Bounded Reasoning and Higher-Order Uncertainty

Willemien Kets

September 15, 2021

Abstract

A standard assumption in game theory is that players have an infinite depth of reasoning: they think about what others think and about what others think that others think, and so on, ad infinitum. However, in practice, players may have a finite depth of reasoning. For example, a player may reason about what other players think, but not about what others think he thinks. This paper proposes a class of type spaces that generalizes the type space formalism due to Harsanyi (1968) so that it can model players with an arbitrary depth of reasoning. I show that the type space formalism does not impose any restrictions on the belief hierarchies that can be modeled, thus generalizing the classic result of Mertens and Zamir (1985). However, there is no universal type space that contains all type spaces.

JEL classification: C700, C720, D800, D830

Keywords: Bounded rationality, games, finite depth of reasoning, belief hierarchies, rationalizability, universal type space.

1 Introduction

In many situations of interest, players face uncertainty about their strategic environment. For example, a bidder in an auction may not know the values of other bidders for the object on sale, and a buyer may be uncertain which technology others adopt. The standard way to model this uncertainty uses the type spaces introduced by Harsanyi (1968): Players are endowed with a set of types, and every type is associated with a belief (i.e., a probability distribution) about the types of the other players and the primitive uncertainty (e.g., the payoff functions of the other players, or their actions). Harsanyi's type space approach provides a simple yet general analytical framework for studying all strategic situations where people have different information.

Implicit in the Harsanyi approach is the assumption that players have an infinite depth of reasoning: each type unwinds into an infinite hierarchy of beliefs that describes what the player thinks about the primitive uncertainty, what he thinks that the other players think, what he thinks that the other players think that their opponents think, and so on, ad infinitum.

Suppose now that a player does not reason "all the way," perhaps because of time constraints or limited cognitive ability. For example, suppose a player thinks about the primitive uncertainty and about what other players think about the primitive uncertainty, but does not think about what other players think about what he thinks. That is, the player has a *finite depth of reasoning*.

A priori, it is not clear that the convenience of a type-based framework remains available when we extend the standard framework to model players with a finite depth of reasoning. After all, type spaces involve a circularity, with types being associated with a belief about types, and it is precisely this circularity that makes that each type unwinds into an infinite hierarchy of beliefs. Is it possible to "block" this unwinding at some finite order?

The *first main contribution* of this paper is to show that this can be done. I introduce a class of type spaces that can model belief hierarchies of arbitrary (finite or infinite) depth. The key idea, which builds on the small-world idea of Savage (1954), is that types can differ in the level of detail they use to describe the relevant uncertainty: types with a high depth of reasoning are associated with a more refined description than types with a shallow depth. The description of the relevant uncertainty can be chosen in such a way that it specifies a player's belief only up to some finite order, thus "blocking" the unwinding of the type.

A natural question is whether the type space approach is sufficiently expressive to model all possible higher-order beliefs. For the Harsanyi case, this was shown by Mertens and Zamir (1985). The *second main contribution* of this paper is to extend this classic result to the present environment: any belief hierarchy (of arbitrary depth) can be represented by a type; and every

type generates a belief hierarchy. So, game-theoretic applications can use the more convenient type space formalism without imposing any restrictions on the belief hierarchies that can be modeled.

The *third main contribution* is to show that, unlike in the Harsanyi case, there is no type space that can capture all possible restrictions on players' beliefs. That is, there is no type space that is universal in the sense that it contains every type space as a belief-closed subset. For the case of rationalizability, this implies that if an analyst is interested in studying behavior across all strategic situations, then he needs to consider all type spaces.

Bounded reasoning has received considerable attention in the experimental and behavioral literature on level-k and cognitive hierarchy models; see Crawford, Costa-Gomes, and Iriberri (2013) for a survey. These models can successfully model behavior in a range of games. However, since they do not model beliefs and behavior separately, they are less suitable for a general analysis of the implications of bounded reasoning. This paper is the first to use Savage's small-world approach to provide a general model of bounded reasoning. An advantage of this approach is that it does not require that players think that others are less sophisticated than they are,[1] unlike other models of bounded reasoning (Strzalecki, 2014; Heifetz and Kets, 2018). This makes it possible to generalize equilibrium concepts to settings where players may be bounded in their reasoning (Kets, 2013).

This paper fits in with the literature that generalizes standard models of higher-order beliefs to model players who may not be fully sophisticated (Epstein and Wang, 1996; Di Tillio, 2008; Ganguli, Heifetz, and Lee, 2016).[2] This literature relaxes the standard assumption that individual preferences are based on beliefs that are representable by a probability measure so as to model players who are not probabilistically sophisticated (e.g., ambiguity averse). I maintain the standard assumption that beliefs can be represented by probability measures, but allow for a richer class of probability measures so as to model players who are bounded in their reasoning about beliefs. The present framework also bears some relation to the literature on unawareness in games (e.g., Heifetz, Meier, and Schipper, 2006) in the sense that it models players who do not think through all aspects of the game. However, the unawareness literature studies players who may be unaware of certain primitives of the game (e.g., actions, other players), while I consider settings where players do not form beliefs beyond a certain order.

The remainder of this paper is organized as follows. Section 2 provides an informal discussion of the results. The formal treatment is in Sections 3–7.

[1] For example, in the type space in Section 2.3, all types have the same finite depth of reasoning.

[2] Relatedly, Ahn (2007) models the beliefs of ambiguity averse players by assuming that at each order k, the kth-order belief of a player is given by a set of probability measures.

2 Heuristic treatment

2.1 Harsanyi type spaces

Two players, Ann (denoted by a) and Bob (b), are uncertain about some aspects of their strategic environment. For example, they may be uncertain about the other's payoff function or action. For concreteness, assume that each player $i = a, b$ has two *attributes*, labeled s_i^1, s_i^2, and that they are uncertain about the attribute of the other player, what the other thinks about their attribute, what the other thinks that they think, etcetera. In games with incomplete information, a player's attribute may refer to his payoff function; in epistemic game theory, it may refer to his action. In either case, the uncertainty can be represented by a Harsanyi type space. Given sets S_a, S_b of attributes for the players, a *Harsanyi type space* specifies a set T_i of *types* for each player $i = a, b$, and each type $t_i \in T_i$ is associated with a (subjective) *belief* π_{t_i}, that is, a probability distribution over the other player's attribute and type.

To give an example, suppose that Ann and Bob each have four types, labeled $t_a^1, t_a^2, t_a^3, t_a^4$, and $t_b^1, t_b^2, t_b^3, t_b^4$, respectively. Say that a type t_i *believes* an event E if it assigns probability 1 to it (i.e., $\pi_{t_i}(E) = 1$). Suppose that type t_a^1 for Ann believes that Bob has attribute $s_b = s_b^1$ and that he has type t_b^2 (i.e., $\pi_{t_a^1}(s_b^1, t_b^2) = 1$). Type t_a^2 believes that Bob has $s_b = s_b^1$ and that he has type t_b^3. Type t_a^3 believes that Bob has $s_b = s_b^2$ and that he has type t_b^4. Type t_a^4 believes that Bob has $s_b = s_b^2$ and that he has type t_b^1. The beliefs for Bob are symmetric. For example, type t_b^1 for Bob believes that Ann has $s_a = s_a^1$ and that she has type t_a^2 (i.e., $\pi_{t_b^1}(s_a^1, t_a^2) = 1$).

Every type unwinds into an infinite sequence of beliefs: For example, type t_a^1 for Ann believes that Bob has $s_b = s_b^1$. This is t_a^1's *first-order belief*, denoted $\mu_{t_a^1}^1$. Type t_a^1 also induces a *second-order belief* $\mu_{t_a^1}^2$, that is, a belief about Bob's first-order belief: t_a^1 believes that Bob believes that she has $s_a = s_a^1$ (as type t_a^1 assigns probability 1 to type t_b^2, which puts probability 1 on s_a^1). Type t_a^1 also induces a *third-order belief* $\mu_{t_a^1}^3$, that is, a belief about Bob's second-order belief: t_a^1 believes that Bob believes that she believes that he has $s_b = s_b^3$ (as t_b^2 puts probability 1 on t_a^3, which assigns probability 1 to s_b^2). It is easy to see that every type induces a kth-order belief $\mu_{t_a^1}^k$ (i.e., a belief about the other player's $(k-1)$th-order belief) for every k. Such an infinite sequence $h_a(t_a^1) := (\mu_{t_a^1}^1, \mu_{t_a^1}^2, \ldots)$ of beliefs is called an *(infinite) belief hierarchy*. Figure 1 shows the first few orders of beliefs generated by Ann's types, where a belief hierarchy generated by a type is represented by the list of statements (e.g., "Bob has $s_b = s_b^2$") that the type believes. Since Harsanyi types induce a belief at all orders, they are said to have an *infinite depth (of reasoning)*.

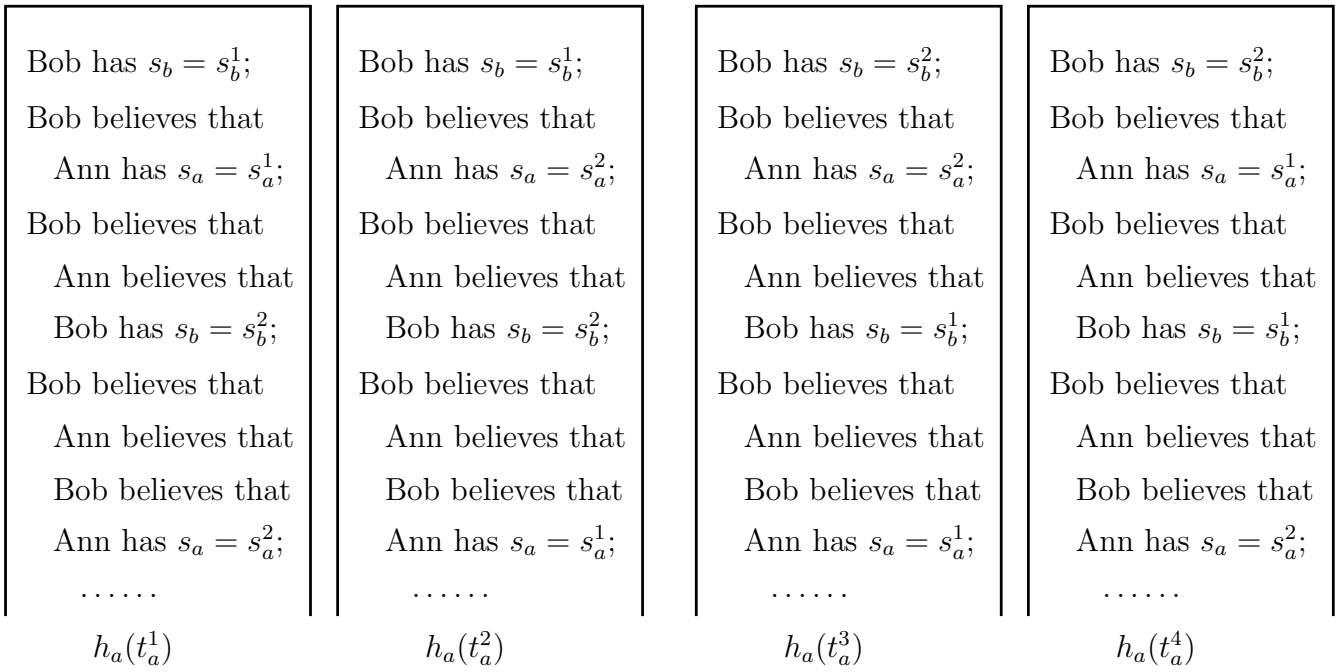

Figure 1: The infinite belief hierarchies generated by the types $t_a^1, t_a^2, t_a^3, t_a^4$.

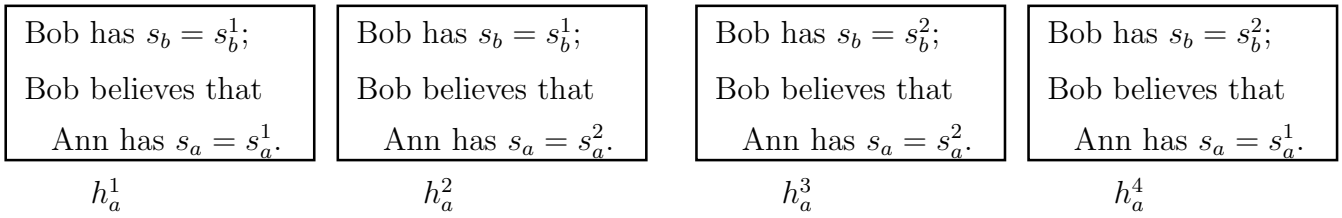

Figure 2: The finite belief hierarchies for Ann.

2.2 Finite belief hierarchies

In practice, players may not reason about the other player's beliefs beyond a certain order. For example, Ann may think about Bob's attribute and about what he thinks, but not about what he thinks that she thinks, or indeed about his beliefs at higher orders. That is, she may have a *finite depth of reasoning*. To give an example, suppose that the first- and a second-order beliefs of both players are as before, but that players do not form a belief at higher orders. This gives four belief hierarchies $h_a^1, h_a^2, h_a^3, h_a^4$ for Ann, as illustrated in Figure 2. Likewise, Bob has four belief hierarchies $h_b^1, h_b^2, h_b^3, h_b^4$ whose beliefs are again symmetric to those for Ann. Since players form a first-order belief and a second-order belief, but no belief at higher order, we say that the players' depth of reasoning equals 2 in this example.

2.3 Types with a finite depth of reasoning

While the beliefs of players with a finite depth can be modeled by writing down their (finite) belief hierarchies explicitly, as we have done in Section 2.2, it is often more convenient to use a type-based framework. To extend the Harsanyi approach to environments where players can have an arbitrary depth of reasoning, it will be instructive to represent the uncertainty that players face by a state space. Recall that a *state (of the world)* for a player is an exhaustive description of the uncertainty that the player faces. A collection of states of the world is a *subjective state space*. In a game-theoretic setting, the state of the world describes the primitive uncertainty (i.e., the attributes of other players) and the beliefs of other players.

This suggests how we can extend the type space formalism so that it can model with an arbitrary depth: Given sets S_a, S_b of attributes for the players, a *type space* specifies a set T_i of types for each player $i = a, b$, and each type $t_i \in T_i$ is associated with a subjective state space Ω_{t_i} and a (subjective) belief π_{t_i} defined on its subjective state space. A Harsanyi type space, like the one we considered in Section 2.1, is a type space where the subjective state space takes a particular form: every state of the world includes the entire belief hierarchy of the other player. As this information is summarized by the other player's type, the subjective state space Ω_{t_i} for a Harsanyi type t_i is $\Omega_{t_i} = \Omega_i^{(\infty)}$, where

$$\Omega_i^{(\infty)} := \left\{(s_{-i}, t_{-i}) : s_{-i} = s_{-i}^1, s_{-i}^2, t_{-i} = t_{-i}^1, t_{-i}^2, t_{-i}^3, t_{-i}^4\right\},$$

where $-i \neq i$ is the player other than i, as is standard. As we have seen, every Harsanyi type t_i unwinds into an infinite hierarchy $h_i(t_i)$ of beliefs; cf. Figure 1.

If players form beliefs only up to some finite order, then they use a coarser description of the uncertainty. To give an example, suppose we want to model a situation where players' depth of reasoning equals 2. That is, Ann forms a belief about Bob's attribute and about Bob's belief about her attribute, but not about Bob's beliefs at higher order, and likewise for Bob. For ease of comparison, assume that beliefs are otherwise the same as before. With some abuse of notation, the types for Ann and Bob are again denoted by $t_a^1, t_a^2, t_a^3, t_a^4$, and $t_b^1, t_b^2, t_b^3, t_b^4$, respectively. As before, type t_a^1 believes that Bob has $s_b = s_b^1$ and that he believes that Ann has $s_a = s_a^1$. That is, t_a^1 assigns probability 1 to the event that Bob has $s_b = s_b^1$ and that he has type t_b^1 or t_b^2 (which both assign probability 1 to s_a^1). However, t_a^1 does not have the language to distinguish the types more finely: t_b^1 and t_b^2 can be distinguished only by their beliefs at order at least 2, and since type t_a^1 has depth 2, it cannot reason about these beliefs. Thus, t_a^1 assigns probability 1 to the event $\{t_b^1, t_b^2\}$ that Bob has type t_b^1 *or* t_b^2, but it cannot assign a probability to any nonempty subset of $\{t_b^1, t_b^2\}$. Likewise, a depth-2 type for Ann can reason about the event that Bob has type t_b^3 *or* t_b^4 (i.e., it can assign a probability to $\{t_b^3, t_b^4\}$), but it cannot reason about any nonempty subset of this event.

Hence, a state of the world for a depth-2 type for player $i = a, b$ is a pair $(s_{-i}, \{t_{-i}, t'_{-i}\})$,[3] where $s_{-i} = s^1_{-i}, s^2_{-i}$ and $\{t_{-i}, t'_{-i}\} = \{t^1_{-i}, t^2_{-i}\}, \{t^3_{-i}, t^4_{-i}\}$, the subjective state space for a depth-2 type in this example is $\Omega_{t_i} = \Omega^{(2)}_i$, where

$$\Omega^{(2)}_i := \left\{(s_{-i}, \{t_{-i}, t'_{-i}\}) : s_{-i} = s^1_{-i}, s^2_{-i}, \{t_{-i}, t'_{-i}\} = \{t^1_{-i}, t^2_{-i}\}, \{t^3_{-i}, t^4_{-i}\}\right\}.$$

As we have seen, a type with subjective state space $\Omega^{(2)}_i$ generates a well-defined first- and second-order belief. For example, type t^1_a for Ann believes that Bob has $s_b = s^1_b$ and that Bob believes Ann has $s_a = s^1_a$ (as $\pi_{t^1_a}(s^1_b, \{t^1_b, t^2_b\}) = 1$, and t^1_b, t^2_b believe that Ann has $s_a = s^1_a$). However, since t^1_b and t^2_b differ in their second-order beliefs, with t^1_b believing that Ann believes that Bob has $s_b = s^1_b$, and t^2_b believing that Ann believes that Bob has $s_b = s^2_b$, type t^1_a does not generate a well-defined third-order belief, and its depth of reasoning equals 2; cf. Figure 2. The same is true for the other types. So, in spite of the fact that type spaces inherently involve a circularity, with each type being associated with a belief about types, the unwinding of a type into a belief hierarchy can be "blocked" at a finite order.

In sum, a player with an infinite depth of reasoning perfectly distinguishes the types of the other player. By contrast, a player with a finite depth of reasoning ignores the distinction between types that differ only in the beliefs these types generate at high order. For example, a type for Ann with depth 2 does not distinguish types for Bob that differ only in their second-order belief (e.g., t^1_b and t^2_b are lumped together into a single set $\{t^1_b, t^2_b\}$). This is precisely the small-world idea of Savage (1954). In Savage's terminology, a small world and a large world are (subjective) state spaces, where a state (of the world) in a small world describes the possible uncertainties a decision-maker faces in less detail than a state in a larger world, by neglecting certain distinctions between states. This means that "a state of the smaller world corresponds not to one state of the larger, but to a *set* of states" (Savage, 1954, p. 9, emphasis added). In the present framework, a player with a lower depth of reasoning makes fewer distinctions between states than a player with a higher depth of reasoning, and thus has a smaller world.

2.4 Choosing the subjective state spaces

Compared to the single-person decision problems considered by Savage (1954), there is an additional layer of complexity here: the states for Bob refer to Ann's beliefs and vice versa. This implies that the subjective state spaces cannot be chosen arbitrarily. Suppose that we want to model a type space in which all types have finite depth k. Then, the subjective state

[3] The standard notation for $(s_i, \{t_i, t'_i\})$ is of course $\{s_i\} \times \{t_i, t'_i\}$. I instead write $(s_i, \{t_i, t'_i\})$ to emphasize that the non-singleton set $\{t_i, t'_i\}$ plays the same role for a finite-depth type as a singleton set for an infinite-depth type.

space for a type for Bob must distinguish the types for Ann if they differ in their mth-order beliefs for some $m \leq k-1$, and lump them together otherwise. But, the mth-order belief generated by a type for Bob depends on the type's subjective state space, which is in turn defined in terms of the higher-order beliefs generated by Ann's types.

So, while we can block the unwinding of a type despite the circularity inherent in the definition of a type space, we seem to encounter another circularity when identifying the appropriate subjective state spaces: the subjective state space for a type for Ann makes reference to the higher-order beliefs of Bob's types, which in turn depend on the subjective state space associated with Bob's types.

However, this circularity is only apparent. It turns out that the subjective state spaces can be ranked in order of their expressivity, and a subjective state space for a finite-depth type can be defined in terms of a subjective state space of a strictly lower rank. As a result, no subjective state space for a finite-depth type is defined in terms of itself, even indirectly, and there is no circularity.

To illustrate, refer back to the example in Section 2.3. By definition, the subjective state space $\Omega_a^{(2)}$ for a depth-2 type for Ann separates the types for Bob if and only if they differ in their first-order belief. That is, $\Omega_a^{(2)}$ lumps together t_b^1 and t_b^2 (which both believe that Ann has $s_a = s_a^1$) as well as t_b^3 and t_b^4 (which both believe that Ann has $s_a = s_a^1$) but distinguishes between $\{t_b^1, t_b^2\}$ and $\{t_b^3, t_b^4\}$. But this is equivalent to saying that $\Omega_a^{(2)}$ separates the types t_b, t_b' for Bob[4] if and only if they differ in their belief $\pi_{t_b}|_{\Omega_b^{(1)}}, \pi_{t_b'}|_{\Omega_b^{(1)}}$ *restricted* to the subjective state space $\Omega_b^{(1)}$ that does not separate the types for Ann, that is,

$$\Omega_b^{(1)} := \left\{(s_a, \{t_a^1, t_a^2, t_a^3, t_a^4\}) : s_a = s_a^1, s_a^2\right\}.$$

For example, types t_b^1 and t_b^2 believe that Ann has $s_a = s_a^1$ and that her type is t_a^1, t_a^2, t_a^3, or t_a^4 (i.e., they assign probability 1 to $(s_a^1, \{t_a^1, t_a^2, t_a^3, t_a^4\}) \in \Omega_b^{(1)}$), and thus have the same belief on $\Omega_b^{(1)}$ (i.e., $\pi_{t_b^1}|_{\Omega_b^{(1)}} = \pi_{t_b^2}|_{\Omega_b^{(1)}}$). Indeed, $\Omega_a^{(2)}$ does not separate t_b^1 and t_b^2. In contrast, t_b^1 and t_b^3 assign probability 1 to $(s_a^1, \{t_a^1, t_a^2, t_a^3, t_a^4\}) \in \Omega_b^{(1)}$ and to $(s_a^2, \{t_a^1, t_a^2, t_a^3, t_a^4\}) \in \Omega_b^{(1)}$, respectively. So, t_b^1 and t_b^3 have different beliefs on $\Omega_b^{(1)}$ (i.e., $\pi_{t_b^1}|_{\Omega_b^{(1)}} \neq \pi_{t_b^3}|_{\Omega_b^{(1)}}$), and they are separated by $\Omega_a^{(2)}$.

