}, h_{-i}^{n-1}) := b_i(t_i)(\{c_{-i}\} \times T_{-i}[h_{-i}^{n-1}])$$

for every $c_{-i} \in C_{-i}$ and every $h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i})$. Here, $h_{-i}^{n-1}(T_{-i}) := \times_{j \neq i} h_j^{n-1}(T_j)$, and for a given $h_{-i}^{n-1} = (h_j^{n-1})_{j \neq i}$ in $h_{-i}^{n-1}(T_{-i})$ we define $T_{-i}[h_{-i}^{n-1}] := \times_{j \neq i} T_j[h_j^{n-1}]$.

By

$$h_i^n(T_i) := \{h_i^n(t_i) \mid t_i \in T_i\} \subseteq \Delta(C_{-i} \times h_{-i}^{n-1}(T_{-i}))$$

we denote the finite set of n-th order beliefs for player i induced by types in T_i.

Finally, for every type $t_i \in T_i$, we denote by

$$h_i(t_i) := (h_i^n(t_i))_{n \in \mathbb{N}}$$

the *belief hierarchy* on the players' choices induced by t_i.

3 Main Result

Our main result provides necessary and sufficient conditions such that two types, from possibly different epistemic models, induce exactly the same belief hierarchy. To state the result formally, we need the following definitions. For two sets A and B, a *correspondence* $f : A \twoheadrightarrow B$ is an object that assigns to every $a \in A$ a nonempty subset $f(a) \subseteq B$. The correspondence $f : A \twoheadrightarrow B$ is called *disjoint* when for all $a, a' \in A$, either $f(a) = f(a')$, or $f(a) \cap f(a') = \emptyset$. For every $a \in A$, we denote by $[a]_f := \{a' \in A \mid f(a') = f(a)\}$ the set of elements that induce the same image under f as a. Moreover, for a given type $t_j^* \in T_j$ and a given player i, we define $T_i^*(t_j^*) := T^*(t_j^*) \cap T_i$, and $T_{-i}^*(t_j^*) := \times_{k \neq i} T_k^*(t_j^*)$. In a similar fashion, we define $R_i^*(r_j^*)$ and $R_{-i}^*(r_j^*)$ for a type $r_j^* \in R_j$.

Theorem 2 (Main Theorem) *Consider a finite static game* $\Gamma = (C_i, u_i)_{i \in I}$, *and two finite epistemic models* $M = (T_i, b_i)_{i \in I}$ *and* $M' = (R_i, \beta_i)$ *for* Γ. *For a given player* j, *consider a type* $t_j^* \in T_j$ *and a type* $r_j^* \in R_j$.

Then, t_j^* *and* r_j^* *induce the same belief hierarchy, if and only if, for every player* i *there is a disjoint correspondence*

$$f_i : T_i^*(t_j^*) \twoheadrightarrow R_i^*(r_j^*)$$

such that $r_j^* \in f_j(t_j^*)$ *and*

$$b_i(t_i)(\{c_{-i}\} \times [t_{-i}]_{f_{-i}}) = \beta_i(r_i)(\{c_{-i}\} \times f_{-i}(t_{-i})) \text{ for all } r_i \in f_i(t_i), \tag{1}$$

for all players i, *all types* $t_i \in T_i^*(t_j^*)$, *all* $c_{-i} \in C_{-i}$, *and all* $t_{-i} \in T_{-i}^*(t_j^*)$.

Here, $f_{-i} : T_{-i}^*(t_j^*) \twoheadrightarrow R_{-i}^*(r_j^*)$ is the induced disjoint correspondence that assigns to every $t_{-i} = (t_k)_{k \neq i}$ in $T_{-i}^*(t_j^*)$ the set $\times_{k \neq i} f_k(t_k)$, which is a subset of $R_{-i}^*(r_j^*)$.

Example. We will now illustrate our main theorem by means of an example. Consider a two-player game $\Gamma = (C_i, u_i)_{i \in I}$ with choice set $C_1 = \{a, b\}$ for player 1, and choice set $C_2 = \{c, d\}$ for player 2. Consider the two epistemic models $M = (T_1, T_2, b_1, b_2)$ and $M' = (R_1, R_2, \beta_1, \beta_2)$ in Table 1. Here, $b_1(t_1) = \frac{1}{2}(c, t_2) + \frac{1}{2}(d, t_2')$ means that type t_1 assigns probability $\frac{1}{2}$ to the opponent's choice-type pair (c, t_2), and probability $\frac{1}{2}$ to the opponent's choice-type pair (d, t_2'). Similarly for the other types in the table.

We will use our main theorem to show that the types t_1 and r_1' induce the same belief hierarchy on choices. Note first that $T_1^*(t_1) = T_1$, $T_2^*(t_1) = T_2$, $R_1^*(r_1') = R_1$ and $R_2^*(r_1') = R_2$. Define the disjoint correspondences $f_1 : T_1^*(t_1) \twoheadrightarrow R_1^*(r_1')$ and $f_2 : T_2^*(t_1) \twoheadrightarrow R_2^*(r_1')$ by

$$\begin{aligned} f_1(t_1) &= f_1(t_1') = \{r_1', r_1''\}, \\ f_1(t_1'') &= \{r_1\}, \\ f_2(t_2) &= f_2(t_2'') = \{r_2'\} \text{ and} \\ f_2(t_2') &= \{r_2, r_2''\}. \end{aligned}$$

Then, $r_1' \in f_1(t_1)$, and it may be verified that these correspondences f_1 and f_2 satisfy condition (1) above. Hence, by Theorem 2, types t_1 and r_1' induce the same belief hierarchy. By the same argument, we can actually conclude that the types t_1, t_1', r_1' and r_1'' all induce the same belief hierarchy, and that the types t_1'' and r_1 induce the same belief hierarchy.

To see why f_1 and f_2 satisfy condition (1), note first that

$$\begin{aligned} [t_1]_{f_1} &= \{t_1, t_1'\}, \ [t_1'']_{f_1} = \{t_1''\}, \\ [t_2]_{f_2} &= \{t_2, t_2''\}, \text{ and } [t_2']_{f_2} = \{t_2'\}. \end{aligned}$$

Epistemic model $M = (T_1, T_2, b_1, b_2)$
$T_1 = \{t_1, t_1', t_1''\}, \qquad T_2 = \{t_2, t_2', t_2''\}$
$b_1(t_1) = \frac{1}{2}(c, t_2) + \frac{1}{2}(d, t_2')$ $b_1(t_1') = \frac{1}{6}(c, t_2) + \frac{1}{3}(c, t_2'') + \frac{1}{2}(d, t_2')$ $b_1(t_1'') = \frac{1}{3}(c, t_2') + \frac{2}{3}(d, t_2'')$
$b_2(t_2) = \frac{1}{4}(a, t_1) + \frac{1}{2}(a, t_1') + \frac{1}{4}(b, t_1'')$ $b_2(t_2') = \frac{1}{8}(a, t_1) + \frac{1}{8}(a, t_1') + \frac{3}{4}(b, t_1'')$ $b_2(t_2'') = \frac{1}{2}(a, t_1) + \frac{1}{4}(a, t_1') + \frac{1}{4}(b, t_1'')$

Epistemic model $M' = (R_1, R_2, \beta_1, \beta_2)$
$R_1 = \{r_1, r_1', r_1''\}, \qquad R_2 = \{r_2, r_2', r_2''\}$
$\beta_1(r_1) = \frac{1}{6}(c, r_2) + \frac{1}{6}(c, r_2'') + \frac{2}{3}(d, r_2')$ $\beta_1(r_1') = \frac{1}{2}(c, r_2') + \frac{1}{8}(d, r_2) + \frac{3}{8}(d, r_2'')$ $\beta_1(r_1'') = \frac{1}{2}(c, r_2') + \frac{3}{8}(d, r_2) + \frac{1}{8}(d, r_2'')$
$\beta_2(r_2) = \frac{1}{4}(a, r_1') + \frac{3}{4}(b, r_1)$ $\beta_2(r_2') = \frac{3}{4}(a, r_1') + \frac{1}{4}(b, r_1)$ $\beta_2(r_2'') = \frac{1}{8}(a, r_1') + \frac{1}{8}(a, r_1'') + \frac{3}{4}(b, r_1)$

Table 1: Illustration of the main theorem

Now, take the type $t_1 \in T_1$, the types $r_1', r_1'' \in f_1(t_1)$, the opponent's type $t_2' \in T_2$, and the opponent's choice d. Then,

$$\begin{aligned} b_1(t_1)(\{d\} \times [t_2']_{f_2}) &= b_1(t_1)(\{d\} \times \{t_2'\}) = \tfrac{1}{2}, \\ \beta_1(r_1')(\{d\} \times f_2(t_2')) &= \beta_1(r_1')(\{d\} \times \{r_2, r_2''\}) = \tfrac{1}{8} + \tfrac{3}{8} = \tfrac{1}{2}, \\ \beta_1(r_1'')(\{d\} \times f_2(t_2')) &= \beta_1(r_1'')(\{d\} \times \{r_2, r_2''\}) = \tfrac{3}{8} + \tfrac{1}{8} = \tfrac{1}{2}. \end{aligned}$$

Hence, we may conclude that

$$b_1(t_1)(\{d\} \times [t_2']_{f_2}) = \beta_1(\hat{r}_1)(\{d\} \times f_2(t_2')) \text{ for all } \hat{r}_1 \in f_1(t_1),$$

which establishes condition (1) for $t_1 \in T_1$, $t_2' \in T_2$ and $d \in C_2$. In the same way, it may be verified that (1) also holds for every other combination of types and choices. Hence, in this way it can be checked that the correspondences f_1 and f_2 indeed satisfy condition (1). So, by Theorem 2 we may indeed conclude that t_1 and r_1' induce the same belief hierarchy.