This principle applies more generally: for any finite depth k, the subjective state space for a type of depth k (which separates types according to their $(k-1)$th-order beliefs) is precisely the subjective state space that separates the types according to their beliefs on some subjective state space. Moreover, these subjective state spaces can be defined recursively. Suppose we

[4]Given a set X, a collection $\mathcal{X}$ of subsets of X *separates* $x, x' \in X$ if there is $B \in \mathcal{X}$ such that $x \in B$ and $x' \notin B$ or vice versa.

have identified a subjective state space that separates the types for Bob according to their first-order beliefs (such as $\Omega_a^{(2)}$). Then, any two types t_a, t'_a for Ann that differ in their belief restricted to this subjective state space (i.e., $\pi_{t_a}|_{\Omega_a^{(2)}} \neq \pi_{t'_a}|_{\Omega_a^{(2)}}$) induce different second-order beliefs. Now, let $\Omega_b^{(3)}$ be the subjective state space that separates the types for Ann if and only if they differ on $\Omega_a^{(2)}$ (i.e., $\Omega_b^{(3)}$ separates t_a and t'_a if and only if $\pi_{t_a}|_{\Omega_a^{(2)}} \neq \pi_{t'_a}|_{\Omega_a^{(2)}}$). Then, by construction, $\Omega_b^{(3)}$ separates the types for Ann according to their second-order belief. Hence, if a type for Bob has depth at least 3, then it can assign a probability to any of the states in $\Omega_b^{(3)}$.

To illustrate, consider the following example. Each player $i = a, b$ has 8 types, labeled $t_i^1, t_i^2, t_i^3, t_i^4, t_i^5, t_i^6, t_i^7, t_i^8$. Suppose that the subjective state space of each type t_i is

$$\Omega_{t_i} := \left\{(s_{-i}, \{t_{-i}, t'_{-i}\}) : s_{-i} = s_{-i}^1, s_{-i}^2, \{t_{-i}, t'_{-i}\} = \{t_{-i}^1, t_{-i}^2\}, \{t_{-i}^3, t_{-i}^4\}, \{t_{-i}^5, t_{-i}^6\}, \{t_{-i}^7, t_{-i}^8\}\right\},$$

and that beliefs are as follows:

$$\begin{array}{llll} \pi_{t_i^1}(s_{-i}^1, \{t_{-i}^1, t_{-i}^2\}) & = 1; & \pi_{t_i^5}(s_{-i}^2, \{t_{-i}^1, t_{-i}^2\}) & = 1; \\ \pi_{t_i^2}(s_{-i}^1, \{t_{-i}^3, t_{-i}^4\}) & = 1; & \pi_{t_i^6}(s_{-i}^2, \{t_{-i}^3, t_{-i}^4\}) & = 1; \\ \pi_{t_i^3}(s_{-i}^1, \{t_{-i}^5, t_{-i}^6\}) & = 1; & \pi_{t_i^7}(s_{-i}^2, \{t_{-i}^5, t_{-i}^6\}) & = 1; \\ \pi_{t_i^4}(s_{-i}^1, \{t_{-i}^7, t_{-i}^8\}) & = 1; & \pi_{t_i^8}(s_{-i}^2, \{t_{-i}^7, t_{-i}^8\}) & = 1. \end{array}$$

Then, type t_a^1 for Ann believes that Bob has $s_b = s_b^1$, that Bob believes that she has $s_a = s_a^1$ (as t_b^1, t_b^2 assign probability 1 to s_a^1), and that Bob believes that she believes that he has $s_b = s_b^1$ (as t_b^1, t_b^2 assign probability 1 to t_a^1, t_a^2, which assign probability 1 to s_b^1). However, since t_b^1 and t_b^2 differ in their third-order belief, t_a^1 does not generate a well-defined fourth-order belief, and the type has depth 3. The same is true for the other types. Now, suppose that for $i = a, b$, we define $\Omega_i^{(1)} := \left\{(s_{-i}, \{t_{-i}^1, t_{-i}^2, t_{-i}^3, t_{-i}^4, t_{-i}^5, t_{-i}^6, t_{-i}^7, t_{-i}^8\}) : s_{-i} = s_{-i}^1, s_{-i}^2\right\}$ to be the subjective state space that does not distinguish any of the types, as before. Then,

$$\Omega_i^{(2)} := \left\{(s_{-i}, Q_{-i}) : s_{-i} = s_{-i}^1, s_{-i}^2, Q_{-i} = \{t_{-i}^1, t_{-i}^2, t_{-i}^3, t_{-i}^4\}, \{t_{-i}^5, t_{-i}^6, t_{-i}^7, t_{-i}^8\}\right\}$$

is the subjective state space that separates the types if and only if they differ in their belief restricted to $\Omega_{-i}^{(1)}$: types $t_{-i}^1, t_{-i}^2, t_{-i}^3, t_{-i}^4$ believe that player i has $s_i = s_i^1$ (and has type $t_i^1, t_i^2, t_i^3, t_i^4, t_i^5, t_i^6, t_i^7$, or t_i^8) while types $t_{-i}^5, t_{-i}^6, t_{-i}^7, t_{-i}^8$ believe that i has $s_i = s_i^2$ (and has type $t_i^1, t_i^2, t_i^3, t_i^4, t_i^5, t_i^6, t_i^7$, or t_i^8). As before, $\Omega_i^{(2)}$ is precisely the subjective state space that separates the types that differ in their first-order beliefs. Then,

$$\Omega_i^{(3)} := \left\{(s_{-i}, \{t_{-i}, t'_{-i}) : s_{-i} = s_{-i}^1, s_{-i}^2, \{t_{-i}, t'_{-i}\} = \{t_{-i}^1, t_{-i}^2\}, \{t_{-i}^3, t_{-i}^4\}, \{t_{-i}^5, t_{-i}^6\}, \{t_{-i}^7, t_{-i}^8\}\right\}$$

is the subjective state space that separates the types if and only if they differ in their belief restricted to $\Omega_{-i}^{(2)}$. For example, types t_a^1 and t_a^2 believe that Bob has $s_b = s_b^1$ and that he has

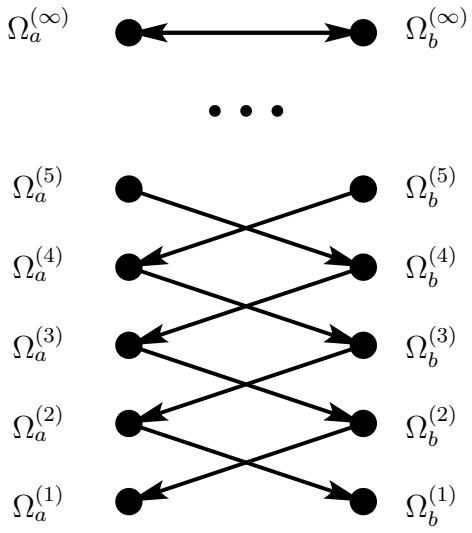

Figure 3: Subjective state spaces. An arrow from Ω_i to Ω_{-i} means that a type with subjective state space Ω_i separates the types according to their belief restricted to Ω_{-i}.

type t_b^1, t_b^2, t_b^3, or t_b^4 (i.e., t_a^1 and t_a^2 assign probability 1 to $(s_b^1, \{t_b^1, t_b^2, t_b^3, t_b^4\}) \in \Omega_a^{(2)}$) while types t_a^3 and t_a^4 believe that Bob has $s_b = s_b^1$ and that he has type t_b^5, t_b^6, t_b^7, or t_b^8 (i.e., t_a^3 and t_a^4 assign probability 1 to $(s_b^1, \{t_b^5, t_b^6, t_b^7, t_b^8\}) \in \Omega_a^{(2)}$). But, $\Omega_b^{(3)}$ is just the subjective state space Ω_{t_b} for Bob's types, which have depth 3. Hence, the subjective state space $\Omega_i^{(3)}$ that separates the types according to their second-order belief is precisely the subjective state space that separates the types according to their beliefs on the subjective state space $\Omega_{-i}^{(2)}$; and $\Omega_{-i}^{(2)}$ is precisely the subjective state space that separates the types according to their beliefs on the "trivial" subjective state space $\Omega_i^{(1)}$ that does not separate any of the types.

This suggests that if a type t_a has a finite depth of reasoning, then its subjective state space Ω_{t_a} is either trivial (i.e., does not separate any of the types) or there is a finite sequence $\Omega_a^{(k)}, \Omega_b^{(k-1)}, \ldots, \Omega_j^{(1)}$ of subjective state spaces such that $\Omega_{t_a} = \Omega_a^{(k)}$, $\Omega_j^{(1)}$ is the trivial subjective state space ($j = a, b$) and for all $m > 1$ and $i = a, b$, $\Omega_i^{(m)}$ separates the types according to their beliefs restricted to $\Omega_{-i}^{(m-1)}$. By definition, if a subjective state space Ω_{t_i} associated with a finite-depth type for player i separates types according to their beliefs on a subjective state space Ω, then Ω does not separate types according to their beliefs on Ω_{t_i}. Hence, there is no circularity for finite-depth types: the subjective state spaces associated with finite-depth types can indeed be ranked. This is illustrated in Figure 3, where the subjective state spaces for each player are ordered vertically.

What about types with an infinite depth of reasoning? Refer back to the Harsanyi type space in Section 2.1. In this case, the subjective state space $\Omega_i^{(\infty)}$ perfectly distinguishes between the types. Consequently, the subjective state space $\Omega_a^{(\infty)}$ separates the types according

to their belief restricted to $\Omega_b^{(\infty)}$. The converse also holds. Thus, for $i = a, b$, $\Omega_i^{(\infty)}$ separates the types t_{-i}, t'_{-i} whenever $\pi_{t_{-i}}|_{\Omega_{-i}^{(\infty)}} \neq \pi_{t'_{-i}}|_{\Omega_{-i}^{(\infty)}}$. This does involve a circularity, but it is a circularity of a particular kind: it is precisely the standard condition on Harsanyi type spaces that the function that maps each type t_i into a belief π_{t_i} be measurable (Proposition 5.1).

Condition (SEP) in Section 5 summarizes the conditions on subjective state spaces discussed above in recursive form. Then:

Proposition 6.3 *Assume condition (SEP). Then, every type generates a belief hierarchy of a well-defined depth.*

Thus, by choosing the set of events types can reason about (i.e., the subjective state spaces), it is possible to define types that can reason about only finitely many orders of beliefs.[5]

2.5 Types and belief hierarchies are equally expressive

For game-theoretic analyses, belief hierarchies are the primary objects of interest, and type spaces are just devices to model these hierarchies. Hence, a natural question is whether the type space approach and the belief hierarchy approach are equally expressive, in the sense that every belief hierarchy can be modeled by a type and vice versa. For Harsanyi type spaces, Mertens and Zamir (1985) provide an affirmative answer. Also in the present context, modeling belief hierarchies with types is without loss of generality:

Theorem 6.1 *Types and belief hierarchies are equally expressive: (1) every belief hierarchy defines a type; and (2) every type generates a belief hierarchy.*

To prove this result, I use a somewhat different proof method than has been used in the literature. The standard proof for the Harsanyi case proceeds by constructing the set of all infinite belief hierarchies and showing that this set of belief hierarchies defines a Harsanyi type space (e.g., Mertens and Zamir, 1985; Brandenburger and Dekel, 1993; Battigalli and Siniscalchi, 1999). The existence of such a "canonical" Harsanyi type space $\mathcal{T}^h$ immediately implies that belief hierarchies and types are equally expressive in the Harsanyi case: since the set of types in the canonical Harsanyi type space consists of all infinite belief hierarchies, the type space also generates all such hierarchies. But, this proof methods proves a stronger result than is necessary: for types and belief hierarchies to be equally expressive, it is not necessary that there is a *single* type space that generates all belief hierarchies. Instead, it suffices to

[5]This fits in with a literature that studies how the measurable structure associated with Harsanyi type spaces can implicitly impose restrictions on reasoning, that is, on belief hierarchies (e.g., Brandenburger and Keisler, 2006; Friedenberg and Meier, 2012; Perea and Kets, 2016).

define a class of spaces of belief hierarchies (rather than only the "canonical" space of belief hierarchies) and to show that every type space corresponds to a space of belief hierarchies and vice versa. The key to proving that every type generates a belief hierarchy is to show that the types' subjective state spaces distinguish the types for the other player based on their higher-order beliefs. While the formal result (Proposition 6.3) requires considerable care, this is essentially guaranteed by condition (SEP), as we have seen. The proof that every belief hierarchy defines a type uses a standard extension argument (Proposition 6.10).

2.6 A universal type space?

In the Harsanyi case, type spaces can be "nested" in the sense that a type space that imposes strong restrictions on players' beliefs can be contained in a type spaces that relaxes these assumptions. For example, a type space that models the assumption that bidders in an auction all have private values for the object on sale can be embedded in a type space where all players believe it is likely that players have private values, believe that others believe this is likely, and so on.

Can type spaces still be nested in this way if players have an arbitrary depth of reasoning? To illustrate the issues, it will be helpful to consider an example. The primitive uncertainty is described by a common attribute (i.e., there exist s^1, s^2 such that $s_i^1 = s^1$ and $s_i^2 = s^2$ for $i = a, b$). Ann's type set is $T_a = \{t_a^1, t_a^2\}$ and Bob's type space is $T_b = \{t_b^1, t_b^2\}$. Type t_a^1 believes that $s = s^1$ but does not reason about Bob's beliefs (i.e., $\Omega_{t_a^1} = \Omega_a^{(1)}$, where $\Omega_a^{(1)} := \{(s^1, \{t_b^1, t_b^2\}), (s^1, \{t_b^1, t_b^2\})\}$). Type t_a^2 believes that $s = s^2$ and does not reason about Bob's beliefs (i.e., $\Omega_{t_a^2} = \Omega_a^{(1)}$). Type t_b^1 believes that $s = s_b^1$ and that Ann believes that $s = s_b^1$ (i.e., $\pi_{t_b^1}(s^1, t_a^1) = 1$), and type t_b^2 believes that $s = s_b^2$ and that Ann believes that $s = s^2$ (i.e., $\pi_{t_b^2}(s^2, t_a^2) = 1$). For future reference, denote this type space by $\mathcal{T}$. This type space models a situation where Bob believes that Ann has correct beliefs about the common attribute. In particular, while Ann does not reason about Bob's belief explicitly, she does rule out that Bob thinks that she is has incorrect beliefs: in any of the states that she considers, Bob thinks she has correct beliefs. Such belief restrictions can arise naturally, for example, if Ann observes the attribute and that Bob observes that she observes the attribute (but Bob does not observe the attribute himself).

An analyst interested in analyzing this situation might want to allow for the possibility that Bob thinks it possible that Ann's beliefs are not correct (perhaps because Ann may be inattentive). He thus considers the following type space: Ann's type set is $\widetilde{T}_a = \{\tilde{t}_a^1, \tilde{t}_a^2\}$ and Bob's type space is $\widetilde{T}_b = \{0,1\} \times \{0,1\}$. Type $\tilde{t}_a^1$ assigns probability 1 to $(s^1, \widetilde{T}_b)$, and $\tilde{t}_a^2$ assigns probability 1 to $(s^2, \widetilde{T}_b)$. The belief for type $\tilde{t}_b = (x, y)$ is defined as follows: if $y = 0$, $\tilde{t}_b$ believes

that Ann has correct beliefs: it assigns probability x to $(s^1, \tilde{t}_a^1)$ and probability $1-x$ to $(s^2, \tilde{t}_a^2)$; if $y = 0$, then $\tilde{t}_b$ believes that Ann's beliefs are not correct: it assigns probability x to $(s^1, \tilde{t}_a^2)$ and probability $1-x$ to $(s^2, \tilde{t}_a^1)$. Denote this type space by $\widetilde{\mathcal{T}}$. This type space embeds the type space $\mathcal{T}$ if we define a pair $\varphi = (\varphi_a, \varphi_b)$ of mappings from $\mathcal{T}$ to $\widetilde{\mathcal{T}}$ by:

$$\begin{aligned} \varphi_a(t_a^1) &:= \tilde{t}_a^1; \quad \varphi_b(t_b^1) := (1,0); \\ \varphi_a(t_a^2) &:= \tilde{t}_a^2; \quad \varphi_b(t_b^2) := (0,0). \end{aligned}$$

Then, φ preserve the types' depth of reasoning and the beliefs at the order that the type can reason about. For example, type t_a^1 and $\varphi_a(t_a^1)$ believe that $s = s^1$ and do not reason about Bob's beliefs (i.e., they have depth 1). Likewise, t_b^1 and $\varphi_b(t_b^1)$ believe that $s = s^1$, that Ann believes that $s = s^1$, and that Ann does not reason about Bob's beliefs. That is, $\varphi := (\varphi_a, \varphi_b)$ is a *type morphism* from $\mathcal{T}$ to $\widetilde{\mathcal{T}}$. The type space $\widetilde{\mathcal{T}}$ includes the images $\varphi_i(t_i)$ of the types t_i in $\mathcal{T}$ but it also includes other types. In particular, it allows for the possibility that Bob believes that Ann has incorrect beliefs. In this sense, the type space $\widetilde{\mathcal{T}}$ relaxes the assumptions on higher-order beliefs implicit in the specification of $\mathcal{T}$.

But, while $\widetilde{\mathcal{T}}$ embeds $\mathcal{T}$ in the sense described above, there is an important sense in which the types t_i in $\mathcal{T}$ and their images $\varphi_i(t_i)$ in $\widetilde{\mathcal{T}}$ have different higher-order beliefs. To wit, the type space $\mathcal{T}$ models a situation where Ann rules out that Bob thinks she has incorrect beliefs: any type in T_a believes that Bob has a type in T_b, and all types in T_b believe that Ann has correct beliefs about the common attribute. But in $\widetilde{\mathcal{T}}$, Ann does not rule out that Bob thinks she has incorrect beliefs: for $t_a \in T_a$, type $\tilde{t}_a = \varphi_a(t_a)$ cannot rule out any of the types in $\widetilde{T}_b$, some of which believe that Ann has incorrect beliefs, because its subjective state space is too coarse to assign zero probability to the types that think that Ann has incorrect beliefs. Thus, the image $\varphi_a(t_a)$ of type t_a thinks possible beliefs that the type t_a rules out, and we say that $\mathcal{T}$ does not form a *belief-closed subset* of $\widetilde{\mathcal{T}}$. In this case, type $t_a \in T_a$ and its image $\varphi_a(t_a)$ under φ generate different belief hierarchies. Intuitively, type morphisms preserve players' beliefs at the orders they can reason about, but fail to preserve beliefs at higher orders because types' subjective state spaces are too coarse to rule out certain beliefs. Hence, $\mathcal{T}$ cannot be embedded in a way that preserves higher-order beliefs.

So, is there type space that embeds all type spaces in a way that preserves higher-order beliefs? Say that a type space $\widetilde{\mathcal{T}}$ is *universal* for a class of type spaces if every type space in the class can be embedded into $\widetilde{\mathcal{T}}$ as a belief-closed subset via a type morphism. Then, we have the following negative result:

Theorem 7.5 *There is no universal type space for the class of all type spaces.*

This result contrasts with the well-known positive result for the Harsanyi case that the canonical Harsanyi type space $\mathcal{T}^h$ is universal for the class of Harsanyi type spaces (Mertens

and Zamir, 1985, Thm. 2.9.(5)). The difference between the Harsanyi case and the general case lies in the feature of type morphisms that they preserve the beliefs of types at orders the types can reason about, but not at higher orders.[6] Since Harsanyi types induce a well-defined kth-order belief at every order k, a type morphism from a Harsanyi type space to a Harsanyi type space preserves beliefs at all orders. Intuitively, the subjective state space associated with a Harsanyi type describes belief hierarchies in full detail, and a Harsanyi type can assign a probability to any event that refers to the other player's higher-order belief. As a result, if a Harsanyi type t_i^H rules out certain events, then its image $\varphi_i(t_i^H)$ under a type morphism φ can rule out these events by assigning zero probability to them. By contrast, if types have a finite depth of reasoning, then type morphisms fail to preserve beliefs at higher orders. As the subjective type space associated with a finite-depth type describes belief hierarchies in limited detail, a finite-depth type can assign a probability only to events that can be expressed in terms of the other player's belief up to some finite order. As a result, the image $\varphi_i(t_i)$ of a finite-depth type t_i under a type morphism $\varphi = (\varphi_a, \varphi_b)$ cannot rule out certain events by assigning zero probability to them. In the example above, type $\varphi_a(t_a)$ cannot rule out that Bob thinks Ann has incorrect beliefs by assigning zero probability to the types that think she has incorrect beliefs because its subjective state space is too coarse to distinguish these types from the types that think she has correct beliefs.

Together, Theorems 6.1 and 7.5 show that two properties that go hand in hand in the Harsanyi case – the equivalence between types and belief hierarchies and the existence of a universal type space – do not concur when players are bounded in their reasoning. Indeed, Theorem 7.5 motivates my alternative approach to proving Theorem 6.1 described in Section 2.5: if there is no universal type space, no type space can generate all belief hierarchies, and it is necessary to consider the class of all spaces of belief hierarchies.

One might think that the requirement that a type space embed all type spaces as a belief-closed subset is too strong: perhaps an analyst who is interested in studying strategic behavior across all strategic situations can use a type space $\widetilde{\mathcal{T}}$ that has the property that there is a type morphism from any type space $\mathcal{T}$ to $\widetilde{\mathcal{T}}$ (even if the image of $\mathcal{T}$ is not belief-closed subspace of $\widetilde{\mathcal{T}}$). In Section 7, I show by example that this is not the case, at least for the commonly-used concept of rationalizability: if $\mathcal{T}$ can be embedded into $\widetilde{\mathcal{T}}$ using a type morphism φ, but the image of $\mathcal{T}$ does not form a belief-closed subset of $\widetilde{\mathcal{T}}$, then there may be a game and a type t_i in $\mathcal{T}$ such that t_i and $\varphi_i(t_i)$ have different rationalizable actions (Example 3). Intuitively, if a type morphism does not preserve higher-order beliefs (i.e., the image of $\mathcal{T}$ is not a belief-closed subset of $\widetilde{\mathcal{T}}$), then the embedded type space captures a different state of affairs than

[6] In particular, the negative result (Theorem 7.5) is not due to the nonexistence of a canonical type space for the general case: such a canonical type space exists (Corollary 6.16).

the original type space, and we have no reason to expect similar behavior in these distinct situations. Thus, if an analyst is interested in studying the rationalizable behavior of players with a finite depth of reasoning across all strategic environments, then he needs to consider all type spaces.

2.7 Outline

The remainder of this paper is organized as follows. After a brief discussion of preliminaries in Section 3, Sections 4 and 5 introduce belief hierarchies and types, respectively. Section 6 shows that these two approaches are equally expressive. Section 7 shows that there is no universal type space for the class of all type spaces. All proofs not included in the main text can be found in the appendices.

3 Preliminaries

To define beliefs over infinite sets, subjective state spaces are replaced by σ-algebras, and probability measures take the place of probability distributions. This section introduces basic concepts and notation.

A *measurable space* is a pair $(X, \mathcal{F})$ where X is an arbitrary set and $\mathcal{F}$ is a σ-algebra on X. An element E of $\mathcal{F}$ is an *event*. Given a measurable space $(X, \mathcal{F})$, a (subjective) *belief* about X is a probability measure μ defined on $(X, \mathcal{F})$. So, $\mu(E) \in [0, 1]$ is the probability that the belief μ assigns to the event E. The σ-algebra on which μ is defined is denoted by $\Sigma(\mu)$. The set of beliefs on $(X, \mathcal{F})$ is denoted by $\Delta(X, \mathcal{F})$, or $\Delta(X)$ if no confusion can result. The set $\Delta(X, \mathcal{F})$ is endowed with the coarsest σ-algebra that contains the sets

$$\big\{\mu \in \Delta(X, \mathcal{F}) : \mu(E) \geq p\big\} : \qquad E \in \mathcal{F}, p \in [0, 1].$$

This σ-algebra, denoted by $\mathcal{F}_{\Delta(X,\mathcal{F})}$, separates beliefs (i.e., probability measures) according to the probabilities they assign to events in $\mathcal{F}$. If X is metrizable, $\mathcal{F}$ is the Borel σ-algebra on X, and $\Delta(X, \mathcal{F})$ is endowed with the weak topology, then $\mathcal{F}_{\Delta(X,\mathcal{F})}$ coincides with the Borel σ-algebra (Heifetz and Samet, 1998).

Given measurable spaces $(X, \mathcal{F}_X)$ and $(Y, \mathcal{F}_Y)$, a function $f : X \to Y$ is *measurable* (with respect to $\mathcal{F}_X$ and $\mathcal{F}_Y$), or $(\mathcal{F}_X, \mathcal{F}_Y)$-*measurable*, if $f^{-1}(E) \in \mathcal{F}_X$ for every event $E \in \mathcal{F}_Y$. The function f is an *isomorphism* (with respect to $\mathcal{F}_X$ and $\mathcal{F}_Y$) if its inverse exists and is measurable. If $f : X \to Y$ is an $(\mathcal{F}_X, \mathcal{F}_Y)$-measurable function, and $\mu \in \Delta(X, \mathcal{F}_X)$ is a belief about X, then the *image measure* associated with μ (induced by f) is the belief $\mu \circ f^{-1}$ about

Y, where

$$\mu \circ f^{-1}(E) = \mu(\{x \in X : f(x) \in E\})$$

for $E \in \mathcal{F}_Y$. Since f is measurable, the image measure $\mu \circ f^{-1}$ is well-defined.

If $(X, \mathcal{F}_X)$ is a measurable space, then any subset $Y \subset X$ has the *relative σ-algebra* (induced by $\mathcal{F}_X$). For any family of measurable spaces $(X_z, \mathcal{F}_z)$, $z \in Z$, the product $\prod_{z \in Z} X_z$ is endowed with the *product σ-algebra* $\bigotimes_{z \in Z} \mathcal{F}_z$. The union $\bigcup_{z \in Z} X_z$ is endowed with the *sum σ-algebra*, that is, the σ-algebra that contains precisely the subsets $E \subseteq \bigcup_{z \in Z} X_z$ such that $E \cap X_z \in \mathcal{F}_z$ for all $z \in Z$ (Kechris, 1995, p. 67).

A measurable space $(X, \mathcal{F})$ is a *standard Borel space* if there is a Polish topology on X such that its Borel σ-algebra coincides with $\mathcal{F}$ (e.g., Kechris, 1995, p. 74). A measurable space $(X, \mathcal{F})$ is an *analytic Borel space* if it is isomorphic to $(A, \mathscr{B}(A))$ for an analytic set A (i.e., a subset of a Polish space that is the continuous image of a Polish space) and its Borel σ-algebra $\mathscr{B}(A)$ (Kechris, 1995, p. 197). A measurable space that is standard Borel is analytic, but the converse need not hold. Many commonly encountered spaces are standard Borel and thus analytic. For example, the measurable spaces associated with Polish or compact metric spaces are standard Borel spaces. The additional generality afforded by analytic Borel spaces will be useful for proving the equivalence results in Section 6.

4 Belief hierarchies

For simplicity, I focus on the case of two players, labeled $i = a, b$. Player a is uncertain about some features of player b (e.g., his action, or his payoff function). This primitive uncertainty is captured by the set S_b of attributes for b. Likewise, the primitive uncertainty for player b is captured by the set S_a of attributes for a. The sets S_a and S_b are assumed to be finite and are endowed with their natural (discrete) σ-algebras, denoted $\mathcal{F}_{S_a}$ and $\mathcal{F}_{S_b}$, respectively.[7] When player i is fixed, the other player is denoted by $-i$, as is standard. For ease of notation, if $\mathcal{F}$ is a σ-algebra on a set X_i associated with player i, then I write $\overline{\mathcal{F}}$ for the product σ-algebra $\mathcal{F}_{S_i} \otimes \mathcal{F}$.

4.1 Models

Higher-order beliefs can be modeled using belief hierarchies. A belief hierarchy specifies a player's belief over the state of nature (e.g., her opponents' strategies), her beliefs over

[7] The results extend to the case of three or more players with minor modifications. The assumption that S_a and S_b are finite is also not critical.

opponents' beliefs, and so on. A model is a collection of belief hierarchies that is belief-closed in the sense that the beliefs in each hierarchy have support in the model. Formally, an ((S_a, S_b)-based) *model* is a pair $\mathcal{H} = (H_a, H_b)$ such that for each player $i = a, b$,

$$H_i = \{(\mu_i^1, \mu_i^2, \ldots) : (\mu_i^1, \ldots, \mu_i^n) \in H_i^n \text{ for all } n\},$$

is a (nonempty) set of *belief hierarchies*, where $H_i^1 \subset \Delta(S_{-i})$ and the sets H_i^m of *mth-order belief hierarchies*, $m = 2, 3, \ldots$, satisfy the following conditions:

(IND) H_i^m is a nonempty subset of $H_{-i}^{m-1} \times \Delta^+(S_{-i} \times H_{-i}^{m-1})$, where

$$\Delta^+(S_{-i} \times H_{-i}^{m-1}) := \bigcup_{\ell=0}^{m-1} \Delta(S_{-i} \times H_{-i}^{m-1}, \overline{\mathcal{F}}_{-i,\ell}^{m-1}),$$

with $\mathcal{F}_{-i,0}^{m-1} := \{H_{-i}^{m-1}, \emptyset\}$ the trivial σ-algebra, $\mathcal{F}_{-i,m-1}^{m-1}$ the relative σ-algebra induced by the product topology, and

$$\mathcal{F}_{-i,\ell}^{m-1} := \left\{ \left\{ (\mu_{-i}^1, \ldots, \mu_{-i}^\ell, \ldots, \mu_{-i}^{m-1}) \in H_{-i}^{m-1} : (\mu_{-i}^1, \ldots, \mu_{-i}^\ell) \in B_{-i}^\ell \right\} : B_{-i}^\ell \in \mathcal{F}_{-i,\ell}^\ell \right\} \tag{4.1}$$

the σ-algebra on H_{-i}^{m-1} that is generated by the projection function from H_{-i}^{m-1} to H_{-i}^ℓ, $\ell < m-1$;

(COH) for every $(\mu_i^1, \ldots, \mu_i^m) \in H_i^m$, $\text{marg}_{S_{-i} \times H_{-i}^{m-2}} \mu_i^m = \mu_i^{m-1}$ if $m > 2$, and $\text{marg}_{S_{-i}} \mu_i^2 = \mu_i^1$ otherwise;[8]

(EXT) for every $(\mu_i^1, \ldots, \mu_i^{m-1}) \in H_i^{m-1}$, there is μ_i^m such that $(\mu_i^1, \ldots, \mu_i^{m-1}, \mu_i^m) \in H_i^m$;

(ANL_H) the measurable space $(H_i^m, \mathcal{F}_{i,m}^m)$ is an analytic Borel space.

Condition (IND) says that H_i^m is defined inductively, with each mth-order belief hierarchy $(\mu_i^1, \ldots, \mu_i^m)$ consisting of an $(m-1)$th-order belief hierarchy $(\mu_i^1, \ldots, \mu_i^{m-1})$ and an *mth-order belief* μ_i^m. A player's mth-order belief can be defined on different σ-algebras $\mathcal{F}_{-i,0}^{m-1}, \ldots \mathcal{F}_{-i,m-1}^{m-1}$. The σ-algebras differ in how finely they distinguish the opponent's belief hierarchies. As I show in Section 4.2 below, the selection of σ-algebras ensures that every belief hierarchy has a well-defined depth of reasoning. Condition (COH) is a standard coherency condition that says that beliefs at different orders do not contradict each other; see, e.g., Siniscalchi (2008) for a discussion. Condition (EXT) states that every $(m-1)$th-order belief hierarchy is extended to an mth-order belief hierarchy. This is always possible: By Proposition A.2 in the appendix,

[8] Given a belief μ on a product space $X \times Y$, $\text{marg}_X \mu$ is its marginal on X.

any set H_i^{m-1} of $(m-1)$th-order belief hierarchies can be extended to a set H_i^m of mth-order belief hierarchies. Condition (ANL_H) imposes a measurable structure on H_i^m.

Given a set H_i of belief hierarchies, let $\mathcal{F}_{H_i}$ be the coarsest σ-algebra on H_i such that the projection function from H_i into H_i^m is measurable for all m whenever H_i^m is endowed with the σ-algebra $\mathcal{F}_{i,m}^m$. This yields analytic Borel spaces:

Lemma 4.1. The measurable spaces $(H_a, \mathcal{F}_{H_a})$ and $(H_b, \mathcal{F}_{H_b})$ are nonempty analytic Borel spaces.

4.2 Depth of reasoning

Belief hierarchies may differ in the events they can assign a probability to, that is, they can be defined on different σ-algebras (cf. (IND)). A belief hierarchy has a well-defined depth if it can either assign a probability to all events that concern the other player's high-order beliefs (in which case it has an infinite depth), or there is some finite k such that it can assign a probability to all events that concern the opponent's belief up to order $k-1$, but not at any higher order (in which case it has finite depth k).

Formally, fix a belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots)$. The belief hierarchy h_i has *depth (of reasoning) at least* 1 if the σ-algebra on which its high-order beliefs are defined distinguish the events that refer to the primitive uncertainty, i.e., for all $m \geq 1$, the σ-algebra $\Sigma(\mu_i^m)$ on which the mth-order belief μ_i^m is defined satisfies $\Sigma(\mu_i^m) \supset \overline{\mathcal{F}}_{-i,0}^{m-1}$. For $k > 1$, say that h_i has *depth of reasoning at least* k if the σ-algebra on which its high-order beliefs are defined distinguishes events that are expressible in terms of the opponent's $(k-1)$th-order beliefs, that is, for all $\ell \leq k-2$, $E \in \overline{\mathcal{F}}_{i,\ell}^{k-2}$, and $p \in [0,1]$,

$$\{(s_{-i}, \mu_{-i}^1, \ldots, \mu_{-i}^{m-1}) \in S_{-i} \times H_{-i}^{m-1} : \Sigma(\mu_{-i}^{k-1}) = \overline{\mathcal{F}}_{i,\ell}^{k-2}, \mu_{-i}^{k-1}(E) \geq p\} \in \Sigma(\mu_i^m).$$

If this condition holds, then the mth-order belief μ_i^m can assign a probability to every event that can be expressed in terms of the opponent's $(k-1)$th-order belief hierarchies. By definition, if a belief hierarchy has depth at least k, then it has depth at least ℓ for $\ell \leq k$.

Definition 4.2. [Depth of reasoning] A belief hierarchy has an *infinite depth (of reasoning)* if it has depth at least m for $m \geq 1$. A belief hierarchy has *depth (of reasoning)* k for $k = 1, 2, \ldots$ if it has depth of reasoning at least k, but not at least $k+1$. If a belief hierarchy has depth k for $k = 1, 2, \ldots$, then it has a *finite depth (of reasoning)*.

The following result is immediate:

Proposition 4.3. Every belief hierarchy has a well-defined depth of reasoning.

Proof. Fix a player $i = a, b$. It is easy to see that for every m, the σ-algebras in (IND) can be ordered by set inclusion:

$$\mathcal{F}_{-i,0}^{m-1} \subset \mathcal{F}_{-i,1}^{m-1} \subset \cdots \subset \mathcal{F}_{-i,m-1}^{m-1}.$$

Hence, every belief hierarchy has depth at least 1. Fix a belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots)$. Then, by conditions (IND) and (COH), one of the following is the case: (i) $\Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,m-1}^{m-1}$ for every m; or (ii) there is $k < \infty$ such that $\Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,m-1}^{m-1}$ for $m \leq k$ and $\Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,k-1}^{m-1} \subsetneq \overline{\mathcal{F}}_{-i,k}^{m-1}$ for $m > k$ (where $\overline{\mathcal{F}}_{-i,0}^{0} := \mathcal{F}_{S_{-i}}$). In the former case, the belief hierarchy has infinite depth by definition. In the latter case, (4.1) implies that h_i has depth k. □

Proposition 4.3 implies that belief hierarchies with a greater depth of reasoning can assign a probability to more events than belief hierarchies with a lower depth.

4.3 Examples

It will be instructive to consider a few examples. The first example is the model constructed by Mertens and Zamir (1985) that contains the belief hierarchies generated by Harsanyi type spaces. This model is canonical in the sense that it does no impose any a priori restrictions on the beliefs players may have beyond the assumption that all players have an infinite depth of reasoning:

Example 1. [The canonical Harsanyi model (Mertens and Zamir, 1985)] For each player $i = a, b$, let $H_i^{h,1} := \Delta(S_{-i})$. For $m = 2, 3, \ldots$, suppose $H_a^{h,m-1}$ and $H_b^{h,m-1}$ have been defined. Then, let $H_a^{h,m}$ be the set of elements $(\mu_a^1, \ldots, \mu_a^m)$ such that (1) $(\mu_a^1, \ldots, \mu_a^{m-1}) \in H_a^{h,m-1}$; (2) μ_a^m is defined on $\overline{\mathcal{F}}_{b,m-1}^{h,m-1}$; and (3) (COH) is satisfied. By construction, conditions (IND)–(ANL_H) are satisfied. Define $H_b^{h,m}$ analogously, and let $H_i^h := \{(\mu_i^1, \mu_i^2, \ldots) : (\mu_i^1, \ldots, \mu_i^n) \in H_i^{h,n}$ for all $n\}$. All belief hierarchies in $\mathcal{H}^h := (H_a^h, H_b^h)$ have an infinite depth of reasoning. ◁

By definition, all belief hierarchies in the canonical Harsanyi model have an infinite depth of reasoning. So, while the canonical Harsanyi model contains all belief hierarchies generated by Harsanyi type spaces, it does not contain any belief hierarchies with a finite depth of reasoning. The following example uses a construction analogous to that in Example 1 to define a model that includes both the belief hierarchies in $\mathcal{H}^h$ as well as belief hierarchies with a finite depth of reasoning.

Example 2. [The canonical model for arbitrary depth] For each player $i = a, b$, let $H_i^{*,1} := \Delta(S_{-i})$. For $m = 2, 3, \ldots$, suppose $H_i^{*,m-1}$ has been defined for each player i. Then, let $H_a^{*,m}$ be the set of elements $(\mu_a^1, \ldots, \mu_b^m)$ such that (1) $(\mu_a^1, \ldots, \mu_a^{m-1}) \in H_a^{*,m-1}$; (2) μ_a^m is

defined on $\overline{\mathcal{F}}^{*,m-1}_{b,\ell}$ for some $\ell \leq m-1$; and (3) (COH) is satisfied. Define $H^{*,m}_b$ analogously. Again, conditions (IND)–(ANL$_H$) are satisfied. Define H^*_a and H^*_b in the usual way. The model $\mathcal{H}^* := (H^*_a, H^*_b)$ contains belief hierarchies of any (finite or infinite) depth of reasoning. ◁

The difference between the canonical model $\mathcal{H}^*$ and the canonical Harsanyi model $\mathcal{H}^h$ is that in the canonical Harsanyi model, beliefs are defined on the finest σ-algebra, while in the canonical model, they can be defined on coarser σ-algebras. The canonical Harsanyi model $\mathcal{H}^h$ can thus be viewed as a submodel of $\mathcal{H}^*$. In fact, it is not difficult to show that $\mathcal{H}^h$ is the submodel of $\mathcal{H}^*$ that is characterized by the event that players have an infinite depth of reasoning and this is common belief (cf. Heifetz and Kets, 2018).

5 Types

Type spaces provide an alternative way to model higher-order beliefs. An ((S_a, S_b)-based) *type space* is a tuple

$$\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$$

that satisfies conditions (SEP) and (ANL$_T$) below. For each player $i = a, b$, T_i is a set of *types*, and $\mathscr{F}_i$ is a collection of σ-algebras on T_i. The function π_i is the *belief map*: it maps each type $t_i \in T_i$ into a belief $\pi_i(t_i)$ over $S_{-i} \times T_{-i}$, where t_i's belief about the other player's type is defined on one of the σ-algebras in $\mathscr{F}_{-i}$. For simplicity, I write π_{t_i} for $\pi_i(t_i)$. Denote σ-algebra on which t_i's belief over the other player's type is defined by $\Sigma_{t_i} \in \mathscr{F}_{-i}$ (i.e., $\pi_{t_i} \in \Delta(S_{-i} \times T_{-i}, \overline{\Sigma}_{t_i})$). So, in the terminology of Section 2, $\overline{\Sigma}_{t_i}$ represents the subjective state space associated with t_i.

To state conditions (SEP) and (ANL$_T$), say that a σ-algebra $\mathcal{F}_i$ on T_i *separates the types according to their belief* on a σ-algebra $\mathcal{F}_{-i}$ on T_{-i}, denoted $\mathcal{F}_i \succ \mathcal{F}_{-i}$, if

$$\left\{t_i \in T_i : E \in \overline{\Sigma}_{t_i}, \pi_{t_i}(E) \geq p\right\} \in \mathcal{F}_i$$

for every event $E \in \overline{\mathcal{F}}_{-i}$ and $p \in [0,1]$. If $\mathcal{F}_i$ is the coarsest σ-algebra that separates the types according to their belief on $\mathcal{F}_{-i}$, then $\mathcal{F}_i$ *strictly separates* the types according to their belief on $\mathcal{F}_{-i}$; this is denoted by $\mathcal{F}_i \succ^* \mathcal{F}_{-i}$. A pair $(\mathcal{F}_a, \mathcal{F}_b)$ of σ-algebras defined on T_a and T_b, respectively, that is such that $\mathcal{F}_a$ separates the types according to their belief on $\mathcal{F}_b$ and vice versa (i.e., $\mathcal{F}_a \succ \mathcal{F}_b$ and $\mathcal{F}_b \succ \mathcal{F}_a$) are a *mutual-separation pair*.

Condition (SEP) imposes restrictions on the class of σ-algebras. As I show in Section 6.1 below, this condition ensures that each type generates a belief hierarchy with a well-defined depth of reasoning.

(SEP) For each player $i = a, b$ and every nontrivial σ-algebra $\mathcal{F}_i \in \mathscr{F}_i$, there is a σ-algebra $\mathcal{F}_{-i} \in \mathscr{F}_{-i}$ such that one of the following holds:

(a) $\mathcal{F}_a$ and $\mathcal{F}_b$ form a mutual-separation pair; or

(b) $\mathcal{F}_i$ strictly separates the types according to their belief on $\mathcal{F}_{-i}$, i.e., $\mathcal{F}_i \succ^* \mathcal{F}_{-i}$.

Condition (ANL$_T$) imposes a measurable structure on the type sets.

(ANL$_T$) There is a mutual-separation pair $(\mathcal{F}_a^{\mathcal{T}}, \mathcal{F}_b^{\mathcal{T}})$ (not necessarily in $\mathscr{F}_a, \mathscr{F}_b$) such that $(T_a, \mathcal{F}_a^{\mathcal{T}})$ and $(T_b, \mathcal{F}_b^{\mathcal{T}})$ are analytic Borel spaces.

This definition generalizes Harsanyi's (1968) classic definition. Recall that an $((S_a, S_b)$-based) *Harsanyi type space* is a tuple $\mathcal{T}^H = (T_a, T_b, \mathcal{F}_a, \mathcal{F}_b, \pi_a, \pi_b)$, where T_a and T_b are sets of types, $\mathcal{F}_a$ and $\mathcal{F}_b$ are σ-algebras on T_a and T_b, respectively, and $(T_a, \mathcal{F}_a)$ and $(T_b, \mathcal{F}_b)$ are analytic Borel spaces. The function π_i maps each type for player i into a belief $\pi_i(t_i) \in \Delta(S_{-i} \times T_{-i}, \overline{\mathcal{F}}_{-i})$, and is measurable (with respect to $\mathcal{F}_i$ and $\mathcal{F}_{\Delta(S_{-i} \times T_{-i}, \overline{\mathcal{F}}_{-i})}$).

Proposition 5.1. Every Harsanyi type space is a type space.

Proof. Let $\mathcal{T}^H = (T_a, T_b, \mathcal{F}_a, \mathcal{F}_b, \pi_a, \pi_b)$ be a Harsanyi type space. The following result implies that $\mathcal{F}_a$ and $\mathcal{F}_b$ form a mutual-separation pair.