4 Some Preparatory Results

Before we prove Theorem 2 we first present some preparatory results that will be useful for the proof. For the preparatory results we present in this section, assume that $\Gamma = (C_i, u_i)_{i \in I}$ is a finite static game, and assume that $M = (T_i, b_i)_{i \in I}$ and $M' = (R_i, \beta_i)_{i \in I}$ are two finite epistemic models for Γ. The proofs of the three lemmas can be found in the proofs section at the end of this paper. The proof of Corollary 4 is given in the present section, as it is very short.

Our first result states that two types which induce the same n-th order belief, also induce the same $(n-1)$-th order belief. As a consequence, two types that share the same n-th order belief, also share the same $(n-1)$-th order belief, the same $(n-2)$-th order belief, ..., the same first-order belief. This results seems quite natural as – on an intuitive level – an n-th order belief is a "more detailed" belief than an $(n-1)$-th order belief, and hence we should be able to derive the $(n-1)$-th order belief if the n-th order belief is given.

Lemma 3 (Identical n-th order beliefs imply identical $(n-1)$-th order beliefs) *For some $n \geq 2$, consider a type $t_i \in T_i$ and a type $r_i \in R_i$ with $h_i^n(t_i) = h_i^n(r_i)$. Then, $h_i^{n-1}(t_i) = h_i^{n-1}(r_i)$.*

Despite the fact that this result is quite intuitive, proving this result still requires a non-trivial piece of work. The reason is that in our setup, the $(n-1)$-th order belief cannot directly be obtained from the n-th order belief by taking a marginal, as is the case in Mertens and Zamir (1985) and Brandenburger and Dekel (1993). To see this, note that in the Mertens-Zamir and the Brandenburger-Dekel framework, the space of uncertainty for the n-th order beliefs can directly be written as the Cartesian product of the space of uncertainty for the $(n-1)$-th order beliefs and some other set. Therefore, we can immediately derive an $(n-1)$-th order belief from an n-th order belief by simply taking the marginal on the space of uncertainty for $(n-1)$-th order beliefs. Such a construction is not possible within our setup, and this makes the proof of Lemma 3 more difficult than one might maybe expect.

Our next result follows rather easily from the lemma above. It states that for every two finite epistemic models, we can always find a number n such that two types induce the same belief hierarchy precisely when they induce the same n-th order belief. Hence, checking the n-th order belief is sufficient for testing whether two types induce the same belief hierarchy or not. This result plays an important role in proving the "only if" direction of Theorem 2.

Corollary 4 (Identical n-th order beliefs imply identical belief hierarchies) *There is some $n \geq 1$ such that for every player i, every $t_i \in T_i$ and $r_i \in R_i$,*

$$h_i(t_i) = h_i(r_i) \text{ if and only if } h_i^n(t_i) = h_i^n(r_i).$$

Proof. Fix a player i. For every $n \geq 1$ define the set

$$A_i^n := \{(t_i, r_i) \in T_i \times R_i \mid h_i^n(t_i) = h_i^n(r_i)\},$$

and let

$$A_i^\infty := \cap_{n \in \mathbb{N}} A_i^n = \{(t_i, r_i) \in T_i \times R_i \mid h_i(t_i) = h_i(r_i)\}.$$

By Lemma 3 it follows that $A_i^n \subseteq A_i^{n-1}$ for all $n \geq 2$. Hence, as both T_i and R_i are finite, there must be some number n_i such that $A_i^{n_i} = A_i^\infty$. This implies that, for every $t_i \in T_i$ and $r_i \in R_i$, $h_i(t_i) = h_i(r_i)$ if and only if $h_i^{n_i}(t_i) = h_i^{n_i}(r_i)$. If we set $n := \max\{n_i \mid i \in I\}$, the proof is complete. ■

Our following result states that, whenever two types t_i and r_i are equivalent – in the sense of inducing the same belief hierarchy – then every type that enters t_i's belief hierarchy must be equivalent to a type that enters r_i's belief hierarchy. Also this result is rather intuitive: Consider two types t_i and r_i that are equivalent, and some type t' that enters t_i's belief hierarchy. Then, t_i must deem possible some type that deems possible some type ... that deems possible some type that deems possible the type t'. In terms of belief hierarchies, this means that t_i deems possible the event that some other player deems possible the event that ... that some other player deems possible the event that his opponent has the belief hierarchy induced by t'. As r_i holds the same belief hierarchy as t_i, also r_i must deem possible the event that some other player deems possible the event that ... that some other player deems possible the event that his opponent has the belief hierarchy induced by t'. In particular, there must be a type entering r_i's belief hierarchy that induces the same belief hierarchy as t' does – exactly what we have to show. Proving this intuitive result still requires some hard work, as the reader will see. But basically the proof provides a formalization of the intuitive argument above. For the purposes of this paper, this result will be crucial for proving the "only if" direction in Theorem 2.

Lemma 5 (Equivalent types deem possible equivalent opponents' types) *Consider a type $t_j^* \in T_i$ and a type $r_j^* \in R_j$ with $h_j(t_j^*) = h_j(r_j^*)$. Then, for every player i, and every $t_i \in T_i^*(t_j^*)$, there is some $r_i \in R_i^*(r_j^*)$ with $h_i(t_i) = h_i(r_i)$.*

Our last preparatory result is a technical property that follows from condition (1) in Theorem 2.

Lemma 6 (Consequence of condition (1)) *Consider a fixed type $t_j^* \in T_j$. For every player i, let $f_i : T_i^*(t_j^*) \twoheadrightarrow R_i^*(r_j^*)$ be a disjoint correspondence such that condition (1) in Theorem 2 is satisfied. Then, for every player i, every $t_i \in T_i^*(t_j^*)$ and every $r_i \in f_i(t_i)$, the belief $\beta_i(r_i)$ only assigns positive probability to opponents' type combinations that are in $f_{-i}(t_{-i})$ for some $t_{-i} \in T_{-i}^*(t_j^*)$.*

With these preparatory results at hand, we are now fully equipped to prove our main theorem.

5 Proof of the Main Result

In this section we give a formal proof of Theorem 2.

(If) Suppose first that for every player i there is a disjoint correspondence $f_i : T_i^*(t_j^*) \twoheadrightarrow R_i^*(r_j^*)$, with $r_j^* \in f_j(t_j^*)$, such that these correspondences satisfy condition (1). We show that $h_j(t_j^*) = h_j(r_j^*)$. In fact, we will show for every player i, every $t_i \in T_i^*(t_j^*)$ and every $r_i \in f_i(t_i)$, that $h_i(t_i) = h_i(r_i)$. In order to show the latter, we prove, by induction on n, that for every player i, every $t_i \in T_i^*(t_j^*)$ and every $r_i \in f_i(t_i)$, we have that $h_i^n(t_i) = h_i^n(r_i)$.

Consider first the case $n = 1$. Take some $t_i \in T_i^*(t_j^*)$ and some $r_i \in f_i(t_i)$. We will show that $h_i^1(t_i) = h_i^1(r_i)$.