Lemma 5.2. The σ-algebras $\mathcal{F}_a$ and $\mathcal{F}_b$ form a mutual-separation pair if and only if the functions π_a and π_b are $(\mathcal{F}_a, \mathcal{F}_{\Delta(S_b \times T_b, \overline{\mathcal{F}}_b)})$-measurable and $(\mathcal{F}_b, \mathcal{F}_{\Delta(S_a \times T_a, \overline{\mathcal{F}}_a)})$-measurable, respectively.

Proof. Fix a player $i = a, b$. The function π_i is $(\mathcal{F}_i, \mathcal{F}_{\Delta(S_{-i} \times T_{-i}, \overline{\mathcal{F}}_{-i})})$-measurable if and only if

$$\{t_i \in T_i : \pi_{t_i}(E) \geq p\} \in \mathcal{F}_i$$

for every $E \in \overline{\mathcal{F}}_{-i}$ (Kechris, 1995, p. 66). So, π_i is $(\mathcal{F}_i, \mathcal{F}_{\Delta(S_{-i} \times T_{-i}, \overline{\mathcal{F}}_{-i})})$-measurable if and only if $\mathcal{F}_i$ separates the types according to their belief on $\mathcal{F}_{-i}$. □

It now follows that $\mathcal{T}^H$ satisfies (SEP) and (ANL$_T$). Hence, $\mathcal{T}^H$ is a type space. □

Proposition 5.1 implies that condition (SEP) weakens the standard condition that belief maps be measurable. Indeed many type spaces are not Harsanyi type spaces, as we have in Section 2.

6 Types and belief hierarchies are equally expressive

The main result of this section, Theorem 6.1, shows that the type space approach and the belief hierarchy approach are equally expressive in the sense that any situation that can be modeled with belief hierarchies can be modeled with types and vice versa. That is, there is a one-to-one relation between models and nonredundant type spaces.[9] Theorem 6.1 can informally be stated as follows:

Theorem 6.1. (Informal) Every nonredundant type space corresponds to a model; and, conversely, every model corresponds to a nonredundant type space.

Theorem 6.1 has the following corollary:

Corollary 6.2. (Informal) Every type generates a belief hierarchy; and, conversely, every belief hierarchy defines a type.

Theorem 6.1 and Corollary 6.2 imply modeling belief hierarchies by types is without loss of generality: type spaces do not impose any restrictions on the belief hierarchies that can be modeled.

The type space approach and the belief hierarchy approach may fail to be equally expressive for at least two reasons. First, there may be belief hierarchies that cannot be extended to a type. This can be the case if no measurable structure is imposed on players' beliefs (Heifetz and Samet, 1999). In such a case, the belief hierarchy approach is more expressive than the type space approach. Section 6.2 shows that in the present setting, where the relevant spaces are assumed to be analytic Borel, every belief hierarchy can be extended to a type.

Second, the class of models must be large enough so that every type space corresponds to a model in the class but not so large that some models do not correspond to type spaces. A key requirement is that the type of bounded reasoning that can be modeled using type spaces can alternatively be captured by models and vice versa.[10] This is not guaranteed: Tsakas (2014) considers players that assign only rational probabilities (i.e., probabilities in $\mathbb{Q}$) to events and shows the surprising result that there are belief hierarchies that assign rational probabilities to every event that correspond to types that assign irrational probabilities to certain events. This implies that "rational" belief hierarchies are more expressive than "rational" types (i.e.,

[9] A type space is nonredundant if no two types induce the same belief hierarchy (Mertens and Zamir, 1985); see Section 6.1 for a formal definition.

[10] Additionally, even if there are no restrictions on players' beliefs and reasoning, types and belief hierarchies may fail to be equally expressive if the measurable structure on models is not coordinated with the measurable structure on type spaces; see Mertens, Sorin, and Zamir (1994). In the present context, this is ensured by conditions (ANL_H) and (ANL_T).

types that assign only rational probabilities to events). In the present context, the key issue is to ensure that the class of events that belief hierarchies of a certain depth can reason about matches the class of events that the corresponding types can reason about. Section 6.1 (in particular, Proposition 6.3) shows that there is a tight connection between the measurable structures on the type sets (given by (SEP)) and on models (given by (IND)) that ensures that any restrictions on players' reasoning that can be modeled by belief hierarchies can be modeled by types and vice versa.

The remainder of this section defines the relevant terms and proves the results formally. Section 6.1 shows that every type space defines a model. Section 6.2 demonstrates that every model defines a type space. Section 6.3 uses these results to state and prove Theorem 6.1 and Corollary 6.2.

6.1 From types to belief hierarchies

This section shows that every type space defines a model. I do so by simultaneously constructing a model for a given type space and the functions that map types into belief hierarchies. Fix a type space $\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$. For each player $i = a, b$ and type $t_i \in T_i$, the type's first-order belief is the marginal of its belief π_{t_i} on S_{-i}, that is:

$$\mu^1_{t_i} := \operatorname{marg}_{S_{-i}} \pi_{t_i}.$$

Let $h_i^{\mathcal{T},1}$ be the function that assigns to each type its first-order belief. That is,

$$h_i^{\mathcal{T},1}(t_i) := \mu^1_{t_i}.$$

Define $H_i^{\mathcal{T},1} := h_i^{\mathcal{T},1}(T_i) \subset \Delta(S_{-i})$ to be the image of $h_i^{\mathcal{T},1}$. Let $\mathcal{F}_{i,1}^{\mathcal{T},1}$ be the relative σ-algebra on $H_i^{\mathcal{T},1}$ induced by $\mathcal{F}_{\Delta(S_{-i})}$.

For $m = 2, 3, \ldots$, suppose that for each player $i = a, b$ and for every $\ell = 1, 2, \ldots, m-1$, the set $H_i^{\mathcal{T},\ell}$ and the function $h_i^{\mathcal{T},\ell}$ (from T_i to $H_i^{\mathcal{T},\ell}$) have been defined, and that $\mathcal{F}_{i,\ell}^{\mathcal{T},\ell}$ is a σ-algebra on $H_i^{\mathcal{T},\ell}$. Fix a player $i = a, b$ and $\ell = 1, \ldots, m-2$. Let

$$\mathcal{F}_{i,\ell}^{\mathcal{T},m-1} := \left\{ \left\{ (\mu_i^1, \ldots, \mu_i^\ell, \ldots, \mu_i^{m-1}) : (\mu_i^1, \ldots, \mu_i^\ell) \in B_i^\ell \right\} : B_i^\ell \in \mathcal{F}_{i,\ell}^\ell \right\}$$

be the σ-algebra on $H_i^{\mathcal{T},m-1}$ generated by the projection function into $H_i^{\mathcal{T},\ell}$ (cf. Eq. (4.1)), and let $\mathcal{F}_{i,0}^{\mathcal{T},m-1}$ be the trivial σ-algebra on $H_i^{\mathcal{T},m-1}$. Define

$$\Delta^+(S_{-i} \times H_{-i}^{\mathcal{T},m-1}) := \bigcup_{\ell=0}^{m-1} \Delta(S_{-i} \times H_{-i}^{\mathcal{T},m-1}, \overline{\mathcal{F}}_{-i,\ell}^{\mathcal{T},m-1}),$$

and define the mth-order belief $\mu_{t_i}^m \in \Delta^+(S_{-i} \times H_{-i}^{\mathcal{T},m-1})$ induced by t_i by

$$\mu_{t_i}^m := \pi_{t_i} \circ \left(\mathrm{Id}_{S_{-i}}, h_{-i}^{\mathcal{T},m-1}\right)^{-1},$$

where Id_X is the identity function on X.[11] That is, for any $E \in \Sigma(\mu_{t_i}^m)$,

$$\mu_{t_i}^m(E) = \pi_{t_i}\left(\left\{(s_{-i}, t_{-i}) : (s_{-i}, h_{-i}^{\mathcal{T},m-1}(t_{-i})) \in E\right\}\right).$$

Define the function $h_i^{\mathcal{T},m}$ from T_i to $H_i^{\mathcal{T},m-1} \times \Delta^+(S_{-i} \times H_{-i}^{\mathcal{T},m-1})$ by

$$h_i^{\mathcal{T},m}(t_i) := \left(h_i^{\mathcal{T},m-1}(t_i), \mu_{t_i}^m\right).$$

Let $H_i^{\mathcal{T},m}$ be the image of $h_i^{\mathcal{T},m}$, and let $\mathcal{F}_{i,m}^{\mathcal{T},m}$ be the relative σ-algebra on $H_i^{\mathcal{T},m}$.

Proposition 6.3. The function $h_i^{\mathcal{T},k}$ is well-defined. That is, $h_i^{\mathcal{T},k}(t_i) \in H_i^{\mathcal{T},k}$ for all $t_i \in T_i$.

The proof of Proposition 6.3 is deferred to the end of this section. A key part of the proof is to establish a *one-to-one* relation between the measurable structure on the sets of belief hierarchies and the measurable structure on the type sets. That this can be done is not obvious: condition (SEP) on the σ-algebras on type spaces makes no reference to belief hierarchies. The proof outline in Section 6.1.1 below discusses how it is nevertheless possible to "match" each σ-algebra on a type set to a σ-algebra on the set of belief hierarchies. This one-to-one relationship will also be central to the equivalence result Theorem 6.1.

We are now ready to define the model induced by the type space $\mathcal{T}$. For every type $t_i \in T_i$, let

$$h_i^{\mathcal{T}}(t_i) := \left(\mu_{t_i}^1, \mu_{t_i}^2, \ldots\right)$$

and

$$H_i^{\mathcal{T}} := \{h_i^{\mathcal{T}}(t_i) : t_i \in T_i\}.$$

Then:

Proposition 6.4. The pair $\mathcal{H}^{\mathcal{T}} := (H_a^{\mathcal{T}}, H_b^{\mathcal{T}})$ is a model.

Proof. It is immediate that conditions (COH) and (EXT) hold, and Proposition 6.3 implies that condition (IND) is satisfied. The proof that (ANL$_H$) holds is relegated to the appendix. □

I refer to the functions $h_a^{\mathcal{T}}$ and $h_b^{\mathcal{T}}$ as *hierarchy mappings*. If the hierarchy mappings are one-to-one, then different types generate different belief hierarchies, and $\mathcal{T}$ is *nonredundant* (Mertens and Zamir, 1985). With some abuse of terminology, if a belief hierarchy $h_i^{\mathcal{T}}(t_i)$ has depth $k \leq \infty$, then I say that type t_i has depth k. By Proposition 4.3 and 6.4, every type has a well-defined depth of reasoning.

[11] For functions $f_1 : X_1 \to Y_1$ and $f_2 : X_2 \to Y_2$, (f_1, f_2) is the function from $X_1 \times X_2$ to $Y_1 \times Y_2$ defined by $(f_1, f_2)(x_1, x_2) = (f_1(x_1), f_2(x_2))$.

6.1.1 Proof of Proposition 6.3

Outline of proof. It is instructive to first consider the special case where $\mathcal{T}$ is a Harsanyi type space. Recall that by Proposition 5.1, a Harsanyi type space $(T_a, T_b, \mathcal{F}_a, \mathcal{F}_b, \pi_a, \pi_b)$ can be viewed as a (general) type space $\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$ if we define $\mathscr{F}_i := \{\mathcal{F}_i\}$.

Claim. Suppose $\mathcal{T} = (T_a, T_b, \mathcal{F}_a, \mathcal{F}_b, \pi_a, \pi_b)$ is a Harsanyi type space. Then, the function $h_i^{\mathcal{T},k}$ is well-defined. ◁

Proof of claim. The proof is standard (Mertens and Zamir, 1985, pp. 5–6). In the Harsanyi case, all types are endowed with the same σ-algebra, and we have $\Sigma_{t_i} = \mathcal{F}_{-i}$ for all $t_i \in T_i$. The key step is to show (inductively) that $h_{-i}^{\mathcal{T},k-1}$ is measurable (with respect to Σ_{t_i} and $\mathcal{F}_{-i,k-1}^{\mathcal{T},k-1}$). This implies that the kth-order belief $\mu_{t_i}^k := \pi_{t_i} \circ (\mathrm{Id}_{S_{-i}}, h_{-i}^{\mathcal{T},k-1})^{-1}$ is defined on $\overline{\mathcal{F}}_{-i,k-1}^{\mathcal{T},k-1}$, and the result follows. That the function $h_{-i}^{\mathcal{T},k-1}$ is measurable follows from the fact that the belief maps π_a and π_b are measurable.[12] □

So, in the Harsanyi case, the belief maps π_a and π_b are measurable, and this implies that for each type t_i, the function $h_{-i}^{\mathcal{T},k-1}$ is measurable with respect to Σ_{t_i} and $\overline{\mathcal{F}}_{i,k-1}^{\mathcal{T},k-1}$. In the general case, types can be endowed with different σ-algebras, and belief maps need not be measurable. As a result, $h_{-i}^{\mathcal{T},k-1}$ may not be measurable with respect to Σ_{t_i} and $\overline{\mathcal{F}}_{i,k-1}^{\mathcal{T},k-1}$ for some types t_i. The proof for Harsanyi type spaces thus does not extend to the general case.

However, kth-order beliefs can be defined on different σ-algebras, just like types may be endowed with different σ-algebras. Accordingly, $h_i^{\mathcal{T},k}$ is well-defined if for every type t_i, there is *some* $m \leq k-1$ such that $h_{-i}^{\mathcal{T},k-1}$ is measurable with respect to Σ_{t_i} and $\overline{\mathcal{F}}_{i,m}^{\mathcal{T},k-1}$. Then, every type t_i generates a well-defined kth-order belief hierarchy even if $h_{-i}^{\mathcal{T},k-1}$ is not measurable with respect to Σ_{t_i} and $\overline{\mathcal{F}}_{i,k-1}^{\mathcal{T},k-1}$.

So, the goal is to "match" the σ-algebra Σ_{t_i} (defined on the type set T_{-i}) to a σ-algebra $\mathcal{F}_{-i,m}^{\mathcal{T},k-1}$, $m = 0, \ldots, k-1$ (defined on the set $H_{-i}^{\mathcal{T},k-1}$ of $(k-1)$th-order belief hierarchies) in such a way that $h_{-i}^{\mathcal{T},k-1}$ is measurable with respect to Σ_{t_i} and $\overline{\mathcal{F}}_{i,m}^{\mathcal{T},k-1}$.

That it is possible to find a match for every type t_i is not obvious. The measurable structure on the type spaces, as given by condition (SEP), does not make reference to belief hierarchies: it makes reference only to other σ-algebras on the type set.[13] I resolve this problem by inductively constructing σ-algebras $\mathcal{Q}_i^m$, $m = 0, 1, \ldots$, on the type sets. I do so by building on the relation

[12] To see this, note that the composition of measurable functions is measurable, and that a function that maps probability measures into image measures is measurable.

[13] Also, while the σ-algebras $\mathcal{F}_{-i,m}^{\mathcal{T},k-1}$, $m = 0, \ldots, k-1$, on the $(k-1)$th-order belief hierarchies can be ordered by set inclusion, condition (SEP) does not impose such order on the σ-algebras on the type sets; see Lemma 6.9 below.

between types and belief hierarchies established in earlier induction steps. The σ-algebras $\mathcal{Q}_i^m$, $m = 0, 1, \ldots$, have a direct relationship with the σ-algebras on the belief hierarchies (Lemma 6.6). I then relate the σ-algebra $\Sigma_{t_{-i}}$ of each type t_{-i} to the σ-algebras $\mathcal{Q}_i^m$, $m = 0, 1, \ldots$ (Lemma 6.9). This allows me to match the σ-algebras in the appropriate way and show that the hierarchy mapping $h_{-i}^{\mathcal{T},k-1}$ is measurable for every type t_i.

Proof of Proposition 6.3. For $i = a, b$, define $\mathcal{Q}_i^0$ to be the trivial σ-algebra on T_i. I show in the appendix that the following hold:

(1) For every $t_i \in T_i$, $h_i^{\mathcal{T},1}(t_i) \in H_i^{\mathcal{T},1}$;

(2) The coarsest σ-algebra $\mathcal{Q}_i^1$ on T_i that separates the types according to their belief on $\mathcal{Q}_{-i}^0$ (i.e. $\mathcal{Q}_i^1 \succ^* \mathcal{Q}_{-i}^0$) satisfies:

(2a) for $n = 0, 1$, the function $h_i^{\mathcal{T},1}$ is measurable with respect to $\mathcal{Q}_i^n$ and $\mathcal{F}_{i,n}^{\mathcal{T},1}$;

(2b) $\mathcal{Q}_i^1$ is the coarsest σ-algebra that contains the sets

$$\{t_i \in T_i : \mu_{t_i}^1 \in B_i^1\}$$

for $B_i^1 \in \mathcal{F}_{i,1}^{\mathcal{T},1}$;

(2c) $\mathcal{Q}_i^1 \supset \mathcal{Q}_i^0$;

(2d) for every $t_i \in T_i$, one of the following is the case: either $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^{\mathcal{T},1}$, or $\Sigma_{t_i} = \mathcal{Q}_{-i}^0$.

For $k = 2, 3, \ldots$, suppose, inductively, that for $i = a, b$, the σ-algebras $\mathcal{Q}_i^0, \ldots, \mathcal{Q}_i^{k-2}$ have been defined, and that

(IH1) For every $t_i \in T_i$, $h_i^{\mathcal{T},k-1}(t_i) \in H_i^{\mathcal{T},k-1}$;

(IH2) The coarsest σ-algebra $\mathcal{Q}_i^{k-1}$ on T_i that separates the types according to their belief on $\mathcal{Q}_{-i}^{k-2}$ (i.e. $\mathcal{Q}_i^{k-1} \succ^* \mathcal{Q}_{-i}^{k-2}$) satisfies:

(IH2a) for $n = 0, 1, \ldots, k-1$, the function $h_i^{\mathcal{T},k-1}$ is measurable with respect to $\mathcal{Q}_i^n$ and $\mathcal{F}_{i,n}^{\mathcal{T},k-1}$;

(IH2b) $\mathcal{Q}_i^{k-1}$ is the coarsest σ-algebra that contains the sets

$$\{t_i \in T_i : (\mu_{t_i}^1, \ldots, \mu_{t_i}^{k-1}) \in B_i^{k-1}\} \tag{6.1}$$

for $B_i^{k-1} \in \mathcal{F}_{i,k-1}^{\mathcal{T},k-1}$;

(IH2c) $\mathcal{Q}_i^{k-1} \supset \mathcal{Q}_i^{k-2}$;

(IH2d) for every type t_i, one of the following is the case: either $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^{\mathcal{T},k-1}$, or there is $n = 0, \ldots, k-2$ such that $\Sigma_{t_i} = \mathcal{Q}_{-i}^{n}$.

This inductive argument allows us to "match" the σ-algebras on the types with those on the belief hierarchies. Induction hypothesis (IH2a) shows that there is a direct relation between the σ-algebra $\mathcal{Q}_i^n$ on the type set T_i and the σ-algebras on the set of belief hierarchies. In fact, induction hypothesis (IH2b) states that $\mathcal{Q}_i^n$ is the coarsest σ-algebra that separates i's types according to their ℓth-order beliefs for $\ell \leq n$. This directly relates $\mathcal{Q}_i^n$ and the σ-algebras $\mathcal{F}_{i,n}^m$ for $m \geq n$. Induction hypothesis (IH2d) then relates the types' σ-algebras Σ_{t_i} to the σ-algebras $\mathcal{Q}_{-i}^n$, $n \geq 0$. In particular, every σ-algebra Σ_{t_i} associated with a type t_i either coincides with some $\mathcal{Q}_{-i}^n \subsetneq \mathcal{Q}_{-i}^{n+1}$, or contains all events in $\mathcal{Q}_{-i}^{k-1}$. In the former case, Σ_{t_i} can be matched with $\mathcal{F}_{-i,k-1}^{\mathcal{T},k-1}$ if $n \geq k-1$ and with $\mathcal{F}_{-i,n}^{\mathcal{T},k-1}$ otherwise. In the latter case, Σ_{t_i} can be matched with $\mathcal{F}_{-i,k-1}^{\mathcal{T},k-1}$. This implies that the hierarchy mappings are well-defined:

Lemma 6.5. For every player $i = a, b$ and type t_i, $h_i^{\mathcal{T},k}(t_i) \in H_i^{\mathcal{T},k}$.

Proof. I show that $h_i^{\mathcal{T},k}(t_i) \in H_i^{\mathcal{T},k-1} \times \Delta^+(S_{-i} \times H_{-i}^{\mathcal{T},k-1})$. By the induction hypothesis (IH1), it suffices to show that the kth-order belief $\mu_{t_i}^k$ is defined on a σ-algebra $\overline{\mathcal{F}}_{-i,m}^{\mathcal{T},k-1}$, $m = 0, \ldots, k-1$. By (IH2d), type t_i's σ-algebra Σ_{t_i} is either finer than $\mathcal{Q}_{-i}^{k-1}$, or there is $m \leq k-2$ such that $\Sigma_{t_i} = \mathcal{Q}_{-i}^m$. It then follows from the (IH2a) that $\mu_{t_i}^k$ is defined on $\overline{\mathcal{F}}_{-i,m}^{\mathcal{T},k-1}$ for some m. □

It remains to complete the induction. Proofs that are not included here can be found in the appendix.

Lemma 6.6. The σ-algebra $\mathcal{Q}_i^k$ is the coarsest σ-algebra that contains the sets

$$\{t_i \in T_i : (\mu_{t_i}^1, \ldots, \mu_{t_i}^k) \in B_i^k\} \tag{6.2}$$

for $B_i^k \in \mathcal{F}_{i,k}^{\mathcal{T},k}$.

Proof. By the induction hypothesis (IH2b), it suffices to show that the σ-algebra generated by the function $h_i^{\mathcal{T},k}$ (i.e., the coarsest σ-algebra that contains the sets in (6.2)) is precisely the coarsest σ-algebra that separates the types according to their belief on the σ-algebra generated by the function $h_{-i}^{\mathcal{T},k-1}$ (i.e., the coarsest σ-algebra that contains the sets in (6.1)). For $m \geq 1$, denote the σ-algebra generated by $h_i^{\mathcal{T},m}$ (and $\mathcal{F}_{i,m}^m$) by $\mathcal{S}_i^{\mathcal{T},m}$, and let $\mathcal{S}_i^{\mathcal{T},0}$ be the trivial σ-algebra on T_i. So, we need to show that $\mathcal{S}_i^{\mathcal{T},k} \succ^* \mathcal{S}_{-i}^{\mathcal{T},k-1}$.

To show this, let $m \leq k$. By the coherency condition (COH), $\mathcal{S}_i^{\mathcal{T},m}$ is the coarsest σ-algebra that contains the sets

$$\left\{t_i \in T_i : \overline{\Sigma}(\mu_{t_i}^m) = \overline{\mathcal{F}}_{-i,n}^{\mathcal{T},m-1}, \mu_{t_i}^m(E) \geq p\right\}$$

for $n = 0, \dots, m-1$, $E \in \overline{\mathcal{F}}_{-i,n}^{\mathcal{T},m-1}$, and $p \in [0,1]$ (Aliprantis and Border, 2005, Lemma 4.23). By Lemma A.1 in the appendix, and using that the σ-algebras $\mathcal{F}_{-i,0}^{\mathcal{T},m-1}, \dots, \mathcal{F}_{-i,m-1}^{\mathcal{T},m-1}$ can be ordered by set inclusion, $\mathcal{S}_i^{\mathcal{T},m}$ is the coarsest σ-algebra that contains the sets

$$\left\{t_i \in T_i : E \in \overline{\Sigma}(\mu_{t_i}^m), \mu_{t_i}^m(E) \geq p\right\}$$

for $E \in \overline{\mathcal{F}}_{-i,m-1}^{\mathcal{T},m-1}$, and $p \in [0,1]$, or, equivalently, the sets

$$\left\{t_i \in T_i : E' \in \overline{\Sigma}_{t_i}, \pi_{t_i}(E') \geq p\right\}$$

for $E \in \mathcal{S}_{-i}^{\mathcal{T},m-1}$ and $p \in [0,1]$. That is, $\mathcal{S}_i^{\mathcal{T},m}$ is the coarsest σ-algebra that separates the types according to their belief on $\mathcal{S}_{-i}^{\mathcal{T},m-1}$. Note that, by (IH1), the σ-algebras $\mathcal{S}_{-i}^{\mathcal{T},m-1}$ and $\mathcal{S}_i^{\mathcal{T},m}$ are well-defined for $m \leq k-1$. So, $\mathcal{Q}_i^k = \mathcal{S}_i^{\mathcal{T},k}$; and by Lemma 6.5, $\mathcal{Q}_i^k$ is well-defined. □

Lemma 6.7. The function $h_i^{\mathcal{T},k}$ is $(\mathcal{Q}_i^n, \mathcal{F}_{i,n}^{\mathcal{T},k})$-measurable for $n = 0, 1, \dots, k$.

Lemma 6.8. The σ-algebra $\mathcal{Q}_i^k$ is at least as fine as $\mathcal{Q}_i^{k-1}$, that is, $\mathcal{Q}_i^k \supset \mathcal{Q}_i^{k-1}$.

This follows directly from Lemma 6.6.

Lemma 6.9. For every type t_i, one of the following is the case: either $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^k$, or there is $n = 0, \dots, k-1$ such that $\Sigma_{t_i} = \mathcal{Q}_{-i}^n$.

Proof. I first show that either $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^k$, or $\Sigma_{t_i} \subseteq \mathcal{Q}_{-i}^k$. If Σ_{t_i} is the trivial σ-algebra $\{T_{-i}, \emptyset\}$, then clearly $\Sigma_{t_i} \subseteq \mathcal{Q}_{-i}^k$. So suppose $\Sigma_{t_i} \neq \{T_{-i}, \emptyset\}$. By condition (SEP), one of the following is the case:

(a) Σ_{t_i} is part of a finite chain, that is, there exist $n < \infty$ and (distinct) σ-algebras $\mathcal{F}_i^1, \mathcal{F}_i^3, \dots, \mathcal{F}_i^n \in \mathscr{F}_i$ and $\mathcal{F}_{-i}^2, \mathcal{F}_{-i}^4 \dots, \mathcal{F}_{-i}^n \in \mathscr{F}_{-i}$ such that

$$\Sigma_{t_i} \succ^* \mathcal{F}_i^1 \succ^* \mathcal{F}_{-i}^2 \succ^* \cdots \succ^* \mathcal{F}_i^n = \{T_i, \emptyset\}$$

if n is odd, and

$$\Sigma_{t_i} \succ^* \mathcal{F}_i^1 \succ^* \mathcal{F}_{-i}^2 \succ^* \cdots \succ^* \mathcal{F}_{-i}^n = \{T_{-i}, \emptyset\}$$

if n is even;

(b) Σ_{t_i} is part of a cycle or infinite chain, that is, there exist σ-algebras $\mathcal{F}_i^1, \mathcal{F}_i^3, \dots \in \mathscr{F}_i$ and $\mathcal{F}_{-i}^2, \mathcal{F}_{-i}^4, \dots \in \mathscr{F}_{-i}$ (not necessarily distinct) such that

$$\Sigma_{t_i} \succ^* \mathcal{F}_i^1 \succ^* \mathcal{F}_{-i}^2 \succ^* \mathcal{F}_i^3 \succ^* \cdots .$$

(c) Σ_{t_i} is part of a mutual-separation pair, that is, there is $\mathcal{F}_i \in \mathscr{F}_i$ such that $\Sigma_{t_i} \succ \mathcal{F}_i$ and vice versa.

For case (a), the result follows from the definition of $\mathcal{Q}_i^n$, $n \leq k$, Lemma 6.6, and Lemma 6.8. I claim that in case (b) or (c), $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^{\mathcal{T},k}$. I present the argument for (b); the argument for (c) is similar and thus omitted. So, suppose (b) is the case. By assumption, Σ_{t_i} separates the types according to their belief on $\mathcal{F}_i^1 \supseteq \mathcal{Q}_i^0$. So, $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^1$. Likewise, $\mathcal{F}_i^1$ separates the types according to their belief on $\mathcal{F}_{-i}^2$. Hence, $\mathcal{F}_i^1 \supseteq \mathcal{Q}_i^1$. Since Σ_{t_i} separates the types according to their belief on $\mathcal{F}_i^1$, $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^2$ (by the definition of $\mathcal{Q}_i^n$). Repeating this argument gives the desired result. □

Lemma 6.9 classifies the types' σ-algebras. If Σ_{t_i} is part of a finite chain (case (a)), then $\Sigma_{t_i} = \mathcal{Q}_i^m$ for some $m < \infty$. Otherwise — if Σ_{t_i} is part of a cycle, infinite chain, or mutual-separation pair, as in case (b) or (c) —, Σ_{t_i} is finer than any σ-algebra $\mathcal{Q}_i^m$ (i.e., $\Sigma_{t_i} \supset \mathcal{Q}_i^m$ for all m).[14] □

Write $\mathcal{Q}_i^\infty$ for the coarsest σ-algebra that contains the sets in $\mathcal{Q}_i^n$, $n \geq 0$.

6.2 From belief hierarchies to types

This section shows that every model defines a type space. Given a model $\mathcal{H} = (H_a, H_b)$, a *standard-form type space* for $\mathcal{H}$ is a type space $\mathcal{T}^\mathcal{H} = (T_a^\mathcal{H}, T_b^\mathcal{H}, \mathscr{F}_a^\mathcal{H}, \mathscr{F}_b^\mathcal{H}, \pi_a^\mathcal{H}, \pi_b^\mathcal{H})$ such that (1) the type set for each player i is the set H_i of belief hierarchies; and (2) for every k, the marginal of the belief π_{h_i} associated with type $h_i = (\mu_i^1, \mu_i^2, \ldots)$ over the first k orders of the other player's beliefs is precisely what it should be, namely μ_i^{k+1}. The next result shows that every model has a standard-form type space:

Proposition 6.10. Let $\mathcal{H}$ be a model. Then, there is a standard-form type space $\mathcal{T}^\mathcal{H}$ for $\mathcal{H}$.

Proof. Let $\mathcal{H} = (H_a, H_b)$ be a model. To construct a standard-form type space, let the type sets for players a and b be the sets H_a and H_b of belief hierarchies, respectively. I show that for each type h_i, there is a unique belief $\pi_{h_i}^\mathcal{H}$ on $S_{-i} \times H_{-i}$ such that its marginal on $S_{-i} \times H_{-i}^{m-1}$ coincides with the mth-order belief induced by h_i. The first result considers belief hierarchies with an infinite depth of reasoning.

Lemma 6.11. Fix a player $i = a, b$ and a belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots) \in H_i$ with an infinite depth of reasoning. Then, there is a unique belief $\pi_{h_i}^\mathcal{H}$ on $\overline{\mathcal{F}}_{H_{-i}}$ such that its marginal on $S_{-i} \times H_{-i}^{m-1}$ equals μ_i^m, $m = 1, 2 \ldots$.

[14] A pair of σ-algebras that are part of a cycle or that belong to different mutual-separation pairs cannot necessarily be ordered by set inclusion, i.e., we could have $\mathcal{F}_i \not\subset \mathcal{F}_i'$ and $\mathcal{F}_i' \not\subset \mathcal{F}_i$.

The proof, which can be found in the appendix, uses an extension theorem due to Choksi (1958). Unlike other extension results in the literature (e.g., Mertens and Zamir, 1985; Brandenburger and Dekel, 1993; Heifetz, 1993; Mertens, Sorin, and Zamir, 1994), it does not require that functions are continuous. This makes it possible to prove the result without introducing continuity assumptions that would obfuscate that the key issue is the relation between the measurable structure on belief hierarchies and types.

The next result concerns types that correspond to belief hierarchies with a finite depth of reasoning. To state the result, define $\mathcal{F}_{i,m}$ to be the σ-algebra on H_i generated by the projection function from H_i into H_i^m when H_i^m is endowed with the σ-algebra $\mathcal{F}_{i,m}^m$. That is, $\mathcal{F}_{i,m}$ contains precisely the sets

$$\big\{(\mu_i^1, \mu_i^2, \ldots) \in H_i : (\mu_i^1, \ldots, \mu_i^m) \in B_i^m\big\}$$

for $B_i^m \in \mathcal{F}_{i,m}^m$. In words, $\mathcal{F}_{i,m}$ separates the belief hierarchies in H_i if they differ in their beliefs at some order $\ell \leq m$, and lumps them together otherwise. This suggests that the belief associated with a belief hierarchy of finite depth k can assign a probability precisely to the events in $\mathcal{F}_{-i,k-1}$. The following result shows that this is indeed the case:

Lemma 6.12. Fix a player $i = a, b$ and a belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots) \in H_i$ that has depth $k < \infty$. Then, there is a unique belief $\pi_{h_i}^{\mathcal{H}}$ on $\overline{\mathcal{F}}_{-i,k-1}$ such that its marginal on $S_{-i} \times H_{-i}^{m-1}$ equals μ_i^m, $m = 1, 2 \ldots$.

Proof. Define $\pi_{h_i}^{\mathcal{H}} \in \Delta(S_{-i} \times H_{-i}, \overline{\mathcal{F}}_{-i,k-1})$ by:

$$E \in \overline{\mathcal{F}}_{-i,k-1} : \qquad \pi_{h_i}^{\mathcal{H}}(E) := \mu_i^k\big((\mathrm{Id}_{S_{-i}}, \mathrm{proj}_{-i,k-1})(E)\big), \tag{6.3}$$

where $\mathrm{proj}_{-i,k-1}$ is the projection function from H_{-i} into H_{-i}^{k-1} when H_{-i}^{k-1} is endowed with the σ-algebra $\mathcal{F}_{-i,k-1}^{k-1}$. As $(\mathrm{Id}_{S_{-i}}, \mathrm{proj}_{-i,k-1})(E) \in \overline{\mathcal{F}}_{-i,k-1}^{k-1}$ for $E \in \overline{\mathcal{F}}_{-i,k-1}$, the belief $\pi_{h_i}^{\mathcal{H}}$ is well-defined. Clearly, $\pi_{h_i}^{\mathcal{H}}$ satisfies the desired properties. To see that it is unique, note that any belief that satisfies the desired properties is determined completely by the kth-order belief μ_i^k, and must therefore satisfy (6.3). □

Lemmas 6.11 and 6.12 can be used to define a type space. Fix a player $i = a, b$. Write $\mathcal{F}_{i,0}$ and $\mathcal{F}_{i,\infty}$ for $\{H_i, \emptyset\}$ and $\mathcal{F}_{H_i}$, respectively, and define

$$\Delta^+(S_{-i} \times H_{-i}) := \Delta(S_{-i} \times H_{-i}, \overline{\mathcal{F}}_{-i,\infty}) \cup \bigcup_{m=0}^{\infty} \Delta(S_{-i} \times H_{-i}, \overline{\mathcal{F}}_{-i,m})$$

for the set of all beliefs and endow $\Delta^+(S_{-i} \times H_{-i})$ with its usual σ-algebra, denoted $\mathcal{F}_{\Delta^+(S_{-i} \times H_{-i})}$. Let $\pi_i^{\mathcal{H}}$ be the function that assigns to each belief hierarchy h_i the belief $\pi_{h_i}^{\mathcal{H}}$ (defined in Lemmas

6.11 and 6.12). So, the belief $\pi^{\mathcal{H}}_{h_i}$ is defined on the σ-algebra $\overline{\Sigma}^{\mathcal{H}}_{h_i} := \overline{\mathcal{F}}_{-i,k-1}$ if h_i has depth k (where $\infty - 1 = \infty$). Finally, define

$$\mathscr{F}^{\mathcal{H}}_i := \{\mathcal{F}_{i,m} : m = 0, 1, \ldots, \infty\}.$$

I claim that

$$\mathcal{T}^{\mathcal{H}} := (H_a, H_b, \mathscr{F}^{\mathcal{H}}_a, \mathscr{F}^{\mathcal{H}}_b, \pi^{\mathcal{H}}_a, \pi^{\mathcal{H}}_b)$$

is a type space. This follows if Conditions (SEP) and (ANL$_T$) are satisfied. The following two results establish this.

Lemma 6.13. The space $\mathcal{T}^{\mathcal{H}}$ satisfies Condition (SEP).

Proof. Fix a player $i = a, b$. I first show that $\mathcal{F}_{i,k} \succ^* \mathcal{F}_{-i,k-1}$ for $k = 1, 2, \ldots$. I prove the result for $k > 1$; the proof for $k = 1$ is similar and thus omitted. By definition, $\mathcal{F}_{i,k} \succ^* \mathcal{F}_{-i,k-1}$ if $\mathcal{F}_{i,k}$ is the coarsest σ-algebra on H_i that contains the sets

$$\{h_i \in H_i : E \in \overline{\Sigma}^{\mathcal{H}}_{h_i}, \pi^{\mathcal{H}}_{h_i}(E) \geq p\} : \qquad E \in \overline{\mathcal{F}}_{-i,k-1}, p \in [0,1]. \tag{6.4}$$

By Lemmas 6.11 and 6.12, it suffices to show that $\mathcal{F}_{i,k}$ is the coarsest σ-algebra that contains the sets

$$\{(\mu^1_i, \mu^2_i, \ldots) \in H_i : E' \in \Sigma(\mu^k_i), \mu^k_i(E') \geq p\} : \qquad E' \in \overline{\mathcal{F}}^{k-1}_{-i,k-1}, p \in [0,1].$$

By Lemma A.1 in the appendix, this is equivalent to showing that $\mathcal{F}_{i,k}$ is the coarsest σ-algebra that contains the sets

$$\{(\mu^1_i, \mu^2_i, \ldots) \in H_i : \Sigma(\mu^k_i) = \overline{\mathcal{F}}^{k-1}_{-i,\ell}, \mu^k_i(E'') \geq p\} : \qquad \ell = 0, \ldots, k-1, E'' \in \overline{\mathcal{F}}^{k-1}_{-i,\ell}, p \in [0,1].$$

This follows from (COH) and (IND). The full argument is presented in the appendix. The appendix also shows that $\mathcal{F}_{a,\infty}$ and $\mathcal{F}_{b,\infty}$ form a mutual-separation pair. □

Lemma 6.14. The space $\mathcal{T}^{\mathcal{H}}$ satisfies Condition (ANL$_T$).

Proof. The result follows from Lemmas 4.1 and 6.13. By Lemma 6.13, $\mathcal{F}_{a,\infty}$ and $\mathcal{F}_{b,\infty}$ form a mutual-separation pair; and by Lemma 4.1, $(H_a, \mathcal{F}_{a,\infty})$ and $(H_b, \mathcal{F}_{b,\infty})$ are analytic Borel spaces. □

So, by Lemmas 6.13 and 6.14, $\mathcal{T}^{\mathcal{H}}$ is a type space; and, by Lemmas 6.11 and Lemma 6.12, it is in fact the standard-form type space for $\mathcal{H}$. By construction, a type h_i with σ-algebra $\overline{\Sigma}^{\mathcal{H}}_{h_i} = \overline{\mathcal{F}}_{-i,k-1}$ has depth k (where $\infty - 1 = \infty$). □

Proposition 6.10 establishes that every model defines a type space. It has an immediate corollary:

Corollary 6.15. The canonical Harsanyi model $\mathcal{H}^h$ in Example 1 has a standard-form type space (and is a Harsanyi type space).

That the canonical Harsanyi model defines a type space is well-known (Mertens and Zamir, 1985). Proposition 6.10 shows that in fact *every* model defines a type space. For example:

Corollary 6.16. The canonical model $\mathcal{H}^*$ in Example 2 has a standard-form type space.

For future reference, denote the standard-form type spaces corresponding to the canonical models $\mathcal{H}^h$ and $\mathcal{H}^*$ by $\mathcal{T}^h$ and $\mathcal{T}^*$, respectively.

6.3 Belief isomorphisms

Sections 6.1 and 6.2 show that every type space defines a model and that every model defines a type space. This section shows that every type space is in fact isomorphic to a model and vice versa. Say that a type space $\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$ and a model $\mathcal{H} = (H_a, H_b)$ are *belief-isomorphic* if for each player i, the hierarchy mapping $h_i^{\mathcal{T}}$ is an isomorphism between T_i and H_i with respect to the σ-algebras $\mathcal{Q}_i^n$ and $\mathcal{F}_{i,n}$, $n = \infty, 0, 1, \ldots$, on the types and belief hierarchies, respectively. Thus, a type space $\mathcal{T}$ and a model $\mathcal{H}$ are belief-isomorphic if every type in $\mathcal{T}$ corresponds to precisely one belief hierarchy in $\mathcal{H}$ and vice versa; and, moreover, every event in $\mathcal{Q}_i^n$, $n = \infty, 0, 1, \ldots$, corresponds to an event in $\mathcal{F}_{i,n}$ and vice versa.

Theorem 6.1. Types and belief hierarchies are equally expressive. That is:

(a) For every nonredundant type space $\mathcal{T}$, there is a model $\mathcal{H}$ such that $\mathcal{T}$ and $\mathcal{H}$ are belief-isomorphic.

(b) For every model $\mathcal{H}$, there is a (nonredundant) type space $\mathcal{T}$ such that $\mathcal{T}$ and $\mathcal{H}$ are belief-isomorphic.

Proof. The proofs of (a) follows if we take $\mathcal{H} = \mathcal{H}^{\mathcal{T}}$ (where $\mathcal{H}^{\mathcal{T}}$ is as defined in Section 6.1) and show that the hierarchy mappings are isomorphisms with respect to the relevant σ-algebras. Likewise, the proof of (b) follows if we take $\mathcal{T}$ to be the standard-form type space $\mathcal{T}^{\mathcal{H}}$ for $\mathcal{H}$ defined in Section 6.2 and show that the hierarchy mappings are isomorphisms. I prove these claims by showing that for any nonredundant type space $\mathcal{T}$, the hierarchy mappings $h_i^{\mathcal{T}} : T_i \to H_i^{\mathcal{T}}$, $i = a, b$, are isomorphisms with respect to the σ-algebras $\mathcal{Q}_i^n$ and $\mathcal{F}_{i,n}$, $n = \infty, 0, 1, \ldots$.

So, let $\mathcal{T}$ be a nonredundant type space. Then, the hierarchy mappings are one-to-one and onto. Fix a player $i = a, b$, and denote the inverse of $h_i^{\mathcal{T}}$ by $g_i^{\mathcal{T}}$. Let $n < \infty$. The hierarchy mapping $h_i^{\mathcal{T}}$ is measurable with respect to $\mathcal{Q}_i^n$ and $\mathcal{F}_{i,n}$ if for each $E \in \mathcal{F}_{i,n}$, we have

$\{t_i \in T_i : h_i^{\mathcal{T}}(t_i) \in E\} \in \mathcal{Q}_i^n$. Fix $E \in \mathcal{F}_{i,n}$. Then, by the definition of $\mathcal{F}_{i,n}$, there is $B_i^n \in \mathcal{F}_{i,n}^n$ such that

$$\{t_i \in T_i : h_i^{\mathcal{T}}(t_i) \in E\} = \{t_i \in T_i : h_i^{\mathcal{T},n}(t_i) \in B_i^n\}.$$

By Lemma 6.6, $\{t_i \in T_i : h_i^{\mathcal{T},n}(t_i) \in B_i^n\} \in \mathcal{Q}_i^n$. So, $h_i^{\mathcal{T}}$ is measurable. Its inverse is measurable with respect to $\mathcal{F}_{i,n}$ and $\mathcal{Q}_i^n$ if for each $E \in \mathcal{Q}_i^n$, we have $\{h_i \in H_i : g_i^{\mathcal{T}}(h_i) \in E\}$. Fix $E \in \mathcal{Q}_i^n$. By Lemma 6.6, there is $B_i^n \in \mathcal{F}_{i,n}^n$ such that

$$\begin{aligned} \left\{h_i \in H_i : g_i^{\mathcal{T}}(h_i) \in E\right\} &= \left\{h_i \in H_i : g_i^{\mathcal{T}}(h_i) \in \{t_i \in T_i : h_i^{\mathcal{T},n}(t_i) \in B_i^n\}\right\} \\ &= \{(\mu_i^1, \mu_i^2, \ldots) \in H_i : (\mu_i^1, \ldots, \mu_i^n) \in B_i^n\}, \end{aligned}$$

and $\{(\mu_i^1, \mu_i^2, \ldots) \in H_i : (\mu_i^1, \ldots, \mu_i^n) \in B_i^n\} \in \mathcal{F}_{i,n}$. That $h_i^{\mathcal{T}}$ is an isomorphism with respect to $\mathcal{Q}_i^\infty$ and $\mathcal{F}_{i,\infty}$ follows from the fact that $\mathcal{Q}_i^\infty$ and $\mathcal{F}_{i,\infty}$ are generated by $\mathcal{Q}_i^n$, $n < \infty$, and $\mathcal{F}_{i,n}$, $n < \infty$, respectively (e.g., Aliprantis and Border, 2005, Coroll. 4.24). □

In words, Theorem 6.1 shows that the "language" associated with types is as expressive as the "language" associated with belief hierarchies: for every type space, there is a model that induces the same belief hierarchies; and, conversely, for every model, there is a type space that generates precisely the belief hierarchies in the model. That the hierarchy mappings are isomorphisms implies that every event that can be expressed in terms of players' types can be expressed in terms of players' belief hierarchies and vice versa.