By definition, $h_i^1(t_i)$ and $h_i^1(r_i)$ are both in $\Delta(C_{-i})$. We define $[T_{-i}^*(t_j^*)]_{f_{-i}} := \{[t_{-i}]_{f_{-i}} \mid t_{-i} \in T_{-i}^*(t_j^*)\}$.We then have, for every $c_{-i} \in C_{-i}$,

$$\begin{aligned} h_i^1(t_i)(c_{-i}) &= b_i(t_i)(\{c_{-i}\} \times T_{-i}) \\ &= b_i(t_i)(\{c_{-i}\} \times T_{-i}^*(t_j^*)) \\ &= \sum_{[t_{-i}]_{f_{-i}} \in [T_{-i}^*(t_j^*)]_{f_{-i}}} b_i(t_i)(\{c_{-i}\} \times [t_{-i}]_{f_{-i}}) \\ &= \sum_{[t_{-i}]_{f_{-i}} \in [T_{-i}^*(t_j^*)]_{f_{-i}}} \beta_i(r_i)(\{c_{-i}\} \times f_{-i}(t_{-i})) \\ &= \beta_i(r_i)(\{c_{-i}\} \times R_{-i}) \\ &= h_i^1(r_i)(c_{-i}), \end{aligned}$$

which implies that $h_i^1(t_i) = h_i^1(r_i)$. Here, the second equality follows from the observation that t_i only assigns positive probability to opponents' type combinations in $T_{-i}^*(t_j^*)$, as $t_i \in T_i^*(t_j^*)$.

The third equality follows from the fact that $T^*_{-i}(t^*_j)$ is the disjoint union of the sets $[t_{-i}]_{f_{-i}}$ where $[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}}$. The fourth equality follows from condition (1). The fifth equality follows from Lemma 6, which states that $\beta_i(r_i)$ only assigns positive probability to opponents' type combinations r_{-i} which are in some set $f_{-i}(t_{-i})$ for some $[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}}$, and the fact that f_{-i} is a disjoint correspondence. This completes the induction start, for $n = 1$.

Consider now some $n \geq 2$ and suppose, by the induction assumption, that for every player i, every $t_i \in T^*_i(t^*_j)$ and every $r_i \in f_i(t_i)$, we have that $h^{n-1}_i(t_i) = h^{n-1}_i(r_i)$. This implies that $h^{n-1}_{-i}(t'_{-i}) = h^{n-1}_{-i}(t_{-i})$ for every $t'_{-i} \in [t_{-i}]_{f_{-i}}$. To see this, let $r_{-i} \in f_{-i}(t_{-i})$, and consider some $t'_{-i} \in [t_{-i}]_{f_{-i}}$. Then, as $f_{-i}(t'_{-i}) = f_{-i}(t_{-i})$, it follows that $r_{-i} \in f_{-i}(t'_{-i})$. By the induction assumption, we thus have that $h^{n-1}_{-i}(t'_{-i}) = h^{n-1}_{-i}(r_{-i}) = h^{n-1}_{-i}(t_{-i})$. As a direct consequence, it holds for every $h^{n-1}_{-i} \in h^{n-1}_{-i}[T_{-i}]$ that

$$T^*_{-i}(t^*_j) \cap T_{-i}[h^{n-1}_{-i}] = \bigcup_{[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}} : h^{n-1}_{-i}(t_{-i}) = h^{n-1}_{-i}} [t_{-i}]_{f_{-i}}, \tag{2}$$

where this union is disjoint,

We will now show that $h^n_i(t_i) = h^n_i(r_i)$, for every $t_i \in T^*_i(t^*_j)$ and every $r_i \in f_i(t_i)$. By construction, we have that $h^n_i(t_i) \in \Delta(C_{-i} \times h^{n-1}_{-i}(T^*_{-i}(t^*_j)))$ and $h^n_i(r_i) \in \Delta(C_{-i} \times h^{n-1}_{-i}(R^*_{-i}(r^*_j)))$, where

$$h^{n-1}_{-i}(T^*_{-i}(t^*_j)) = \times_{k \neq i}\{h^{n-1}_k(t_k) \mid t_k \in T^*_k(t^*_j)\},$$

and similarly for $h^{n-1}_{-i}(R^*_{-i}(r^*_j))$.

For every $c_{-i} \in C_{-i}$ and every $h^{n-1}_{-i} \in h^{n-1}_{-i}(T^*_{-i}(t^*_j))$ we have

$$\begin{aligned} h^n_i(t_i)(c_{-i}, h^{n-1}_{-i}) &= b_i(t_i)(\{c_{-i}\} \times T_{-i}[h^{n-1}_{-i}]) \\ &= b_i(t_i)(\{c_{-i}\} \times (T^*_{-i}(t^*_j) \cap T_{-i}[h^{n-1}_{-i}])) \\ &= \sum_{[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}} : h^{n-1}_{-i}(t_{-i}) = h^{n-1}_{-i}} b_i(t_i)(\{c_{-i}\} \times [t_{-i}]_{f_{-i}}) \\ &= \sum_{[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}} : h^{n-1}_{-i}(t_{-i}) = h^{n-1}_{-i}} \beta_i(r_i)(\{c_{-i}\} \times f_{-i}(t_{-i})) \\ &= \beta_i(r_i)(\{c_{-i}\} \times R_{-i}[h^{n-1}_{-i}]) \\ &= h^n_i(r_i)(c_{-i}, h^{n-1}_{-i}), \end{aligned}$$

which implies that $h^n_i(t_i) = h^n_i(r_i)$ for all $r_i \in f_i(t_i)$. Here, the second equality follows from the fact that $b_i(t_i)$ only assigns positive probability to opponents' type combinations in $T^*_{-i}(t^*_j)$, as $t_i \in T^*_i(t^*_j)$. The third equality follows from (2). The fourth equality follows from condition (1). The fifth equality follows from Lemma 6, which states that $\beta_i(r_i)$ only assigns positive probability to opponents' type combinations r_{-i} which are in some set $f_{-i}(t_{-i})$ for some $[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}}$, and the induction assumption, which implies that $h^{n-1}_{-i}(r_{-i}) = h^{n-1}_{-i}(t_{-i})$ for all $t_{-i} \in T^*_{-i}(t^*_j)$ and all $r_{-i} \in f_{-i}(t_{-i})$.

Hence, we have shown that $h^n_i(t_i) = h^n_i(r_i)$ for all $t_i \in T^*_i(t^*_j)$ and all $r_i \in f_i(t_i)$. By induction on n, we may conclude that $h_i(t_i) = h_i(r_i)$ for all players i, all $t_i \in T^*_i(t^*_j)$ and all $r_i \in f_i(t_i)$. In particular, since $r^*_j \in f_j(t^*_j)$, we may conclude that $h_j(t^*_j) = h_j(r^*_j)$, which was to show.

(Only if) Suppose now that $h_j(t^*_j) = h_j(r^*_j)$. We prove that for every player i there is a disjoint correspondence $f_i : T^*_i(t^*_j) \twoheadrightarrow R^*_i(r^*_j)$, with $r^*_j \in f_j(t^*_j)$, such that these correspondences satisfy condition (1).

As $h_j(t^*_j) = h_j(r^*_j)$, we know by Lemma 5 that for every player i, and every $t_i \in T^*_i(t^*_j)$, there is at least one $r_i \in R^*_i(r^*_j)$ with $h_i(t_i) = h_i(r_i)$. For every player i, let $f_i : T^*_i(t^*_j) \twoheadrightarrow R^*_i(r^*_j)$ be the correspondence given by

$$f_i(t_i) := \{r_i \in R^*_i(r^*_j) \mid h_i(r_i) = h_i(t_i)\}$$

for every $t_i \in T^*_i(t^*_j)$. Then, by the insight above, $f_i(t_i)$ is non-empty for every $t_i \in T^*_i(t^*_j)$. Moreover, for every $t_i, t'_i \in T^*_i(t^*_j)$ either $f_i(t_i) = f_i(t'_i)$ or $f_i(t_i) \cap f_i(t'_i) = \emptyset$. So, f_i is a disjoint correspondence. Note also that $r^*_j \in f_j(t^*_j)$ since $h_j(r^*_j) = h_j(t^*_j)$.

It remains to prove that these correspondences f_i satisfy the condition (1).