Theorem 6.1 immediately proves Corollary 6.2:

Corollary 6.2. Every belief hierarchy corresponds to a type and vice versa. That is,

(a) For every type t_i in a type space $\mathcal{T}$, there is a model $\mathcal{H}$ and a belief hierarchy h_i in $\mathcal{H}$ such that h_i is precisely the belief hierarchy $h_i^{\mathcal{T}}(t_i)$ induced by t_i.

(b) For every belief hierarchy h_i, there is a type space $\mathcal{T}$ and a type in $\mathcal{T}$ such that the belief hierarchy $h_i^{\mathcal{T}}(t_i)$ generated by t_i is precisely h_i.

7 A universal type space?

7.1 A negative result

This section shows that if players can have a finite depth of reasoning, then there is no universal type space, unlike in the Harsanyi case. For the remainder of this paper, I restrict attention to nonredundant type spaces; so, in the remainder of the paper, every type space is nonredundant. Throughout this section, $\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$ and $\widetilde{\mathcal{T}} = (\widetilde{T}_a, \widetilde{T}_b, \widetilde{\mathscr{F}}_a, \widetilde{\mathscr{F}}_b, \widetilde{\pi}_a, \widetilde{\pi}_b)$ are two (S_a, S_b)-based type spaces.

The definition of a universal type space uses type morphisms, which are defined as follows:

Definition 7.1. [Mertens and Zamir, 1985, Def. 2.6] A *type morphism* from $\mathcal{T}$ to $\widetilde{\mathcal{T}}$ is a pair $\varphi = (\varphi_a, \varphi_b)$ of functions, where for each player i, φ_i is a function from T_i to $\widetilde{T}_i$ that satisfies the following:

(a) φ_i *preserves depth (of reasoning)*: for every type $t_i \in T_i$, $\varphi_i(t_i) \in \widetilde{T}_i$ and t_i have the same depth of reasoning;

(b) φ_i *preserves beliefs*: for every type $t_i \in T_i$ and every $E \in \overline{\Sigma}_{\varphi_i(t_i)}$, we have $(\mathrm{Id}_{S_{-i}}, \varphi_{-i})^{-1}(E) \in \overline{\Sigma}_{t_i}$ and

$$\widetilde{\pi}_{\varphi_i(t_i)}(E) = \pi_{t_i} \circ (\mathrm{Id}_{S_{-i}}, \varphi_{-i})^{-1}(E).$$

Definition 7.2. [Mertens and Zamir, 1985, Def. 2.15] A type space $\mathcal{T}$ can be *embedded in a type space* $\widetilde{\mathcal{T}}$ *as a belief-closed subset* if there is a type morphism φ from $\mathcal{T}$ to $\widetilde{\mathcal{T}}$ and for every type $t_i \in T_i$, $\widetilde{\pi}_{\varphi_i(t_i)}$ has support in $S_{-i} \times \varphi_{-i}(T_{-i})$.

Definition 7.3. [**Universal type space**] Fix a class $\mathscr{C}$ of type spaces. Then, a type space $\widetilde{\mathcal{T}} \in \mathscr{C}$ is *universal for* $\mathscr{C}$ if every $\mathcal{T} \in \mathscr{C}$ can be embedded in $\widetilde{\mathcal{T}}$ as a belief-closed subset.

So, universality imposes two requirements: given a class $\mathscr{C}$ of type spaces, a type space $\widetilde{\mathcal{T}}$ is universal for $\mathscr{C}$ if for every type space $\mathcal{T} \in \mathscr{C}$, there is a type morphism from $\mathcal{T}$ to $\widetilde{\mathcal{T}}$, *and* this type morphism embeds $\mathcal{T}$ into $\widetilde{\mathcal{T}}$ as a belief-closed subset.[15] In the case of (nonredundant) Harsanyi type spaces, the former implies the latter under appropriate assumptions on the measurable structure (Mertens and Zamir, 1985, Thm. 2.9.5; see Battigalli and Friedenberg, 2009, App. A for a detailed discussion).

As is well-known, a universal Harsanyi type space exists. To state the result, let $\mathscr{C}^*$ be the class of (nonredundant) type spaces, and let $\mathscr{C}^h \subsetneq \mathscr{C}^*$ be the class of (nonredundant) Harsanyi type spaces.

Proposition 7.4. [**Existence universal Harsanyi type space (Mertens and Zamir, 1985, Thm. 2.9(5))**] The canonical Harsanyi type space $\mathcal{T}^h$ is universal for the class $\mathscr{C}^h$ of Harsanyi type spaces.

Proof. The proof is standard. Fix a nonredundant Harsanyi type space $\mathcal{T} = (T_a, T_b, \mathcal{F}_a, \mathcal{F}_b, \pi_a, \pi_b)$. It is straightforward to define mappings $\tilde{h}_i^{\mathcal{T}}$, $i = a, b$, that map the types in $\mathcal{T}$ into a belief hierarchy in $\mathcal{H}^h$ analogously to the hierarchy mappings $h_i^{\mathcal{T}}$ defined in Section 6. Then, $\varphi_i := \tilde{h}_i^{\mathcal{T}}$

[15] This definition is equivalent to defining the universal type space to be the type space that generate all belief hierarchies for a given class of models (proof available upon request).

is one-to-one and measurable. Moreover, the canonical Harsanyi model $\mathcal{H}^h$ defines a Harsanyi type space $\mathcal{T}^h = (T_a^h, T_b^h, \mathcal{F}_a^h, \mathcal{F}_b^h, \pi_a^h, \pi_b^h)$ (Corollary 6.15) and $\varphi := (\varphi_a, \varphi_b)$ is a type morphism from $\mathcal{T}$ to $\mathcal{T}^h$. It remains to show that for each type $t_i \in T_i$, the belief associated with $\varphi_i(t_i)$ has support in $S_{-i} \times \varphi_{-i}(T_{-i})$. Fix a type $t_i \in T_i$. By Theorem 6.1 of Mackey (1957), there is $B \in \mathcal{F}_{-i}$ such that $\pi_{t_i}(S_{-i} \times B) = 1$ and B endowed with its relative σ-algebra is a standard Borel space. I claim that $\varphi_{-i}(B)$ is measurable for $\varphi_i(t_i)$ and $\pi^h_{\varphi_i(t_i)}(S_{-i} \times \varphi_{-i}(B)) = 1$. The former claim follows from Theorem 3.2 of Mackey (1957). The latter claim follows from the fact that φ_a and φ_b preserve beliefs. □

However, the analogue of Proposition 7.4 for the class $\mathscr{C}^*$ of all type spaces does not hold: the canonical type space $\mathcal{T}^*$ is not universal for $\mathscr{C}^*$. In fact, *no* type space is universal for $\mathscr{C}^*$ if there is nontrivial primitive uncertainty:

Theorem 7.5. [Non-existence universal type space] Suppose $|S_a|, |S_b| \geq 2$. Then, there is no universal type space for $\mathscr{C}^*$.

Proof. Suppose by contradiction that $\mathcal{T}^U = (T_a^u, T_b^u, \mathscr{F}_a^u, \mathscr{F}_b^u, \pi_a^u, \pi_b^u)$ is a universal type space for $\mathscr{C}^*$, and fix $s_i^1, s_i^2 \in S_i$ for $i = a, b$. I show that there are two type spaces that cannot both be embedded in $\mathcal{T}^U$ as a belief-closed subset, a contradiction. For ease of notation, I write $\mu(E)$ for $\text{marg}_X \mu(E)$ if μ is a belief on $X \times Y$ and E is a measurable subset of X.

Consider the type space $\mathcal{T}$ defined as follows: for each player $i = a, b$, $T_i = \{t_i^1, t_i^2, t_i^3, t_i^4\}$, and for each $t_i \in T_i$, Σ_{t_i} is the coarsest σ-algebra that contains the sets $\{t_{-i}^1, t_{-i}^2\}$ and $\{t_{-i}^3, t_{-i}^4\}$. Beliefs are given by:

$$\begin{aligned}
\pi_{t_i^1}(s_{-i}^1, \{t_{-i}^1, t_{-i}^2\}) &= \tfrac{1}{2} & \qquad \pi_{t_i^1}(s_{-i}^2, \{t_{-i}^1, t_{-i}^2\}) &= \tfrac{1}{2};\\
\pi_{t_i^2}(s_{-i}^1, \{t_{-i}^3, t_{-i}^4\}) &= \tfrac{1}{2} & \qquad \pi_{t_i^2}(s_{-i}^2, \{t_{-i}^3, t_{-i}^4\}) &= \tfrac{1}{2};\\
\pi_{t_i^3}(s_{-i}^1, \{t_{-i}^1, t_{-i}^2\}) &= \tfrac{1}{4} & \qquad \pi_{t_i^3}(s_{-i}^2, \{t_{-i}^1, t_{-i}^2\}) &= \tfrac{3}{4};\\
\pi_{t_i^4}(s_{-i}^1, \{t_{-i}^3, t_{-i}^4\}) &= \tfrac{1}{4} & \qquad \pi_{t_i^4}(s_{-i}^2, \{t_{-i}^3, t_{-i}^4\}) &= \tfrac{3}{4};
\end{aligned}$$

where the notation is as in Section 2. For example, type t_a^1 for Ann assigns equal probability to s_b^1 and s_b^2, and believes that Bob assigns equal probability to s_a^1 and s_a^2. It is easy to check that all types in $\mathcal{T}$ have depth 2. Clearly, $\mathcal{T} \in \mathscr{C}^*$. Let $\varphi = (\varphi_a, \varphi_b)$ be a type morphism from $\mathcal{T}$ into $\mathcal{T}^U$.

Also, consider a type space $\mathcal{T}' = (T_a', T_b', \mathscr{F}_a', \mathscr{F}_b', \pi_a', \pi_b') \in \mathscr{C}^*$ that includes a type $\tilde{t}_b \in T_b'$ whose belief satisfies:

$$\begin{aligned}
\pi'_{\tilde{t}_b}(\{s_a^1\} \times \{t_a \in T_a' : \pi'_{t_a}(s_b^1) = 1\}) &= \tfrac{1}{2};\\
\pi'_{\tilde{t}_b}(\{s_a^2\} \times \{t_a \in T_a' : \pi'_{t_a}(s_b^1) = 1\}) &= \tfrac{1}{2}.
\end{aligned}$$

That is, $\tilde{t}_b$ is a type that assigns equal probability to s_a^1 and s_a^2 and that believes that Ann assigns probability 1 to s_b^1. (Such a type space exists: for example, take $\mathcal{T}' = \mathcal{T}^*$ or $\mathcal{T}' = \mathcal{T}^h$.) Importantly, type $\tilde{t}_b$ has a different belief about Ann's belief about S_b than the types in $\mathcal{T}$: type $\tilde{t}_b$ believes that Ann assigns probability 1 to s_b^1 while the types in $\mathcal{T}$ assign positive probability to both s_a^1 and s_a^2. Let $\varphi' = (\varphi'_a, \varphi'_b)$ be a type morphism from $\mathcal{T}'$ to $\mathcal{T}^U$.

Let

$$B^u := \{t_b^u \in T_b^u : \pi_{t_b^u}^u(s_a^1) = \tfrac{1}{2}, \pi_{t_b^u}^u(s_a^2) = \tfrac{1}{2}\}.$$

be the set of types for Bob in $\mathcal{T}^U$ that assign equal probability to s_a^1 and s_a^2. Since the type morphism φ from $\mathcal{T}$ to $\mathcal{T}^U$ preserves depth, $\mathcal{T}^U$ includes types of depth 2. Denote the σ-algebra on T_i^u associated with a type $t_{-i}^u \in T_{-i}^u$ of depth 2 by $\mathcal{F}_{i,1}^u$.

Since φ embeds $\mathcal{T}$ in $\mathcal{T}^U$ as a belief-closed subset, there is $E_{t_a^1} \in \mathcal{F}_{b,1}^u$ such that $E_{t_a^1} \subset \varphi_b(T_b)$ and the image $\varphi_a(t_a^1)$ of t_a^1 in $\mathcal{T}^U$ assigns probability 1 to $E_{t_a^1}$ (i.e., $\pi_{\varphi_a(t_a^1)}^u(E_{t_a^1}) = 1$). By Lemma B.7 in the appendix,

$$B^u \subset E_{t_a^1}. \tag{7.1}$$

In words, type t_a^1 is certain that Bob assigns equal probability to s_a^1 and s_a^2.

I claim that there is a type in B^u that is not in $\varphi_b(T_b)$. This contradicts (7.1), given that $E_{t_a^1} \subset \varphi_b(T_b)$. To show this, let $t_b^u := \varphi'_b(\tilde{t}_b)$ be the image of $\tilde{t}_b$ in $\mathcal{T}^U$. As φ' preserves beliefs, $\pi_{t_b^u}^u(s_a^1) = \frac{1}{2}$ and $\pi_{t_b^u}^u(s_a^2) = \frac{1}{2}$. So, $t_b^u \in B^u$. I show that $t_b^u \notin \varphi_b(T_b)$. If the depth of reasoning of t_b^u is not equal to 2, then we are done, since φ preserves depth. So, suppose that t_b^u has depth 2. Since φ preserves beliefs, $t_b^u \neq \varphi_b(t_b^3), \varphi_b(t_b^4)$. So, it suffices to show that $t_b^u \neq \varphi_b(t_b^1), \varphi_b(t_b^2)$. I present the argument for t_b^1; the proof for t_b^2 is similar and thus omitted.

Lemma B.8 in the appendix uses that φ' preserves beliefs to show that

$$\pi_{t_b^u}^u(\{t_a \in T_a^u : \pi_{t_a^u}^u(s_b^1) = 1\}) = 1.$$

Then, since φ preserves beliefs,

$$(\varphi_a)^{-1}(\{t_a^u \in T_a^u : \pi_{t_a^u}^u(s_b^1) = 1\}) \supset \{t_a \in T_a : \pi_{t_a}(s_b^1) = \tfrac{1}{2}, \pi_{t_a}(s_b^2) = \tfrac{1}{2}\}. \tag{7.2}$$

But,

$$(\varphi_a)^{-1}(\{t_a^u \in T_a^u : \pi_{t_a^u}^u(s_b^1) = 1\}) = \{t_a \in T_a : \pi_{t_a}(s_b^1) = 1\}.$$

So, (7.2) is equivalent to saying that any type in T_a that assigns equal probability to s_a^1 and s_a^2 assigns probability 1 to s_b^1, a contradiction. □

Theorem 7.5 shows that there is no type space that can simultaneously model all restrictions on players' beliefs. The basic insight is simple. If a type space does not rule out certain beliefs

for the players, then the type space cannot model strategic situations where players do rule out these beliefs. For example, the type space $\mathcal{T}$ in the proof describes a strategic situation where it is ruled out that Bob is certain that Ann believes that Bob has $s_b = s_b^1$. In particular, the type space $\mathcal{T}$ implicitly assumes that Ann rules out that Bob believes that Ann believes that Bob has $s_b = s_b^1$. However, in a (candidate) universal type space, this cannot be ruled out since a universal type space must also accommodate beliefs such as those of the type $\tilde{t}_b$ in $\mathcal{T}'$.

This issue does not arise with Harsanyi type spaces: different Harsanyi type spaces, capturing different information structures, can be contained in a larger (Harsanyi) type space: the universal Harsanyi type space $\mathcal{T}^h$ is simply the union of all Harsanyi type spaces. In essence, when players have a infinite depth, then their "language" (or subjective state space, in the terminology of Section 2) is sufficiently fine so that they can simply assign probability 0 to events that are ruled out, giving the positive result for the class of Harsanyi type spaces (Proposition 7.4). By contrast, when players have a finite depth of reasoning, then their "language" is too coarse to assign zero probability to events that are ruled out, and the union of all type space is not a type space. This leads to the negative result Theorem 7.5 for the class of all type spaces.

Theorem 7.5 does not rely on any special assumptions. For example, it does not require that a "candidate" universal space has any special properties (e.g., is canonical). Instead, it applies to all (nonredundant) type spaces. Moreover, it is not hard to see that a negative result obtains for any subclass $\mathscr{C} \subset \mathscr{C}^*$ of type spaces that contains a type space that is not a Harsanyi type space. So, it is not possible to obtain a positive result by excluding certain "pathological" type spaces from $\mathscr{C}^*$. In particular, while the proof uses types with depth 2, a similar proof applies for types of any finite depth.

7.2 Rationalizable behavior across type spaces

This section shows that if an analyst who is interested in studying rationalizable behavior across all strategic situations needs to consider all type spaces. Specifically, I show by example that if there is type morphism $\varphi = (\varphi_a, \varphi_b)$ from a type space $\mathcal{T}$ to a type space $\widetilde{\mathcal{T}}$, but it does not preserve higher-order beliefs, then there can be types t_i in $\mathcal{T}$ such that t_i and its image $\varphi_i(t_i)$ under φ have different rationalizable actions. Together with Theorem 7.5, this implies that if players can have a finite depth of reasoning, then there is no single type space that an analyst can use to study rationalizable behavior across all type spaces.

I start with some definitions. Each player $i = a, b$ has a finite set A_i of actions. The primitive uncertainty is given by a common attribute $s \in S$, where S is a finite set (i.e., $S_a = S_b = S$). The payoff $u_i(\alpha_i, \alpha_{-i}, s)$ to a player $i = a, b$ depend on his own action $\alpha_i \in A_i$, the action $\alpha_{-i} \in A_{-i}$ of the other player, and the common attribute. Given a S-based type space

$\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$, a *conjecture* for type $t_i \in T_i$ is a function $\sigma_{-i} : S \times T_{-i} \to \Delta(A_{-i})$ that is $\overline{\Sigma}_{t_i}$-measurable.

For the purpose of this section, it suffices to consider finite type spaces, that is, type spaces in which each player has finitely many types. In that case, the σ-algebra $\overline{\Sigma}_{t_i}$ associated with a type can be represented by its subjective state space Ω_{t_i} (Section 2), which can be viewed as a partition Π_{t_i} of $S_{-i} \times T_{-i}$. Then, given a conjecture σ_{-i}, the interim expected utility for a type t_i of action $\alpha_i \in A_i$ is

$$V_{t_i}(\alpha_i, \sigma_{-i}) := \sum_{Q_{-i} \in \Pi_{t_i}} \pi_{t_i}(Q_{-i}) \cdot \sum_{(s,t_{-i}) \in Q_{-i}} u_i(\alpha_i, \sigma_{-i}(s, t_{-i}), s).$$

These definitions coincide with the standard definitions in the Harsanyi case.

I focus on (interim correlated) rationalizability (Dekel, Fudenberg, and Morris, 2007). Given a finite type space $\mathcal{T} = (T_a, T_b, \mathscr{F}_a, \mathscr{F}_b, \pi_a, \pi_b)$, for each player $i = a, b$ and type $t_i \in T_i$, define $R_i^{\mathcal{T},0}(t_i) := A_i$. For $m > 0$, define

$$R_i^{\mathcal{T},m}(t_i) := \left\{ \alpha_i \in A_i : \begin{array}{cc} & \text{there is } \sigma_{-i} : S_{-i} \times T_{-i} \to \Delta(A_{-i}) \text{ s.t.} \\ (1) & \sigma_{-i} \text{ is measurable w.r.t. } \overline{\Sigma}_{t_i}; \\ (2) & \sigma_{-i}(s, t_{-i})(\alpha_{-i}) > 0 \text{ implies that } \alpha_{-i} \in R_j^{\mathcal{T},m-1}(t_{-i}); \\ (3) & \alpha_i \in \arg\max_{\alpha_i'} V_{t_i}(\alpha_i', \sigma_{-i}); \end{array} \right\}.$$

to be the set of best replies for t_i to the $(m-1)$th-order rationalizable actions of the other player. The set of rationalizable actions for t_i is then $R_i^{\mathcal{T}}(t_i) := \bigcap_m R_i^{\mathcal{T},m}(t_i)$. In the Harsanyi case, this definition is precisely the standard definition (Dekel, Fudenberg, and Morris, 2007). The requirement that conjectures be measurable for the type ensures (i.e., (1) in the definition of $R_i^{\mathcal{T},m}(t_i)$ above) ensures that the type's expected payoff is well-defined

If a type morphism $\varphi = (\varphi_a, \varphi_b)$ from a type space $\mathcal{T}$ to a type space $\widetilde{\mathcal{T}}$ preserves higher-order beliefs, then for any type t_i in $\mathcal{T}$, t_i and its image $\varphi_i(t_i)$ have the same rationalizable actions (Dekel, Fudenberg, and Morris, 2007, Lemma 1). However, the next example demonstrates that if φ does not preserve higher-order beliefs, then this need not be the case: if there is type morphism $\varphi = (\varphi_a, \varphi_b)$ from a type space $\mathcal{T}$ to a type space $\widetilde{\mathcal{T}}$, but the image of $\mathcal{T}$ is not a belief-closed subset of $\widetilde{\mathcal{T}}$, then there can be types t_i in $\mathcal{T}$ such that t_i and its image $\varphi_i(t_i)$ under φ have different rationalizable actions.

Example 3. Consider the following game:

s^1

	α_1	α_2	α_3
α_1	1,1	1,0	0,0
α_2	0,1	0,0	0,0
α_3	0,0	0,0	1,1

s^2

	α_1	α_2	α_3
α_1	0,0	0,1	0,0
α_2	1,0	1,1	0,0
α_3	0,0	0,0	1,1

s^3

	α_1	α_2	α_3
α_1	1,1	0,0	0,0
α_2	0,0	1,1	0,0
α_3	0,0	0,0	1,1

Clearly, for any type space, action α_3 is rationalizable for every type. I show that for other actions, it may depend on the precise specification of the information structure whether the action is rationalizable.

First consider the following type space, denoted $\mathcal{T}$: The type sets for Ann and Bob are $T_a = \{t_a^1, t_a^2, t_a^3\}$ and $T_b = \{t_b^1, t_b^2\}$, respectively. The types for Ann are endowed with the trivial σ-algebra (i.e., for $t_a \in T_a$, $\Sigma_{t_a} = \{T_b, \emptyset\}$). Type t_a^1 assigns probability 1 to (s^1, T_b), type t_a^2 assigns probability 1 to (s^2, T_b), and t_a^3 assigns probability 1 to (s^3, T_b) (where again the notation is as in Section 2). The types for Bob are endowed with the finest σ-algebra (i.e., for $t_b \in T_b$, Σ_{t_b} contains all subsets of T_a). Type t_b^1 assigns probability 1 to (s^1, t_a^1) and type t_b^2 assigns probability 1 to (s^3, t_a^1). Then, the types for Ann have depth 1, and the types for Bob have an infinite depth. For this type space, action α_1 is rationalizable for type t_a^3 under the conjecture that t_b^1 and t_b^2 choose α_1.

Next consider the following type space, denoted $\widetilde{\mathcal{T}}$: The type sets for Ann and Bob are $\widetilde{T}_a = \{\tilde{t}_a^1, \tilde{t}_a^2, \tilde{t}_a^3\}$ and $\widetilde{T}_b = \{\tilde{t}_b^1, \tilde{t}_b^2, \tilde{t}_b^3, \tilde{t}_b^4\}$, respectively. As before, the types for Ann are endowed with the trivial σ-algebra (i.e., for $\tilde{t}_a \in \widetilde{T}_a$, $\Sigma_{\tilde{t}_a} = \{\widetilde{T}_b, \emptyset\}$); and type $\tilde{t}_a^1$ assigns probability 1 to $(s^1, \widetilde{T}_b)$, type $\tilde{t}_a^2$ assigns probability 1 to $(s^2, \widetilde{T}_b)$, and $\tilde{t}_a^3$ assigns probability 1 to $(s^3, \widetilde{T}_b)$. Again, the types for Bob are endowed with the finest σ-algebra (i.e., for $\tilde{t}_b \in \widetilde{T}_b$, $\Sigma_{\tilde{t}_b}$ contains all subsets of $\widetilde{T}_a$); and type $\tilde{t}_b^1$ assigns probability 1 to $(s^1, \tilde{t}_a^1)$ and type $\tilde{t}_b^2$ assigns probability 1 to $(s^3, \tilde{t}_a^1)$. Type $\tilde{t}_b^3$ assigns probability 1 to (s^2, t_a^2), and type $\tilde{t}_b^4$ assigns probability $\frac{1}{2} : \frac{1}{2}$ to $(s^2, t_a^2) : (s^3, t_a^2)$. Again, the types for Ann have depth 1, and the types for Bob have an infinite depth.

It is easy to see that $\varphi = (\varphi_a, \varphi_b)$ with $\varphi_i(t_i^m) = \tilde{t}_i^m$, $i = a, b$, is a type morphism that embeds $\mathcal{T}$ into $\widetilde{\mathcal{T}}$. However, the image of $\mathcal{T}$ in $\widetilde{\mathcal{T}}$ under φ does not form a belief-closed subset. For example, types t_a^1, t_a^2, t_a^3 rule out that Bob thinks that $s = s^2$; but types $\varphi_a(t_a^1), \varphi_a(t_a^2), \varphi_a(t_a^3)$ do not.

The rationalizability correspondence is not invariant under φ: For $\widetilde{\mathcal{T}}$, there is no conjecture for Ann such that α_1 is rationalizable for $\tilde{t}_a^3 = \varphi_a(t_a^3)$. In particular, $\tilde{t}_a^3$ cannot rule out that Bob chooses an action to which α_1 is not a best response: types $\tilde{t}_b^3$ and $\tilde{t}_b^4$ can rationally choose α_2 or α_3, but not α_1. ◁

The example demonstrates that types can have different rationalizable actions even if one is the image of the other under a type morphism: if the image of a type space under a type morphism does not form a belief-closed subset, then the types in the embedded type space think possible more beliefs than the types in the original type space, and this has implications for the actions that a type can rationalize.[16] Intuitively, if the strategic situation is modeled

[16] In the context of Harsanyi type spaces, Friedenberg and Meier (2016) likewise show that Bayesian-Nash

by $\mathcal{T}$, Ann can rule out that Bob chooses α_2, but not if the situation is modeled by $\widetilde{\mathcal{T}}$. As a result, Ann can rationalize α_1 only if certain beliefs for Bob are ruled out. Again, the example does not hinge on the particular assumptions, such as that types have depth 1 or that players have finitely many types. In particular, the type space $\widetilde{\mathcal{T}}$ can be replaced by the canonical type space $\mathcal{T}^*$. Likewise, a similar example can be applied where Ann's types in $\widetilde{\mathcal{T}}$ can rule out the types for Bob that are not included in the original type space $\mathcal{T}$ but thinks Bob may think possible certain beliefs that are ruled out in $\mathcal{T}$, etcetera.

Theorem 7.5 and Example 3 imply that if players can have a finite depth of reasoning, then there is no type space that can simultaneously model players' rationalizable behavior across all type spaces, unlike in the Harsanyi case, where the canonical Harsanyi type space includes all Harsanyi type spaces as a belief-closed subset. Thus, if an analyst is interested in studying the rationalizable behavior of players with a finite depth of reasoning, then he needs to consider all type spaces. This is natural: if players can have a finite depth of reasoning, then a type space embedded in a larger type space captures a different state of affairs than the original type space, and we have no reason to expect similar behavior in these two situations.

Appendix A Preliminary results

Some preliminary definitions and results will be helpful. A measurable space $(X, \mathcal{F})$ is *separated* if there is $\mathcal{G} \subset \mathcal{F}$ such that for any pair x, y of distinct points in X, there is $F \in \mathcal{G}$ such that $x \in F$, $y \notin F$. Given a collection $\mathcal{E}$ of subsets of$X$, $\sigma(\mathcal{E})$ is the σ-algebra *generated by* $\mathcal{E}$, that is, the coarsest σ-algebra that contains the sets in E. If $\mathcal{F}$ is separated and is generated by a countable collection $\mathcal{E}$, then $\mathcal{F}$ is *countably generated*. A σ-algebra $\mathcal{F}$ that is countably generated is *countably separated*: there is a (countable) subset $\mathcal{E} \subset \mathcal{F}$ such that every $x, y \in X$, $x \neq y$, there is $E \in \mathcal{F}$ such that $x \in E$, $y \notin E$.

The first auxiliary result relates different σ-algebras. To state the result, let X be a nonempty set, and let $\mathscr{S}$ be a nonempty collection of σ-algebras on X. Let $\Delta(X, \mathscr{S}) := \bigcup_{\mathcal{F} \in \mathscr{S}} \Delta(X, \mathcal{F})$ be the collection of probability measures that are defined on some σ-algebra in $\mathscr{S}$. Let $\mathcal{A}$ be the family of sets of the form

$$\{\mu \in \Delta(X, \mathscr{S}) : \Sigma(\mu) = \mathcal{F}, \mu(E) \geq p\} : \qquad \mathcal{F} \in \mathscr{S}, E \in \mathcal{F}, p \in [0, 1],$$

equilibrium need not be invariant under type morphisms. An important difference is that in the present context, predictions fail to be invariant under a type morphism φ only if it does not preserve higher-order beliefs (i.e., the image of φ does not form a belief-closed subset). By contrast, in the environments that Friedenberg and Meier consider, type morphisms preserve higher-order beliefs.

and let $\mathcal{A}'$ be the family of sets of the form

$$\{\mu \in \Delta(X, \mathscr{S}) : E \in \Sigma(\mu), \mu(E) \geq p\} : \qquad \mathcal{F} \in \mathscr{S}, E \in \mathcal{F}, p \in [0,1].$$

In general, the σ-algebras $\sigma(\mathcal{A})$ and $\sigma(\mathcal{A}')$ generated by $\mathcal{A}$ and $\mathcal{A}'$ may be different. However, the next result shows that they coincide in an important class of cases:

Lemma A.1. Suppose $\mathscr{S}$ is countable and forms a filtration, and suppose $\mathscr{S}$ has a minimal element.[17] Then $\sigma(\mathcal{A}) = \sigma(\mathcal{A}')$.

Proof. I first show that $\sigma(\mathcal{A}') \subset \sigma(\mathcal{A})$. It suffices to show that $\mathcal{A}' \subset \sigma(\mathcal{A})$. Fix $\mathcal{F} \in \mathscr{S}$, $E \in \mathcal{F}$, and $p \in [0,1]$, and define

$$F' := \{\mu \in \Delta(X, \mathscr{S}) : E \in \Sigma(\mu), \mu(E) \geq p\},$$

so that $F' \in \mathcal{A}'$. It is immediate that $F' \in \sigma(\mathcal{A})$: Since for every $\mathcal{F}' \in \mathscr{S}$, either $E \in \mathcal{F}'$ or $E \notin \mathcal{F}'$, F' is a countable union of sets in $\mathcal{A}$:

$$F' = \bigcup_{\mathcal{F}' \in \mathscr{S} : E \in \mathcal{F}'} \{\mu \in \Delta(X, \mathscr{S}) : \Sigma(\mu) = \mathcal{F}', \mu(E) \geq p\}.$$

Hence, $F' \in \sigma(\mathcal{A})$.

I next show the reverse inclusion, that is, $\sigma(\mathcal{A}) \subset \sigma(\mathcal{A}')$. Again, fix $\mathcal{F} \in \mathscr{S}$, $E \in \mathcal{F}$, and $p \in [0,1]$, and define

$$F := \{\mu \in \Delta(X, \mathscr{S}) : \Sigma(\mu) = \mathcal{F}, \mu(E) \geq p\},$$

so that $F \in \mathcal{A}$. If we show that $\Delta(X, \mathcal{F})$ is an element of $\sigma(\mathcal{A}')$, then we are done, because F is then the intersection of two elements of $\sigma(\mathcal{A}')$:

$$F = \{\mu \in \Delta(X, \mathscr{S}) : E \in \Sigma(\mu), \mu(E) \geq p\} \cap \Delta(X, \mathcal{F}).$$

It remains to show that $\Delta(X, \mathcal{F}) \in \sigma(\mathcal{A}')$. Using that $\mathscr{S}$ is a countable filtration with a minimum element $\underline{\mathcal{F}}$, the σ-algebras in $\mathscr{S}$ can be labeled as

$$\underline{\mathcal{F}} =: \mathcal{F}_1 \subsetneq \mathcal{F}_2 \subsetneq \cdots$$

Then,

$$\Delta(X, \mathcal{F}_1) = \Delta(X, \mathscr{S}) \setminus \{\mu \in \Delta(X, \mathscr{S}) : E_2 \in \Sigma(\mu), \mu(E_2) \geq 0\}$$

for any $E_2 \in \mathcal{F}_2 \setminus \mathcal{F}_1$, so $\Delta(X, \mathcal{F}_1) \in \sigma(\mathcal{A}')$. For $k > 1$, assume that $\Delta(X, \mathcal{F}_1), \ldots, \Delta(X, \mathcal{F}_{k-1}) \in \sigma(\mathcal{A}')$. Then,

$$\Delta(X, \mathcal{F}_k) = \Delta(X, \mathscr{S}) \setminus \Big(\{\mu \in \Delta(X, \mathscr{S}) : E_{k+1} \in \Sigma(\mu), \mu(E_{k+1}) \geq 0\} \cup \Delta(X, \mathcal{F}_1) \cup \cdots \cup \Delta(X, \mathcal{F}_{k-1})\Big)$$

[17] That is, there is $\underline{\mathcal{F}} \in \mathscr{S}$ such that $\underline{\mathcal{F}} \subset \mathcal{F}$ for all $\mathcal{F} \in \mathscr{S}$.

for any $E_{k+1} \in \mathcal{F}_{k+1} \setminus \mathcal{F}_k$, so $\Delta(X, \mathcal{F}_k) \in \sigma(\mathcal{A}')$. Since this holds for every k, and $\mathcal{F} = \mathcal{F}^k$ for some k, the event $\Delta(X, \mathcal{F})$ belongs to $\sigma(\mathcal{A}')$. □

I next prove the claim in Section 4.1 that every set of $(m-1)$th-order belief hierarchies can be extended to a set of mth-order belief hierarchies:

Proposition A.2. Suppose H_a^{m-1} and H_b^{m-1} satisfy conditions (IND), (COH), (EXT),and (ANL_H). Then there exist H_a^m and H_b^m that extend H_a^{m-1} and H_b^{m-1}, respectively, and that satisfy (IND), (COH), (EXT),and (ANL_H).

Proof. I first show that every $(m-1)$th-order belief hierarchy can be extended to an mth-order belief hierarchy:

Lemma A.3. For every $(m-1)$th-order belief hierarchy $(\mu_i^1, \ldots, \mu_i^{m-1}) \in H_i^{m-1}$, there is an mth-order belief $\mu_i^m \in \Delta^+(S_{-i} \times H_{-i}^{m-1})$ so that the resulting mth-order belief hierarchy $(\mu_i^1, \ldots, \mu_i^m)$ satisfies (COH).

Proof. Fix $i = a, b$, and let $(\mu_i^1, \ldots, \mu_i^{m-1}) \in H_i^{m-1}$. I will define a function η_i^m from H_i^{m-1} into $H_i^{m-1} \times \Delta^+(S_{-i} \times H_{-i}^{m-1})$, and use that to show that there is a nonempty set $H_i^m \subset H_i^{m-1} \times \Delta^+(S_{-i} \times H_{-i}^{m-1})$ that extends H_i^{m-1} in the appropriate sense. It will be convenient to write $\overline{\mathcal{F}}_{-i,0}^0$ for $\mathcal{F}_{S_{-i}}$. Also, recall that Id_X is the identity function on X, and denote the projection function from H_i^k to H_i^m, $m < k$, by $\text{proj}_{i,m}^k$.

First suppose $\Sigma(\mu_i^{m-1}) = \overline{\mathcal{F}}_{-i,\ell}^{m-2}$ for some $\ell < m-2$. It is easy to check that there is a unique belief $\mu_i^m \in \Delta(S_{-i} \times H_{-i}^{m-1}, \overline{\mathcal{F}}_{-i,\ell}^{m-1})$ such that $\text{marg}_{S_{-i} \times H_{-i}^{m-2}} \mu_i^m = \mu_i^{m-1}$.[18] Define $\eta_i^m(\mu_i^1, \ldots, \mu_i^{m-1}) := (\mu_i^1, \ldots, \mu_i^m)$.

Next suppose that $\Sigma(\mu_i^{m-1}) = \overline{\mathcal{F}}_{-i,m-2}^{m-2}$. I prove the result for $m > 2$; the proof for $m = 2$ is similar, and thus omitted. I define a belief μ_i^m in $\Delta(S_{-i} \times H_{-i}^{m-1}, \overline{\mathcal{F}}_{-i,m-1}^{m-1})$ in such a way that the beliefs μ_i^{m-1} and μ_i^m are coherent. By Corollary 6 of Lubin (1974), there is a belief $\mu_i^m \in \Delta(S_{-i} \times H_{-i}^{m-1}, \overline{\mathcal{F}}_{-i,m-1}^{m-1})$ and a function $g_{-i}^{m-1} : S_{-i} \times H_{-i}^{m-2} \to S_{-i} \times H_{-i}^{m-1}$ such that

- $(\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1}) \circ g_{-i}^{m-1}$ is the identity function on $S_{-i} \times H_{-i}^{m-2}$;
- for each $E \in \overline{\mathcal{F}}_{-i,m-1}^{m-1}$,

$$(\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1})(E \cap g_{-i}^{m-1}(S_{-i} \times H_{-i}^{m-2})) \in \overline{\mathcal{F}}_{-i,m-2}^{m-2};$$

and

$$\mu_i^m(E) = \mu_i^{m-1}\big((\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1})(E \cap g_{-i}^{m-1}(S_{-i} \times H_{-i}^{m-2}))\big).$$

[18] This belief is defined by $\mu_i^m(E) = \mu_i^{m-1}((\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1})(E))$ for $E \in \overline{\mathcal{F}}_{-i,\ell}^{m-1}$.

I claim that the belief thus defined is coherent, that is, $\text{marg}_{S_{-i}\times H_{-i}^{m-2}}\mu_i^m = \mu_i^{m-1}$. To see this, let $E \in \overline{\mathcal{F}}_{-i,m-2}^{m-2}$. Then,

$$\begin{aligned}\text{marg}_{S_{-i}\times H_{-i}^{m-2}}\mu_i^m(E) &= \mu_i^m \circ \left(\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1}\right)^{-1}(E) \\ &= \mu_i^{m-1} \circ \left(g_{-i}^{m-1}\right)^{-1} \circ \left((\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1}\right)^{-1}(E) \\ &= \mu_i^{m-1}(E),\end{aligned}$$

where the last line uses that $(\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1}) \circ g_{-i}^{m-1}$ is the identity. Define $\eta_i^m(\mu_i^1, \ldots, \mu_i^{m-1}) := (\mu_i^1, \ldots, \mu_i^m)$. □

We are now ready to prove Proposition A.2. Let $i = a, b$, and let η_i^m be as above. Define $H_i^m := \eta_i^m(H_i^{m-1})$, and let $\mathcal{F}_{i,m}^m$ be the relative σ-algebra on $H_i^m \subset H_i^{m-1} \times \Delta^+(S_{-i} \times H_{-i}^{m-1})$. Clearly, $(H_i^m, \mathcal{F}_{i,m}^m)$ satisfies conditions (IND), (COH) and (EXT). It remains to show that Condition (ANL_H) is satisfied. This follows from Theorem 4.2 of Mackey (1957) if we show that $\mathcal{F}_{i,m}^m$ is countably generated and that η_i^m is one-to-one and measurable with respect to $\mathcal{F}_{i,m-1}^{m-1}$ and $\mathcal{F}_{i,m}^m$. By Lemma B.6, $\mathcal{F}_{i,m}^m$ is countably generated. Also, η_i^m is clearly one-to-one. It remains to show that η_i^m is $(\mathcal{F}_{i,m-1}^{m-1}, \mathcal{F}_{i,m}^m)$-measurable. To show this, let $E \in \overline{\mathcal{F}}_{-i,\ell}^{m-1}$ for some $\ell < m-1$ and $p \in [0,1]$. Then,

$$\begin{aligned}(\eta_i^m)^{-1}\big(\{(\mu_i^1, \ldots, \mu_i^m) \in H_i^m : \Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,\ell}^{m-1}, \mu_i^m(E) \geq p\}\big) = \\ \{(\mu_i^1, \ldots, \mu_i^{m-1}) \in H_i^{m-1} : \Sigma(\mu_i^{m-1}) = \overline{\mathcal{F}}_{-i,\ell}^{m-2}, \mu_i^{m-1}(E') \geq p\},\end{aligned}$$

where $E' = (\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-2})(E) \in \overline{\mathcal{F}}_{-i,\ell}^{m-2}$. Clearly, this set is in $\mathcal{F}_{-i,m-1}^{m-1}$. Next, let $E \in \overline{\mathcal{F}}_{-i,m-1}^{m-1}$ and $p \in [0,1]$. Then,

$$\begin{aligned}&(\eta_i^m)^{-1}\big(\{(\mu_i^1, \ldots, \mu_i^m) \in H_i^m : \Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,m-1}^{m-1}, \mu_i^m(E) \geq p\}\big) \\ &= \big\{(\mu_i^1, \ldots, \mu_i^{m-1}) \in H_i^{m-1} : \eta_i^m(\mu_i^1, \ldots, \mu_i^{m-1}) \in \\ &\qquad \{(\mu_i^1, \ldots, \mu_i^m) \in H_i^m : \Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,m-1}^{m-1}, \mu_i^m(E) \geq p\}\big\} \\ &= \big\{(\mu_i^1, \ldots, \mu_i^{m-1}) \in H_i^{m-1} : \Sigma(\mu_i^{m-1}) = \overline{\mathcal{F}}_{-i,m-2}^{m-2}, \\ &\qquad \mu_i^{m-1}\big((\text{Id}_{S_{-i}}, \text{proj}_{-i,m-2}^{m-1})(E \cap g_{-i}^{m-1}(S_{-i} \times H_{-i}^{m-2}))\big) \geq p\big\}.\end{aligned}$$

Again, this set is in $\mathcal{F}_{-i,m-1}^{m-1}$, and the claim follows. So, given a set of $(m-1)$th-order belief hierarchies for each player, there exists sets of mth-order belief hierarchies for both players. □

Appendix B Proofs

B.1 Proof of Lemma 4.1

Suppose that $(H_a^m, \mathcal{F}_{a,m}^m)$ and $(H_b^m, \mathcal{F}_{b,m}^m)$, $m = 1, 2, \ldots$, satisfy (IND)–(ANL$_H$). If H_i^m is nonempty for every m, then H_i is nonempty (Bourbaki, 1970, p. 198). It remains to show that $(H_a, \mathcal{F}_{H_a})$ and $(H_b, \mathcal{F}_{H_b})$ are analytic Borel spaces. Throughout, I write $\mathscr{B}(X)$ for the Borel σ-algebra associated with a topological space X.

The argument proceeds in a number of steps. Step 1 establishes that H_i can be viewed as a Souslin spaces, where a *Souslin space* is a Hausdorff space that is the continuous image of a Polish space (e.g., Bogachev, 2007, Ch. 6). This implies that H_i is the measurable image of an analytic Borel space. I use this in Step 2 to show that H_i is an analytic Borel space.

Step 1: Souslin spaces Fix a player $i = a, b$. I first show that there is a topology τ_i^m on H_i^m such that (H_i^m, τ_i^m) is a Souslin space such that $\mathcal{F}_{i,m}^m$ coincides with the associated Borel σ-algebra $\mathscr{B}(H_i^m)$. First, a straightforward inductive argument shows that there is a topology τ_i^m on H_i^m such that (H_i^m, τ_i^m) is separable metrizable and $\mathscr{B}(H_i^m) = \mathcal{F}_{i,m}^m$.[19] So, in particular, (H_i^m, τ_i^m) is Hausdorff. It thus suffices to show that H_i^m is the continuous image of a Polish space. Since $(H_i^m, \mathcal{F}_{i,m}^m) = (H_i^m, \mathscr{B}(H_i^m))$ is an analytic Borel space (by (ANL$_H$)), it is isomorphic to $(A_i^m, \mathscr{B}(A_i^m)$ for an analytic set A_i^m. Denote the isomorphism from A_i^m to H_i^m by ψ_i^m. So, A_i^m is a subset of a Polish space Y_i^m, and $A_i^m = f_i^m(Z_i^m)$ for a continuous function f_i^m into Y_i^m whose domain is a Polish space Z_i^m. Denote the topology on Z_i^m by $\tau_{Z_i^m}$ By Theorem 13.11 of Kechris (1995), there is a Polish topology $\tau_{Z_i^m}^*$ on Z_i^m such that $\tau_{Z_i^m}^* \supset \tau_{Z_i^m}$ that induces the same Borel σ-algebra as $\tau_{Z_i^m}$ and $\psi_i^m \circ f_i^m$ is a continuous function from $(Z_i^m, \tau_{Z_i^m}^*)$ to (H_i^m, τ_i^m). Hence, (H_i^m, τ_i^m) is a Souslin space.