By Corollary 4, there is some $n \geq 1$ such that, for every player i, every $t_i \in T_i$ and every $r_i \in R_i$,

$$h_i(t_i) = h_i(r_i) \text{ if and only if } h^n_i(t_i) = h^n_i(r_i).$$

Now, choose $n \geq 1$ in this way. Then, for every $t_{-i} \in T^*_{-i}(t^*_j)$ we have that

$$\begin{aligned} [t_{-i}]_{f_{-i}} &= \{t'_{-i} \in T^*_{-i}(t^*_j) \mid f_{-i}(t'_{-i}) = f_{-i}(t_{-i})\} \\ &= \{t'_{-i} \in T^*_{-i}(t^*_j) \mid h_{-i}(t'_{-i}) = h_{-i}(t_{-i})\} \\ &= \{t'_{-i} \in T^*_{-i}(t^*_j) \mid h^n_{-i}(t'_{-i}) = h^n_{-i}(t_{-i})\}. \end{aligned} \tag{3}$$

Now, choose some $t_i \in T^*_i(t^*_j)$, some $t_{-i} \in T^*_{-i}(t^*_j)$, and some $r_i \in f_i(t_i)$. As, by construction, $f_i(t_i) := \{r'_i \in R^*_i(r^*_j) \mid h_i(r'_i) = h_i(t_i)\}$, it follows that $h_i(r_i) = h_i(t_i)$. For every $c_{-i} \in C_{-i}$, we have that

$$
\begin{aligned}
b_i(t_i)(\{c_{-i}\} \times [t_{-i}]_{f_{-i}}) &= b_i(t_i)(\{c_{-i}\} \times \{t'_{-i} \in T^*_{-i}(t^*_j) \mid h^n_{-i}(t'_{-i}) = h^n_{-i}(t_{-i})\}) \\
&= b_i(t_i)(\{c_{-i}\} \times \{t'_{-i} \in T_{-i} \mid h^n_{-i}(t'_{-i}) = h^n_{-i}(t_{-i})\}) \\
&= b_i(t_i)(\{c_{-i}\} \times T_{-i}[h^n_{-i}(t_{-i})]) \\
&= h^{n+1}_i(t_i)(c_{-i}, h^n_{-i}(t_{-i})) \\
&= h^{n+1}_i(r_i)(c_{-i}, h^n_{-i}(t_{-i})) \\
&= \beta_i(r_i)(\{c_{-i}\} \times R_{-i}[h^n_{-i}(t_{-i})]) \\
&= \beta_i(r_i)(\{c_{-i}\} \times \{r_{-i} \in R_{-i} \mid h^n_{-i}(r_{-i}) = h^n_{-i}(t_{-i})\}) \\
&= \beta_i(r_i)(\{c_{-i}\} \times \{r_{-i} \in R_{-i} \mid h_{-i}(r_{-i}) = h_{-i}(t_{-i})\}) \\
&= \beta_i(r_i)(\{c_{-i}\} \times \{r_{-i} \in R^*_{-i}(r^*_j) \mid h_{-i}(r_{-i}) = h_{-i}(t_{-i})\}) \\
&= \beta_i(r_i)(\{c_{-i}\} \times f_{-i}(t_{-i}))
\end{aligned}
$$

which establishes condition (1). Here, the first equality follows from (3). The second equality follows from the observation that $b_i(t_i)$ only assigns positive probability to opponents' type combinations in $T^*_{-i}(t^*_j)$, since $t_i \in T^*_i(t^*_j)$. The third equality follows from the definition of $T_{-i}[h^n_{-i}(t_{-i})]$. The fourth equality follows from the definition of $h^{n+1}_i(t_i)$. The fifth equality follows from the observation above that $h_i(r_i) = h_i(t_i)$, which implies that $h^{n+1}_i(r_i) = h^{n+1}_i(t_i)$. The sixth equality follows from the definition of $h^{n+1}_i(r_i)$. The seventh equality follows from the definition of $R_{-i}[h^n_{-i}(t_{-i})]$. The eighth equality follows from the fact that, by the choice of n, $h^n_{-i}(r_{-i}) = h^n_{-i}(t_{-i})$ if and only if $h_{-i}(r_{-i}) = h_{-i}(t_{-i})$. The nineth equality follows from the observation that $\beta_i(r_i)$ only assigns positive probability to opponents' type combinations in $R^*_{-i}(r^*_j)$, as $r_i \in R^*_i(r^*_j)$. The tenth equality follows from the fact that $f_{-i}(t_{-i}) = \{r_{-i} \in R_{-i} \mid h_{-i}(r_{-i}) = h_{-i}(t_{-i})\}$. The proof is hereby complete. ■

6 Special Cases

In this section we will discuss the implications of our main result for some interesting special cases. We first look at the case where two different types within an epistemic model always induce different belief hierarchies, and afterwards we discuss the case where we compare two types from the same epistemic model.

6.1 Non-Redundant Type Spaces

Say that an epistemic model $M = (T_i, b_i)_{i \in I}$ is *non-redundant* if different types induce different belief hierarchies. That is, for every player i, and every two different types $t_i, t'_i \in T_i$, we have that $h_i(t_i) \neq h_i(t'_i)$. Now, consider a special setting in which the two epistemic models

$M = (T_i, b_i)_{i\in I}$ and $M' = (R_i, \beta_i)_{i\in I}$ in Theorem 2 are non-redundant. Suppose we find some correspondences f_i which satisfy condition (1). Then, we know from the proof of Theorem 2 that for every $t_i \in T_i^*(t_j^*)$, all the types in $f_i(t_i)$ induce the same belief hierarchy as t_i. However, as both epistemic models are non-redundant, it must be the case that (a) $f_i(t_i)$ only contains one type, and (b) that there is no other $t_i' \in T_i^*(t_j^*)$ with $f_i(t_i') = f_i(t_i)$. This means that f_i must be a *one-to-one function* from $T_i^*(t_j^*)$ to $R_i^*(r_j^*)$, which maps every $t_i \in T_i^*(t_j^*)$ to a single type $f_i(t_i)$ in $R_i^*(r_j^*)$, and which maps different types in $T_i^*(t_j^*)$ to different types in $R_i^*(r_j^*)$. We thus obtain the following result.

Corollary 7 (Non-Redundant Type Spaces) *Consider a finite static game $\Gamma = (C_i, u_i)_{i\in I}$, and two non-redundant finite epistemic models $M = (T_i, b_i)_{i\in I}$ and $M' = (R_i, \beta_i)_{i\in I}$ for Γ. For a given player j, consider a type $t_j^* \in T_j$ and a type $r_j^* \in R_j$.*

Then, t_j^* and r_j^* induce the same belief hierarchy, if and only if, for every player i there is a one-to-one function

$$f_i : T_i^*(t_j^*) \to R_i^*(r_j^*)$$

such that $f_j(t_j^*) = r_j^*$ and

$$b_i(t_i)(c_{-i}, t_{-i}) = \beta_i(f_i(t_i))(c_{-i}, f_{-i}(t_{-i})) \tag{4}$$

for all players i, all types $t_i \in T_i^*(t_j^*)$, all $c_{-i} \in C_{-i}$, and all $t_{-i} \in T_{-i}^*(t_j^*)$.

The main theorem can also be applied to other special cases of this kind. Consider, for instance, a setting where the epistemic model M is redundant, but where M' is non-redundant. Then, the correspondences f_i from Theorem 2 are actually *functions* from $T_i^*(t_j^*)$ to $R_i^*(r_j^*)$ – that is, the image sets $f_i(t_i)$ are single-valued – but these functions f_i will no longer be one-to-one. The analogue to condition (1) would then state that

$$b_i(t_i)(c_{-i}, [t_{-i}]_{f_{-i}}) = \beta_i(f_i(t_i))(c_{-i}, f_{-i}(t_{-i})),$$

where $[t_{-i}]_{f_{-i}} = \{t'_{-i} \in T_{-i}^*(t_j^*) \mid f_{-i}(t'_{-i}) = f_{-i}(t_{-i})\}$.

6.2 Two Types Within the Same Epistemic Model

In Theorem 2 there is nothing that prevents us from choosing the second epistemic model, M', equal to the first one, M. In that case, we would be checking whether two types t_j^* and r_j^* within the *same* epistemic model $M = (T_i, b_i)_{i\in I}$ would induce the same belief hierarchy or not.

Now, if $T^*(t_j^*) \neq T^*(r_j^*)$, it would be *as if* we would still be considering two different epistemic models. To see this, note that the restriction of M to $T^*(t_j^*)$ is an epistemic model by itself, since $T^*(t_j^*)$ is belief-closed, and so is the restriction of M to $T^*(r_j^*)$. Let $\hat{M}$ be the restriction of M to $T^*(t_j^*)$, and $\hat{M}'$ the restriction of M to $T^*(r_j^*)$. Then, the two models $\hat{M}$ and $\hat{M}'$ would be different since $T^*(t_j^*) \neq T^*(r_j^*)$. As Theorem 2 and its proof only depend on $\hat{M}$ and $\hat{M}'$, it would indeed be as if we would apply Theorem 2 to two different epistemic models.

So, let us concentrate on the special case where $M = M'$ and $T^*(t_j^*) = T^*(r_j^*)$. What would Theorem 2 imply for that case? Suppose that for every player i we have a correspondence $f_i : T_i^*(t_j^*) \twoheadrightarrow R_i^*(r_j^*)$ such that these correspondences satisfy condition (1) in Theorem 2. As $R_i^*(r_j^*) = T_i^*(t_j^*)$, these would be correspondences from $T_i^*(t_j^*)$ to itself.

By Theorem 2, we know that condition (1) implies that $h_j(t_j^*) = h_j(r_j^*)$. Now, define for every player i a new correspondence $P_i : T_i^*(t_j^*) \twoheadrightarrow T_i^*(t_j^*)$ by

$$P_i(t_i) = \{t_i' \in T_i^*(t_j^*) \mid h_i(t_i') = h_i(t_i)\}.$$

Then, we know from the proof of Theorem 2 that these correspondences P_i satisfy condition (1) as well. Moreover, as $h_j(t_j^*) = h_j(r_j^*)$, we conclude that $r_j^* \in P_j(t_j^*)$.

Now, it can easily be verified that the collection of sets $\{P_i(t_i) \mid t_i \in T_i^*(t_j^*)\}$ actually corresponds to a *partition* of the set $T_i^*(t_j^*)$, with $t_i \in P_i(t_i)$ for every $t_i \in T_i^*(t_j^*)$. So, the correspondence P_i actually is a partition, where $P_i(t_i)$ denotes the partition element to which t_i belongs.

Moreover, for every $t_i \in T_i^*(t_j^*)$ we have, by definition, that

$$[t_i]_{P_i} = \{t_i' \in T_i^*(t_j^*) \mid P_i(t_i') = P_i(t_i)\} = P_i(t_i),$$

since P_i is a partition. But then, condition (1) is equivalent to stating that

$$b_i(t_i)(\{c_{-i}\} \times P_{-i}(t_{-i})) = b_i(t_i')(\{c_{-i}\} \times P_{-i}(t_{-i})) \text{ for all } t_i' \in P_i(t_i), \tag{5}$$

where $P_{-i}(t_{-i}) := \times_{k \neq i} P_k(t_k)$ for every $t_{-i} = (t_k)_{k \neq i}$ in $T_{-i}^*(t_j^*)$.

So, the conclusion is that, whenever there are correspondences f_i that satisfy condition (1), then there are also partitions P_i of the sets $T_i^*(t_j^*)$ that satisfy condition (5). The converse is also true: If there are partitions P_i that satisfy (5), then these partitions P_i are, in particular, correspondences that satisfy condition (1). We therefore arrive at the following result.

Corollary 8 (Types Within the Same Epistemic Model) *Consider a finite static game $\Gamma = (C_i, u_i)_{i \in I}$, and a finite epistemic model $M = (T_i, b_i)_{i \in I}$ for Γ. For a given player j, consider two types $t_j^*, r_j^* \in T_j$ with $T^*(t_j^*) = T^*(r_j^*)$.*

Then, t_j^* and r_j^* induce the same belief hierarchy, if and only if, for every player i there is a partition P_i of the set $T_i^*(t_j^*)$ such that $r_j^* \in P_j(t_j^*)$ and

$$b_i(t_i)(\{c_{-i}\} \times P_{-i}(t_{-i})) = b_i(t_i')(\{c_{-i}\} \times P_{-i}(t_{-i})) \text{ for all } t_i' \in P_i(t_i)$$

for all players i, all types $t_i \in T_i^*(t_j^*)$, all $c_{-i} \in C_{-i}$, and all $t_{-i} \in T_{-i}^*(t_j^*)$.

So, the necessary and sufficient condition in this case reads as follows: Two types t_j^* and r_j^* induce the same belief hierarchy precisely when we are able to partition the type sets $T_i^*(t_j^*)$ into equivalence classes such that (a) t_j^* and r_j^* are in the same equivalence class, and (b) different types within the same equivalence class always hold the same belief on the opponents' choices and the opponents' equivalence classes of types.

7 Possible Extensions

My conjecture is that Theorem 2 can be extended in various different directions, some of which I will discuss now.

Infinite type spaces. An important restriction we impose in the present framework is that the epistemic models contain finitely many types only.

In fact, what we *really* need for Theorem 2 to work is that the sets $T^*(t_j^*)$ and $R^*(r_j^*)$ – containing the types that enter t_j^*'s belief hierarchy and r_j^*'s belief hierarchy – are finite. So, even if the epistemic model M itself contains infinitely many types, Theorem 2 would still be valid as long as $T^*(t_j^*)$ and $R^*(r_j^*)$ contain finitely many types only. Also the proof would remain unchanged in this case.

To see this, note that the restriction of M to $T^*(t_j^*)$ is an epistemic model itself, since $T^*(t_j^*)$ is a belief-closed subspace of M. By this, we mean that every type in $T^*(t_j^*)$ only assigns positive probability to opponents' types that are in in $T^*(t_j^*)$ as well. In fact, $T^*(t_j^*)$ is the smallest belief-closed subspace of M that contains the type t_j^*. Similarly, also the restriction of M' to $R^*(r_j^*)$ is an epistemic model, by the same arguments. But then, instead of considering the whole epistemic models M and M', we could just consider their restrictions to $T^*(t_j^*)$ and $R^*(r_j^*)$, which are finite epistemic models. As Theorem 2 and the proof take place entirely within these subspaces $T^*(t_j^*)$ and $R^*(r_j^*)$, Theorem 2 would still be valid in this case.

But what happens if we allow $T^*(t_j^*)$ or $R^*(r_j^*)$ to contain *infinitely* many types? Can we then still obtain a characterization result *similar* to Theorem 2. I believe so, but certain proof techniques in the present paper will no longer work, as they heavily depend on the assumption that $T^*(t_j^*)$ and $R^*(r_j^*)$ are finite. For instance, Corollary 4 – which plays an important role in the proof – will no longer hold in that case, even if we would restrict attention to the epistemic submodels induced by $T^*(t_j^*)$ and $R^*(t_j^*)$. Moreover, topological issues start playing a role when we allow $T^*(t_j^*)$ and $R^*(r_j^*)$ to be infinite, especially when these sets are not countable.

But how severe is the restriction that $T^*(t_j^*)$ and $R^*(r_j^*)$ must be finite? This, of course, depends on the purpose one has in mind. It is well-known that many concepts in game theory, like *rationalizability* (Bernheim (1984), Pearce (1984)), *common belief in rationality* (Brandenburger and Dekel (1987), Tan and Werlang (1988)), *Nash equilibrium* (Nash (1950, 1951)), the *Dekel-Fudenberg procedure* (Dekel and Fudenberg (1990)), *proper rationalizability* (Schuhmacher (1999), Asheim (2001)) and *common belief in future rationality* (Perea (2014)) can be characterized by means of finite epistemic models. Perea (2012) shows that even concepts like *iterated* *assumption of rationality* (Brandenburger, Friendenberg and Keisler (2008)) and *common strong belief in rationality* (Battigalli and Siniscalchi (2002)), which are originally defined within complete – and hence infinite – type spaces, can also be characterized within finite epistemic models. Therefore, finite epistemic models would in principle suffice if the purpose is to investigate any of these concepts. But there may be other scenarios where infinite epistemic models turn out to be indispensable. For these scenarios it would then be crucial to have an analogue of Theorem 2 for the case where $T^*(t_j^*)$ and $R^*(r_j^*)$ are infinite.

Finite belief hierarchies. In many situations of interest, it may simply be too demanding to require that players hold *infinitely* many levels of belief. It is therefore important to model type spaces in which certain types only hold beliefs up to a certain level n. See Kets (2010, 2013) and Heifetz and Kets (2013) for a thorough analysis of this phenomenon. An interesting question is whether Theorem 2 can be extended to such settings where some types only induce a finite belief hierarchy. In particular, consider two epistemic models M and M' containing types with a finite belief hierarchy, and consider two types, t_j^* in M and r_j^* in M', which only hold beliefs up to some level n. Can we find necessary and sufficient conditions such that t_j^* and r_j^* induce exactly the same (finite) belief hierarchy? I believe this may be an interesting problem to be addressed in future research.

Beliefs about other parameters. The belief hierarchies and the epistemic models in this paper concern only the players' beliefs about the *choices* in the game. But what if we also wish to model beliefs about *other* parameters in the game, for instance beliefs about the opponents' utilities? Theorem 2 can easily be adapted to such an alternative structure, simply by replacing the set of opponents' choices – which now serves as the primary space of uncertainty for a player – by a different set which includes the opponents' utilities, or any other parameters one wishes to include. As far as I can see, the proof would look exactly the same, as long as one keeps the epistemic models finite.

Alternative notions of belief. One could also try to extend Theorem 2 to more general notions of belief, such as *lexicographic beliefs* (Blume, Brandenburger and Dekel (1991a, 1991b)) and *conditional beliefs* in dynamic games (Ben-Porath (1997), Battigalli and Siniscalchi (1999, 2002)). If one keeps the epistemic models finite, I expect that a similar result – and a similar proof – should be possible for these settings as well.

Kripke-Aumann structures. Epistemic models with types are not the only way to encode belief hierarchies. One can also use models with states of the world, à la Kripke (1963) and Aumann (1976), to represent belief hierarchies. My feeling is that Theorem 2 can be adapted to such models as well.

8 Proofs

Proof of Lemma 3. We prove the statement by induction on n. Consider first the case where $n = 2$.

Take two types $t_i \in T_i$ and $r_i \in R_i$ with $h_i^2(t_i) = h_i^2(r_i)$. We show that $h_i^1(t_i) = h_i^1(r_i)$. Remember that $h_i^1(t_i)$ and $h_i^1(r_i)$ are both in $\Delta(C_{-i})$. For every $c_{-i} \in C_{-i}$,

$$\begin{aligned} h_i^1(t_i)(c_{-i}) &= b_i(t_i)(\{c_{-i}\} \times T_{-i}) \\ &= \sum_{h_{-i}^1 \in h_{-i}^1(T_{-i}) \cup h_{-i}^1(R_{-i})} b_i(t_i)(\{c_{-i}\} \times T_{-i}[h_{-i}^1]) \\ &= \sum_{h_{-i}^1 \in h_{-i}^1(T_{-i}) \cup h_{-i}^1(R_{-i})} h_i^2(t_i)(c_{-i}, h_{-i}^1) \\ &= \sum_{h_{-i}^1 \in h_{-i}^1(T_{-i}) \cup h_{-i}^1(R_{-i})} h_i^2(r_i)(c_{-i}, h_{-i}^1) \\ &= \sum_{h_{-i}^1 \in h_{-i}^1(T_{-i}) \cup h_{-i}^1(R_{-i})} \beta_i(r_i)(\{c_{-i}\} \times R_{-i}[h_{-i}^1]) \\ &= \beta_i(r_i)(\{c_{-i}\} \times R_{-i}) \\ &= h_i^1(r_i)(c_{-i}), \end{aligned}$$

which implies that $h_i^1(t_i) = h_i^1(r_i)$. Here, the fourth equality follows from the assumption that $h_i^2(t_i) = h_i^2(r_i)$.

Take now some $n \geq 3$, and suppose that the statement is true for $n-1$, for all players i. Consider some type $t_i \in T_i$ and some type $r_i \in R_i$ with $h_i^n(t_i) = h_i^n(r_i)$. We show that $h_i^{n-1}(t_i) = h_i^{n-1}(r_i)$.

Let

$$T_{-i}(t_i) := \{t_{-i} \in T_{-i} \mid b_i(t_i)(C_{-i} \times \{t_{-i}\}) > 0\}$$

be the set of opponents' type combinations to which t_i assigns positive probability. Similarly, we define $R_{-i}(r_i)$. To show that $h_i^{n-1}(t_i) = h_i^{n-1}(r_i)$, we first prove that

$$h_{-i}^{n-1}(T_{-i}(t_i)) = h_{-i}^{n-1}(R_{-i}(r_i)), \tag{6}$$

and, for every $h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i})$,

$$h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = h_{-i}^{n-2}(R_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\} \tag{7}$$

for some $h_{-i}^{n-2} \in h_{-i}^{n-2}(T_{-i})$. Here, $h_{-i}^{n-1}(T_{-i}(t_i)) = \{h_{-i}^{n-1}(t_{-i}) \mid t_{-i} \in T_{-i}(t_i)\}$ and $h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}(t_{-i}) \mid t_{-i} \in T_{-i}[h_{-i}^{n-1}]\}$, and similarly for R_{-i}.

We first show (6). By definition, for every $h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i})$,

$$\begin{aligned} h_i^n(t_i)(C_{-i} \times \{h_{-i}^{n-1}\}) &= b_i(t_i)(C_{-i} \times T_{-i}[h_{-i}^{n-1}]) \\ &= b_i(t_i)(C_{-i} \times \{t_{-i} \in T_{-i} \mid h_{-i}^{n-1}(t_{-i}) = h_{-i}^{n-1}\}) \\ &= b_i(t_i)(C_{-i} \times \{t_{-i} \in T_{-i}(t_i) \mid h_{-i}^{n-1}(t_{-i}) = h_{-i}^{n-1}\}), \end{aligned} \tag{8}$$

where the third equality follows from the fact that $b_i(t_i)$ only assigns positive probability to types in $T_{-i}(t_i)$. In fact, $b_i(t_i)$ assigns positive probability *precisely* to those types that are in $T_{-i}(t_i)$. Hence, it follows from (8) that $h_i^n(t_i)(C_{-i} \times \{h_{-i}^{n-1}\}) > 0$ if and only if there is some $t_{-i} \in T_{-i}(t_i)$ with $h_{-i}^{n-1}(t_{-i}) = h_{-i}^{n-1}$, which is the case precisely when $h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i))$.

In a similar way, it follows that $h_i^n(r_i)(C_{-i} \times \{h_{-i}^{n-1}\}) > 0$ if and only if $h_{-i}^{n-1} \in h_{-i}^{n-1}(R_{-i}(r_i))$. Since, by the induction assumption, $h_i^n(t_i) = h_i^n(r_i)$, it follows that $h_{-i}^{n-1}(T_{-i}(t_i)) = h_{-i}^{n-1}(R_{-i}(r_i))$, and hence (6) holds.

We now prove (7). We first show that $h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}$ for some $h_{-i}^{n-2} \in h_{-i}^{n-2}(T_{-i})$. Take two type combinations $t_{-i}, t'_{-i} \in T_{-i}[h_{-i}^{n-1}]$. That is, $h_{-i}^{n-1}(t_{-i}) = h_{-i}^{n-1}(t'_{-i}) = h_{-i}^{n-1}$. Then, by the induction assumption, it follows that $h_{-i}^{n-2}(t_{-i}) = h_{-i}^{n-2}(t'_{-i})$. So, all type combinations in $T_{-i}[h_{-i}^{n-1}]$ induce the same combination of $(n-1)$-th order beliefs, which we call h_{-i}^{n-2}. So, $h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}$.

Next we show that $h_{-i}^{n-2}(R_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}$ as well. Take some $r_{-i} \in R_{-i}[h_{-i}^{n-1}]$ and some $t_{-i} \in T_{-i}[h_{-i}^{n-1}]$. As $h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}$, it follows that $h_{-i}^{n-2}(t_{-i}) = h_{-i}^{n-2}$. Since $h_{-i}^{n-1}(r_{-i}) = h_{-i}^{n-1} = h_{-i}^{n-1}(t_{-i})$, it follows by the induction assumption that $h_{-i}^{n-2}(r_{-i}) = h_{-i}^{n-2}(t_{-i}) = h_{-i}^{n-2}$. So, we may conclude that $h_{-i}^{n-2}(R_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}$. We have thus shown (7).

We now prove that $h_i^{n-1}(t_i) = h_i^{n-1}(r_i)$. By (7) we know that for every $h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i})$ there is some $h_{-i}^{n-2} \in h_{-i}^{n-2}(T_{-i})$ with $h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}$. For every $h_{-i}^{n-2} \in h_{-i}^{n-2}(T_{-i}(t_i))$, we define

$$h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2}) := \{h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i)) \mid h_{-i}^{n-2}(T_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}\}.$$

In the same way, we define

$$h_{-i}^{n-1}(R_{-i}(r_i) \mid h_{-i}^{n-2}) := \{h_{-i}^{n-1} \in h_{-i}^{n-1}(R_{-i}(r_i)) \mid h_{-i}^{n-2}(R_{-i}[h_{-i}^{n-1}]) = \{h_{-i}^{n-2}\}\}.$$

By (6) and (7) it immediately follows that $h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2}) = h_{-i}^{n-1}(R_{-i}(r_i) \mid h_{-i}^{n-2})$.

Moreover, for every $h_{-i}^{n-2} \in h_{-i}^{n-2}(T_{-i})$,

$$T_{-i}[h_{-i}^{n-2}] \cap T_{-i}(t_i) = \bigcup_{h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2})} T_{-i}[h_{-i}^{n-1}] \cap T_{-i}(t_i). \tag{9}$$

So, for every $c_{-i} \in C_{-i}$ and $h_{-i}^{n-2} \in h_{-i}^{n-2}(T_{-i})$,

$$\begin{aligned}
h_i^{n-1}(t_i)(c_{-i}, h_{-i}^{n-2}) &= b_i(t_i)(\{c_{-i}\} \times T_{-i}[h_{-i}^{n-2}]) \\
&= b_i(t_i)(\{c_{-i}\} \times (T_{-i}[h_{-i}^{n-2}] \cap T_{-i}(t_i)) \\
&= \sum_{h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2})} b_i(t_i)(\{c_{-i}\} \times (T_{-i}[h_{-i}^{n-1}] \cap T_{-i}(t_i)) \\
&= \sum_{h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2})} b_i(t_i)(\{c_{-i}\} \times T_{-i}[h_{-i}^{n-1}]) \\
&= \sum_{h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2})} h_i^n(t_i)(c_{-i}, h_{-i}^{n-1}) \\
&= \sum_{h_{-i}^{n-1} \in h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2})} h_i^n(r_i)(c_{-i}, h_{-i}^{n-1}) \\
&= \sum_{h_{-i}^{n-1} \in h_{-i}^{n-1}(R_{-i}(r_i) \mid h_{-i}^{n-2})} h_i^n(r_i)(c_{-i}, h_{-i}^{n-1}) \\
&= h_i^{n-1}(r_i)(c_{-i}, h_{-i}^{n-2}),
\end{aligned}$$

which implies that $h_i^{n-1}(t_i) = h_i^{n-1}(r_i)$. Here, the second equality follows from the fact that $b_i(t_i)$ only assigns positive probability to type combinations in $T_{-i}(t_i)$. The third equality follows from (9). The fourth equality follows, again, from the fact that $b_i(t_i)$ only assigns positive probability to type combinations in $T_{-i}(t_i)$. The fifth equality follows from the definition of $h_i^n(t_i)$. The sixth equality follows from the assumption that $h_i^n(t_i) = h_i^n(r_i)$. The seventh equality follows from the observation above that $h_{-i}^{n-1}(T_{-i}(t_i) \mid h_{-i}^{n-2}) = h_{-i}^{n-1}(R_{-i}(r_i) \mid h_{-i}^{n-2})$. The eighth equality follows from mimicking the first five equalities, in reverse order, to

$$= \sum_{h_{-i}^{n-1} \in h_{-i}^{n-1}(R_{-i}(r_i) \mid h_{-i}^{n-2})} h_i^n(r_i)(c_{-i}, h_{-i}^{n-1}).$$

By induction on n, the proof is complete. ■

Proof of Lemma 5. For every player i and every $m \geq 1$, let $T_i^m(t_j^*) := T^m(t_j^*) \cap T_i$. Similarly, we define $R_i^m(r_j^*)$. We prove, by induction on m, that for every player i, and every $t_i \in T_i^m(t_j^*)$, there is some $r_i \in R_i^m(r_j^*)$ with $h_i(t_i) = h_i(r_i)$. From Corollary 4 we know that there is some $n \geq 1$ such that, for all players i, for every $t_i \in T_i$ and $r_i \in R_i$,

$$h_i(t_i) = h_i(r_i) \text{ if and only if } h_i^n(t_i) = h_i^n(r_i). \tag{10}$$

So, it is sufficient to show that for every $t_i \in T_i^m(r_j^*)$ there is some $r_i \in R_i^m(r_j^*)$ with $h_i^n(t_i) = h_i^n(r_i)$.

Consider first the case where $m = 1$. By definition, $T_j^1(t_j^*) = \{t_j^*\}$ and $R_j^1(r_j^*) = \{r_j^*\}$. As, by assumption, $h_j(t_j^*) = h_j(r_j^*)$, the statement holds for $T_j^1(t_j^*)$ and $R_j^1(r_j^*)$.

Take now some arbitrary player $i \neq j$. Then, by definition,

$$T_i^1(t_j^*) = \{t_i \in T_i \mid b_j(t_j^*)(C_{-j} \times \{t_i\} \times T_{-ij}) > 0,$$

and

$$R_i^1(r_j^*) = \{r_i \in R_i \mid \beta_j(r_j^*)(C_{-j} \times \{r_i\} \times R_{-ij}) > 0.$$

Take some arbitrary $t_i \in T_i^1(t_j^*)$. Then, $b_j(t_j^*)(C_{-j} \times \{t_i\} \times T_{-ij}) > 0$. Hence, there must be some $t_{-ij} \in T_{-ij}$ such that $b_j(t_j^*)(C_{-j} \times \{(t_i, t_{-ij})\}) > 0$. Let $t_{-j} := (t_i, t_{-ij})$. Hence, $b_j(t_j^*)(C_{-j} \times \{t_{-j}\}) > 0$. Now, choose n as in (10). Then,

$$\begin{aligned} h_j^{n+1}(t_j^*)(C_{-j} \times \{h_{-j}^n(t_{-j})\}) &= b_j(t_j^*)(C_{-j} \times T_{-j}[h_{-j}^n(t_{-j})]) \\ &\geq b_j(t_j^*)(C_{-j} \times \{t_{-j}\}) > 0, \end{aligned}$$

where the first inequality follows from the fact that $t_{-j} \in T_{-j}[h_{-j}^n(t_{-j})]$. As $h_j(t_j^*) = h_j(r_j^*)$, we must have that

$$h_j^{n+1}(t_j^*)(C_{-j} \times \{h_{-j}^n(t_{-j})\}) = h_j^{n+1}(r_j^*)(C_{-j} \times \{h_{-j}^n(t_{-j})\}),$$

and hence $h_j^{n+1}(r_j^*)(C_{-j} \times \{h_{-j}^n(t_{-j})\}) > 0$. Therefore,

$$\begin{aligned} h_j^{n+1}(r_j^*)(C_{-j} \times \{h_{-j}^n(t_{-j})\}) &= \beta_j(r_j^*)(C_{-j} \times R_{-j}[h_{-j}^n(t_{-j})]) \\ &= \beta_j(r_j^*)(C_{-j} \times \{r_{-j} \in R_{-j} \mid h_{-j}^n(r_{-j}) = h_{-j}^n(t_{-j})\}) > 0. \end{aligned}$$

Hence, there must be some $r_{-j} \in R_{-j}$ such that $\beta_j(r_j^*)(C_{-j} \times \{r_{-j}\}) > 0$ and $h_{-j}^n(r_{-j}) = h_{-j}^n(t_{-j})$. Let $r_{-j} = (r_i, r_{-ij})$. As $\beta_j(r_j^*)(C_{-j} \times \{r_{-j}\}) > 0$, it follows that $r_i \in R_i^1(r_j^*)$. Moreover, as $h_{-j}^n(r_{-j}) = h_{-j}^n(t_{-j})$, it follows that $h_i^n(r_i) = h_i^n(t_i)$. By (10) it then follows that $h_i(t_i) = h_i(r_i)$. So, we see that for every $t_i \in T_i^1(t_j^*)$ there is some $r_i \in R_i^1(r_j^*)$ with $h_i(t_i) = h_i(r_i)$. This completes the induction start, with $m = 1$.

Take now some $m \geq 2$ and suppose, by the induction assumption, that for every player i, and every $t_i \in T_i^{m-1}(t_j^*)$, there is some $r_i \in R_i^{m-1}(r_j^*)$ with $h_i(t_i) = h_i(r_i)$. We prove that for every player i, and every $t_i \in T_i^m(t_j^*)$, there is some $r_i \in R_i^m(r_j^*)$ with $h_i(t_i) = h_i(r_i)$.

Choose some $t_i \in T_i^m(t_j^*)$. Then, either $t_i \in T_i^{m-1}(t_j^*)$, or there is some player $k \neq i$ and some $t_k \in T_k^{m-1}(t_j^*)$ with $b_k(t_k)(C_{-k} \times \{t_i\} \times T_{-ik}) > 0$.

Consider first the case where $t_i \in T_i^{m-1}(t_j^*)$. Then, by the induction assumption, there is some $r_i \in R_i^{m-1}(r_j^*)$ with $h_i(t_i) = h_i(r_i)$. As $R_i^{m-1}(r_j^*) \subseteq R_i^m(r_j^*)$, it follows that there is some $r_i \in R_i^m(r_j^*)$ with $h_i(t_i) = h_i(r_i)$, which was to show.

Consider next the case where there is some player $k \neq i$ and some $t_k \in T_k^{m-1}(t_j^*)$ with $b_k(t_k)(C_{-k} \times \{t_i\} \times T_{-ik}) > 0$. By our induction assumption, we know that there is some

$r_k \in R_k^{m-1}(r_j^*)$ with $h_k(t_k) = h_k(r_k)$. Moreover, as $b_k(t_k)(C_{-k} \times \{t_i\} \times T_{-ik}) > 0$, there is some $t_{-ik} \in T_{-ik}$ such that $b_k(t_k)(C_{-k} \times \{(t_i, t_{-ik})\}) > 0$. Let $t_{-k} := (t_i, t_{-ik})$. Then, $b_k(t_k)(C_{-k} \times \{t_{-k}\}) > 0$. Choose n as in (10). Then,

$$\begin{aligned} h_k^{n+1}(t_k)(C_{-k} \times \{h_{-k}^n(t_{-k})\}) &= b_k(t_k)(C_{-k} \times T_{-k}[h_{-k}^n(t_{-k})]) \\ &\geq b_k(t_k)(C_{-k} \times \{t_{-k}\}) > 0, \end{aligned}$$

where the first inequality follows from the fact that $t_{-k} \in T_{-k}[h_{-k}^n(t_{-k})]$. As $h_k(t_k) = h_k(r_k)$, we must have that

$$h_k^{n+1}(t_k)(C_{-k} \times \{h_{-k}^n(t_{-k})\}) = h_k^{n+1}(r_k)(C_{-k} \times \{h_{-k}^n(t_{-k})\}),$$

and hence $h_k^{n+1}(r_k)(C_{-k} \times \{h_{-k}^n(t_{-k})\}) > 0$. Therefore,

$$\begin{aligned} h_k^{n+1}(r_k)(C_{-k} \times \{h_{-k}^n(t_{-k})\}) &= \beta_k(r_k)(C_{-k} \times R_{-k}[h_{-k}^n(t_{-k})]) \\ &= \beta_k(r_k)(C_{-k} \times \{r_{-k} \in R_{-k} \mid h_{-k}^n(r_{-k}) = h_{-k}^n(t_{-k})\}) > 0. \end{aligned}$$

Hence, there must be some $r_{-k} \in R_{-k}$ such that $\beta_k(r_k)(C_{-k} \times \{r_{-k}\}) > 0$ and $h_{-k}^n(r_{-k}) = h_{-k}^n(t_{-k})$. Let $r_{-k} = (r_i, r_{-ik})$. As $\beta_k(r_k)(C_{-k} \times \{r_{-k}\}) > 0$, it follows that $r_i \in R_i^1(r_k)$. Since $r_k \in R_k^{m-1}(r_j^*)$, we conclude that $r_i \in R_i^m(r_j^*)$. Moreover, as $h_{-k}^n(r_{-k}) = h_{-k}^n(t_{-k})$, it follows that $h_i^n(r_i) = h_i^n(t_i)$. By (10) it then follows that $h_i(t_i) = h_i(r_i)$. Remember that $r_i \in R_i^m(r_j^*)$. So, we see that for every $t_i \in T_i^m(t_j^*)$ there is some $r_i \in R_i^m(r_j^*)$ with $h_i(t_i) = h_i(r_i)$. This completes the induction step.

By induction on m, we can thus conclude that for every $t_i \in T_i^*(t_j^*)$ there is some $r_i \in R_i^*(r_j^*)$ with $h_i(t_i) = h_i(r_i)$. This completes the proof. ■

Proof of Lemma 6. We define

$$[T_{-i}^*(t_j^*)]_{f_{-i}} := \{[t_{-i}]_{f_{-i}} \mid t_{-i} \in T_{-i}^*(t_j^*)\}$$

as the set of equivalence classes in $T_{-i}^*(t_j^*)$ induced by f_{-i}. Then, for every $t_i \in T_i^*(t_j^*)$ and every $r_i \in f_i(t_i)$, we have that

$$\begin{aligned} 1 &= b_i(t_i)(C_{-i} \times T_{-i}) \\ &= b_i(t_i)(C_{-i} \times T_{-i}^*(t_j^*)) \\ &= \sum_{[t_{-i}]_{f_{-i}} \in [T_{-i}^*(t_j^*)]_{f_{-i}}} b_i(t_i)(C_{-i} \times [t_{-i}]_{f_{-i}}) \\ &= \sum_{[t_{-i}]_{f_{-i}} \in [T_{-i}^*(t_j^*)]_{f_{-i}}} \beta_i(r_i)(C_{-i} \times f_{-i}(t_{-i})). \end{aligned}$$

Here, the second equality follows from the fact that $b_i(t_i)$ only assigns positive probability to opponents' type combinations in $T^*_{-i}(t^*_j)$, as $t_i \in T^*_i(t^*_j)$. The third equality follows from the observation that $T^*_{-i}(t^*_j)$ is the disjoint union of the sets $[t_{-i}]_{f_{-i}}$, where $[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]$. The fourth equality, finally, follows from condition (1).

The equations above imply that

$$\sum_{[t_{-i}]_{f_{-i}} \in [T^*_{-i}(t^*_j)]_{f_{-i}}} \beta_i(r_i)(C_{-i} \times f_{-i}(t_{-i})) = 1,$$

from which it follows that $\beta_i(r_i)$ only assigns positive probability to opponents' type combinations that are in $f_{-i}(t_{-i})$ for some $t_{-i} \in T^*_{-i}(t^*_j)$. ■

References

[1] Asheim, G.B. (2001), Proper rationalizability in lexicographic beliefs, *International Journal of Game Theory* **30,** 453–478.

[2] Aumann, R.J. (1976), Agreeing to disagree, *Annals of Statistics* **4**, 1236–1239.

[3] Battigalli, P. and M. Siniscalchi (1999), Hierarchies of conditional beliefs and interactive epistemology in dynamic games, *Journal of Economic Theory* **88**, 188–230.

[4] Battigalli, P. and M. Siniscalchi (2002), Strong belief and forward induction reasoning, *Journal of Economic Theory* **106,** 356–391.

[5] Ben-Porath, E. (1997), Rationality, Nash equilibrium and backwards induction in perfect-information games, *Review of Economic Studies* **64,** 23–46.

[6] Bernheim, B.D. (1984), Rationalizable strategic behavior, *Econometrica* **52**, 1007–1028.

[7] Blume, L.E., Brandenburger, A. and E. Dekel (1991a), Lexicographic probabilities and choice under uncertainty, *Econometrica* **59**, 61–79.

[8] Blume, L.E., Brandenburger, A. and E. Dekel (1991b), Lexicographic probabilities and equilibrium refinements, *Econometrica* **59,** 81–98.

[9] Brandenburger, A. and E. Dekel (1987), Rationalizability and correlated equilibria, *Econometrica* **55**, 1391–1402.

[10] Brandenburger, A. and E. Dekel (1993), Hierarchies of beliefs and common knowledge, *Journal of Economic Theory* **59**, 189–198.

[11] Brandenburger, A., Friedenberg, A. and J. Keisler (2008), Admissibility in games, *Econometrica* **76**, 307–352.

[12] Dekel, E. and D. Fudenberg (1990), Rational behavior with payoff uncertainty, *Journal of Economic Theory* **52**, 243–267.

[13] Harsanyi, J.C. (1962), Bargaining in ignorance of the opponent's utility function, *Journal of Conflict Resolution* **6**, 29–38.

[14] Harsanyi, J.C. (1967–1968), Games with incomplete information played by "bayesian" players, I–III, *Management Science* **14**, 159–182, 320–334, 486–502.

[15] Heifetz, A. and W. Kets (2013), Robust multiplicity with a grain of naiveté, Working paper.

[16] Kets, W. (2010), Bounded reasoning and higher-order uncertainty, Working paper, Northwestern University.

[17] Kets, W. (2013), Finite depth of reasoning and equilibrium play in games with incomplete information, Working paper, Northwestern University.

[18] Kripke, S. (1963), A semantical analysis of modal logic I: Normal modal propositional calculi, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **9**, 67–96.

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

[20] Nash, J.F. (1950), Equilibrium points in N-person games, *Proceedings of the National Academy of Sciences of the United States of America* **36,** 48–49.

[21] Nash, J.F. (1951), Non-cooperative games, *Annals of Mathematics* **54**, 286–295.

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

[23] Perea, A. (2012), *Epistemic Game Theory: Reasoning and Choice,* Cambridge University Press.

[24] Perea, A. (2014), Belief in the opponents' future rationality, *Games and Economic Behavior* **83,** 231–254.

[25] Schuhmacher, F. (1999), Proper rationalizability and backward induction, *International Journal of Game Theory* **28**, 599–615.

[26] Tan, T. and S.R.C. Werlang (1988), The bayesian foundations of solution concepts of games, *Journal of Economic Theory* **45**, 370–391.