Let τ_i be the coarsest (weakest) topology on H_i that makes the projection mappings into H_i^m, $m \geq 1$, continuous. Then, (H_i, τ_i) is a Souslin space (Bourbaki, 1998, p. IX.63), and it is straightforward to show that $\mathscr{B}(H_i) = \mathcal{F}_{H_i}$.

Step 2: Analytic spaces By Step 1, H_i is the image $f(H_i)$ of an analytic Borel space under a measurable function f. The result then follows from Theorem 5.1 of Mackey (1957) if $(H_i, \mathcal{F}_{H_i})$ is countably generated. But this follows directly from the fact that $(H_i^m, \mathcal{F}_i^m)$ is countably generated for $m = 1, 2, \ldots$. □

[19] This requires using the sum topology on the disjoint union of topological spaces (e.g., Kechris, 1995, p. 3, 13).

B.2 Proof of Proposition 6.3 (cont.)

Here I prove the results for the base case of the induction (i.e., $k = 1$).

Lemma B.1. The function $h_i^{\mathcal{T},1}$ is well-defined. That is, $h_i^{\mathcal{T},1}(t_i) \in H_i^{\mathcal{T},1}$.

Proof. Immediate from the definitions. □

Let $\mathcal{Q}_i^1$ be the coarsest σ-algebra that separates the types according to their belief on $\mathcal{Q}_{-i}^0$.

Lemma B.2. The σ-algebra $\mathcal{Q}_i^1$ is generated by the function $h_i^{\mathcal{T},1}$, that is,

$$\mathcal{Q}_i^1 = \left\{ \left\{ t_i \in T_i : h_i^{\mathcal{T},1}(t_i) \in B_i^1 \right\} : B_i^1 \in \mathcal{F}_{i,1}^{\mathcal{T},1}. \right.$$

The proof follows directly from the definitions.

Lemma B.3. The function $h_i^{\mathcal{T},1}$ is $(\mathcal{Q}_i^n, \mathcal{F}_{i,n}^{\mathcal{T},1})$-measurable for $n = 0, 1$.

Proof. The result holds trivially for $n = 0$, as $\mathcal{Q}_i^0$ and $\mathcal{F}_{i,0}^{\mathcal{T},1}$ are both trivial σ-algebras. So, let $n = 1$. By standard arguments (Aliprantis and Border, 2005, Coroll. 4.24), the function $h_i^{\mathcal{T},1}$ is $(\mathcal{Q}_i^1, \mathcal{F}_{i,1}^{\mathcal{T},1})$-measurable if and only if

$$\{t_i \in T_i : E \in \Sigma_{t_i}, \pi_{t_i}(E) \geq p\} \in \mathcal{Q}_i^1$$

for $E \in \overline{\mathcal{Q}}_{-i}^0$ and $p \in [0, 1]$. But this follows from the definition. □

Lemma B.4. The σ-algebra $\mathcal{Q}_i^1$ is at least as fine as $\mathcal{Q}_i^0$, that is, $\mathcal{Q}_i^1 \supset \mathcal{Q}_i^0$.

Proof. Immediate, as $\mathcal{Q}_i^0$ is the trivial σ-algebra. □

Lemma B.5. For every type t_i, one of the following is the case: either $\Sigma_{t_i} \supseteq \mathcal{Q}_{-i}^{\mathcal{T},1}$, or $\Sigma_{t_i} = \mathcal{Q}_{-i}^0$.

Proof. The proof is identical to the proof of Lemma 6.9, and thus omitted. □

B.2.1 Proof of Lemma 6.7

The result holds trivially for $n = 0$. Also, for $n = k$, it follows from Lemma 6.6. So, suppose that $n = 1, \ldots, k-1$. By Lemma A.1 in the appendix, the σ-algebra $\mathcal{F}_{i,n}^{\mathcal{T},k}$ is the coarsest σ-algebra on $H_i^{\mathcal{T},k}$ that contains the sets

$$\{(\mu_i^1, \ldots, \mu_i^k) \in H_i^{\mathcal{T},k} : E \in \Sigma(\mu_i^n), \mu_i^n(E) \geq p\}$$

for $E \in \overline{\mathcal{F}}^{n-1}_{-i,m}$, $m = 0, 1, \ldots, n-1$, and $p \in [0,1]$. By standard arguments (e.g., Aliprantis and Border, 2005, Lemma 4.23), the function $h_i^{\mathcal{T},k}$ is $(\mathcal{Q}_i^n, \mathcal{F}_{i,n}^{\mathcal{T},k})$-measurable if and only if

$$\{t_i \in T_i : E \in \Sigma(\mu_{t_i}^n), \mu_{t_i}^n(E) \geq p\} \in \mathcal{Q}_i^n$$

for $E \in \overline{\mathcal{F}}^{n-1}_{-i,m}$, $m = 0, 1, \ldots, n-1$, and $p \in [0,1]$, where we recall that $\mu_{t_i}^n = \pi_{t_i} \circ (\mathrm{Id}_{S_{-i}}, h_{-i}^{\mathcal{T},n-1})^{-1}$.

Fix $m = 0, 1, \ldots, n-1$. By (IH2a), we have that $(h_{-i}^{\mathcal{T},n-1})^{-1}(E') \in \mathcal{Q}_{-i}^m$ for $E' \in \mathcal{F}_{-i,m}^{n-1}$. So, it suffices to show that

$$\{t_i \in T_i : E'' \in \Sigma_{t_i}, \pi_{t_i}(E'') \geq p\} \in \mathcal{Q}_i^n$$

for $E'' \in \overline{\mathcal{Q}}_{-i}^m$ and $p \in [0,1]$. By (IH2c), $\mathcal{Q}_{-i}^{n-1} \supseteq \mathcal{Q}_{-i}^m$. Hence, as $\mathcal{Q}_i^n$ separates the types according to their belief on $\mathcal{Q}_{-i}^{n-1}$, it also separates the types according to their belief on $\mathcal{Q}_{-i}^m$, and the result follows. □

B.3 Proof of Proposition 6.4 (cont.)

I show that $(H_i^{\mathcal{T},k}, \mathcal{F}_{i,k}^{\mathcal{T},k})$ is an analytic Borel space. This follows from Theorem 5.1 of Mackey (1957) and (ANL_T) if we show the following:

- the σ-algebra $\mathcal{F}_{i,k}^{\mathcal{T},k}$ is countably separated; and
- $h_i^{\mathcal{T},k}$ is measurable with respect to $\mathcal{F}_i^{\mathcal{T}}$ and $\mathcal{F}_{i,k}^{\mathcal{T},k}$;

where countably separated σ-algebras are defined in Appendix A and $\mathcal{F}_i^{\mathcal{T}}$ is as in (ANL_T). The first claim follows from the following lemma:

Lemma B.6. For $m = 1, 2, \ldots$, the σ-algebra $\mathcal{F}_{i,m}^{\mathcal{T},m}$ is countably generated.

Proof. The proof is by induction. Fix a player $i = a, b$. The σ-algebra $\mathcal{F}_{i,1}^{\mathcal{T},1}$ is generated by the sets

$$\{\mu_i^1 \in H_i^{\mathcal{T},1} : \mu_i^1(E_1) \geq p_1, \ldots, \mu_i^1(E_x) \geq p_x\} : \qquad E_1, \ldots, E_x \in \mathcal{F}_{S_{-i}}, p_1, \ldots, p_x \in \mathbb{Q}.$$

Clearly, the collection of sets of this form is countable. Denote the collection of sets of this form by $\mathcal{A}_{i,1}^1$. Then, $\mathcal{A}_{i,1}^1$ is a semiring. It is also separating: Suppose $\mu_i^1, \nu_i^1 \in H_i^1$ such that $\mu_i^1 \neq \mu_i^2$. Then there is $E \in \mathcal{F}_{S_{-i}}$ and $p \in \mathbb{Q}$ such that $\mu_i^1(E) < p \leq \nu_i^1(E)$.

For $m = 2, 3, \ldots$, suppose that for each player i and $\ell \leq m-1$, $\mathcal{A}_{i,\ell}^\ell$ is a countable semiring that is separating and that generates $\mathcal{F}_{i,\ell}^{\mathcal{T},\ell}$. It is straightforward to check that for all ℓ, $\mathcal{F}_{i,\ell}^{\mathcal{T},m-1}$ is generated by a countable semiring; denote this semiring by $\mathcal{A}_{i,\ell}^{m-1}$. Fix a player $i = a, b$ and define $\mathcal{A}_{i,m}^m$ to be the collection of sets $A_{i,m-1}^{m-1} \times F$, where $A_{i,m-1}^{m-1} \in \mathcal{A}_{i,m-1}^{m-1}$ and F is of the form

$$\{\mu_i^m \in \Delta^+(S_{-i} \times H_{-i}^{\mathcal{T},m-1}) : \Sigma(\mu_i^m) = \overline{\mathcal{F}}_{-i,\ell}^{\mathcal{T},m-1}, \mu_i^m(E_1) \geq p_1, \ldots, \mu_i^m(E_x) \geq p_x\}$$
$$\ell = 0, \ldots, m-1, E_1, \ldots, E_x \in \overline{\mathcal{F}}_{-i,\ell}^{\mathcal{T},m-1}, p_1, \ldots, p_x \in \mathbb{Q}. \quad \text{(B.1)}$$

It is easy to verify that the collection of sets of the form (B.1) is a countable semiring, and it follows that $\mathcal{A}_{i,m}^m$ is a countable semiring. By Lemma 2 of Liu (2009) (and using that the product σ-algebra is generated by the semiring of measurable rectangles), $\mathcal{A}_{i,m}^m$ generates $\mathcal{F}_{i,m}^{\mathcal{T},m}$. The proof that $\mathcal{F}_{i,m}^{\mathcal{T},m}$ is separated is similar to the proof for $m = 1$ and is thus omitted. □

To see that $h_i^{\mathcal{T},k}$ is measurable with respect to $\mathcal{F}_i^{\mathcal{T}}$ and $\mathcal{F}_{i,k}^{\mathcal{T},k}$, note that it is measurable with respect to $\mathcal{Q}_i^k$ and $\mathcal{F}_{i,k}^{\mathcal{T},k}$ (by Lemma 6.7) and that $\mathcal{F}_i^{\mathcal{T}} \supset \mathcal{Q}_i^k$ (by Lemma 6.9). □

B.4 Proof of Lemma 6.11

Some preliminary notation will be useful. A class $\mathscr{C}$ of subsets of a set X is *compact* if for any sequence $C_1, C_2, \ldots$ in $\mathscr{C}$, $\bigcap_{n=1}^{\infty} C_n = \emptyset$ implies that $\bigcap_{n=1}^{N} C_n = \emptyset$ for some $N < \infty$. A compact class $\mathscr{C}$ *approximates* a probability measure μ on a σ-algebra $\mathcal{F}$ on X if for each event $E \in \mathcal{F}$ and $\varepsilon > 0$, there is $C \in \mathscr{C}$ such that $C \in \mathcal{F}$, $C \subset E$, and $\mu(E \setminus C) < \varepsilon$. This is a measure-theoretic version of the topological property of tightness (also called inner regularity),[20] which plays a key role in extension results for topological spaces.

Fix a belief hierarchy $(\mu_i^1, \mu_i^2, \ldots)$ with an infinite depth. I first show that for every m, there is a compact class that approximates the mth-order belief μ_i^m. This will allow me to apply an extension result due to Choksi (1958).

Step 1: Approximating compact classes. It will be convenient to define

$$\Omega_i^m := \begin{cases} S_{-i} \times H_{-i}^{m-1} & \text{if } m > 1; \\ S_{-i} & \text{otherwise} \end{cases}$$

to be the space of uncertainty for the mth-order belief μ_i^m. As noted in the proof of Lemma A.3, for each m, there is a topology τ_i^m on Ω_i^m such that Ω_i^m is separable metrizable and its Borel σ-algebra coincides with the original σ-algebra (viz., $\mathcal{F}_{S_{-i}}$ if $m-1$, and $\mathcal{F}_{S_{-i}} \otimes \mathcal{F}_{-i,m-1}^{m-1}$ otherwise). For $m \geq 1$, let $\mathscr{C}_i^m$ consist of Ω_i^m and the compact subsets of Ω_i^m (in τ_i^m). Then, $\mathscr{C}_i^m$ is clearly a compact class. I claim that $\mathscr{C}_i^m$ approximates μ_i^m on $\overline{\mathcal{F}}_{-i,m-1}^{m-1}$ (where $\overline{\mathcal{F}}_{-i,0}^{0} := \mathcal{F}_{S_{-i}}$). To see this, note that $\mathscr{C}_i^m \subset \overline{\mathcal{F}}_{-i,m-1}^{m-1}$, as compact sets in a metrizable space are closed and $\overline{\mathcal{F}}_{-i,m-1}^{m-1}$ coincides with the Borel σ-algebra. By Theorem 6.1 of Mackey (1957), there is $D_i^m \in \overline{\mathcal{F}}_{-i,m-1}^{m-1}$ such that $\mu_i^m(D_i^m) = 1$ and D_i^m (endowed with the relative σ-algebra $\mathcal{F}_D$) is a standard Borel space. Let $\mathscr{D}_i^m$ be the collection of subsets of D_i^m that contains D_i^m itself as well as its compact subsets (in the relative topology τ_D on D_i^m induced by τ_i^m); as before, this is a compact class.

[20] Recall that a Borel probability measure μ is *tight* if for every Borel set E, $\mu(E)$ can be approximated by $\mu(K)$ for some compact set K (e.g., Parthasarathy, 2005).

Then, by Theorem III.3.2 of Parthasarathy (2005), μ_i^m can be approximated on $\mathcal{F}_D$ by $\mathscr{D}_i^m$. By Theorem 13.11 of Kechris (1995), we can choose τ_D such that every subset of D_i^m that is compact in τ_D is compact in τ_i^m. Hence, μ_i^m is approximated on $\overline{\mathcal{F}}_{-i,m-1}^{m-1}$ by $\mathscr{C}_i^m$.

Step 2: Extension. For $m \geq 1$, let $\mathscr{C}_i^m$ be the approximating compact class on Ω_i^m defined above. The result that every belief hierarchy with an infinite depth has an associated "canonical" belief over the belief hierarchies of the other player now follows from Theorem 3.1 of Choksi (1958) if we show the following for all m and $\ell < m$:

(i) The projection of $\mathscr{C}_i^m$ on Ω_i^m into Ω_i^{m-1} is a subset of $\mathscr{C}_i^{m-1}$ on Ω_i^{m-1}; and

(ii) for every $(s_{-i}, \mu_{-i}^1, \ldots, \mu_{-i}^{\ell-1}) \in \Omega_i^\ell$,

$$\mathscr{C}_i^m \cap \left\{(s'_{-i}, \nu_{-i}^1, \ldots, \nu_{-i}^{m-1}) \in \Omega_i^m : (s'_{-i}, \nu_{-i}^1, \ldots, \nu_{-i}^{\ell-1}) = (s_{-i}, \mu_{-i}^1, \ldots, \mu_{-i}^{\ell-1})\right\}$$

is a compact class.

To see that (i) holds, note that the projection of a compact set is compact. It follows directly from the definitions that (ii) holds. □

B.5 Proof of Lemma 6.13 (cont.)

Case $k < \infty$ (cont.). I show that $\mathcal{F}_{i,k}$ is the coarsest σ-algebra that contains the sets

$$\{(\mu_i^1, \mu_i^2, \ldots) \in H_i : \Sigma(\mu_i^k) = \overline{\mathcal{F}}_{-i,\ell}^{k-1}, \mu_i^k(E'') \geq p\} : \qquad \ell = 0, \ldots, k-1, E'' \in \overline{\mathcal{F}}_{-i,\ell}^{k-1}, p \in [0,1].$$

To show this, note that by (COH) and (IND), $\mathcal{F}_{i,k}$ is the coarsest σ-algebra that contains the sets

$$\{(\mu_i^1, \mu_i^2, \ldots) \in H_i : \mu_i^k \in B\} : \qquad B \in \mathcal{F}_{\Delta^+(S_{-i} \times H_{-i})}.$$

The result then follows by noting that $\mathcal{F}_{\Delta^+(S_{-i} \times H_{-i})}$ is the coarsest σ-algebra that contains the sets

$$\{\mu_i \in \Delta^+(S_{-i} \times H_{-i}) : \Sigma(\mu_i) = \overline{\mathcal{F}}_{-i,\ell}^{k-1}, \mu_i(F) \geq p\} \qquad \ell = 0, 1, \ldots, k-1, F \in \overline{\mathcal{F}}_{-i,\ell}^{k-1}, p \in [0,1];$$

and that taking inverse images preserves σ-algebras (e.g., Aliprantis and Border, 2005, Lemma 4.23).

Case $k = \infty$. I show that $\mathcal{F}_{a,\infty} \succ^* \mathcal{F}_{b,\infty}$ and vice versa. This implies that $\mathcal{F}_{a,\infty}$ and $\mathcal{F}_{b,\infty}$ form a mutual-separation pair. Using that the σ-algebras in $\mathscr{F}_i^{\mathcal{H}}$ form a filtration, this is equivalent to showing that $\mathcal{F}_{i,\infty}$ is the coarsest σ-algebra that contains the sets

$$\{h_i \in H_i : E \in \overline{\Sigma}_{h_i}^{\mathcal{H}}, \pi_{h_i}(E) \geq p\} \qquad E \in \overline{\mathcal{F}}_{-i,m}, m = 0, 1, \ldots, \infty, p \in [0,1].$$

By Lemma A.1 in Appendix A, this holds if and only if $\mathcal{F}_{i,\infty}$ is the coarsest σ-algebra that contains the sets

$$\{h_i \in H_i : \overline{\Sigma}_{h_i}^{\mathcal{H}} = \overline{\mathcal{F}}_{-i,m}, \pi_{h_i}(E) \geq p\} \qquad m = 0, 1, \ldots, \infty, E \in \overline{\mathcal{F}}_{-i,m}, p \in [0,1].$$

By definition, $\overline{\mathcal{F}}_{-i,\infty}$ is the coarsest σ-algebra that makes the projection mappings $\text{proj}_{-i,m}$ from H_{-i} into H_{-i}^m measurable (when H_{-i}^m is endowed with the σ-algebra $\mathcal{F}_{-i,m}^m$) for all m. So, by Lemma 2 of Liu (2009), it suffices to show that $\mathcal{F}_{i,\infty}$ is the coarsest σ-algebra that contains the sets

$$\{h_i \in H_i : \overline{\Sigma}_{h_i}^{\mathcal{H}} = \overline{\mathcal{F}}_{-i,\ell}, \pi_{h_i}(E) \geq p\} : \ell = 0, 1, \ldots, \infty, E \in \overline{\mathcal{F}}_{-i,m}, m \leq \ell, m < \infty, p \in [0,1]. \tag{B.2}$$

That $\mathcal{F}_{i,\infty}$ contains the sets in (B.2) follows from (6.4) and the fact that $\mathcal{F}_{i,\infty} \supset \mathcal{F}_{i,m}$ for all $m < \infty$. Since $\mathcal{F}_{i,\infty}$ is the coarsest σ-algebra that makes the projection functions $\text{proj}_{i,m}$ measurable, it is, in fact, the coarsest σ-algebra that contains the sets in (B.2). □

B.6 Proof of Theorem 7.5 (cont.)

By the proof of Proposition 6.3, the σ-algebra $\mathcal{F}_{i,1}^u$ on T_i^u associated with a type $t_{-i} \in T_{-i}^u$ of depth 2 is the coarsest σ-algebra that contains the sets

$$\{t_i \in T_i^u : \pi_{t_i}^u(E) \geq p\} : \qquad E \in \mathcal{F}_{S_{-i}}, p \in [0,1].$$

The proof uses the following lemmas.

Lemma B.7. Suppose $E_{t_a^1} \in \mathcal{F}_{b,1}^u$ is such that $\pi_{\varphi_a(t_a^1)}^u(E_{t_a^1}) = 1$. Then, $B^u \subset E_{t_a^1}$.

Proof. Clearly, $B^u \in \mathcal{F}_{b,1}^u$. Moreover, there is no nonempty proper subset F of B^u such that $F \in \mathcal{F}_{b,1}^u$. So, as $E_{t_a}^1 \in \mathcal{F}_{b,1}^u$ $E_{t_a^1} \supset B^u$ or $E_{t_a^1}$ and B^u are disjoint. Since φ preserves beliefs,

$$\begin{aligned} (\varphi_b)^{-1}(B^u) &= \{t_b^1, t_b^2\} \\ (\varphi_b)^{-1}(E_{t_a^1}) &\supset \{t_b^1, t_b^2\}. \end{aligned}$$

Conclude that $E_{t_a^1} \supset B^u$. □

Lemma B.8. We have

$$\pi_{t_b^u}(\{t_a \in T_a^u : \pi_{t_a}^u(s_b^1) = 1\}) = 1.$$

Proof. Observe that $\{t_a \in T_a^u : \pi_{t_a}^u(s_b^1) = 1\} \in \mathcal{F}_{a,1}^u$. Suppose by contradiction that

$$\pi_{t_b^u}(\{t_a \in T_a^u : \pi_{t_a}^u(s_b^1) \neq 1\}) > 0.$$

Then, since φ' preserves beliefs,

$$\pi'_{\tilde{t}_b}((\varphi'_a)^{-1}(\{t_a \in T_a^u : \pi_{t_a}^u(s_b^1) \neq 1\})) > 0,$$

and therefore

$$\{t'_a \in T'_a : \pi'_{t_a}(s_b^1) = 1\} \cap \{t'_a \in T'_a : \varphi'_a(t'_a) \in \{t_a^u \in T_a^u : \pi_{t_a^u}^u(s_b^1) \neq 1\}\} \neq \emptyset.$$

But, since φ' preserves beliefs, this is equivalent to

$$\{t'_a \in T'_a : \pi'_{t'_a}(s_b^1) = 1\} \cap \{t'_a \in T'_a : \pi'_{t'_a}(s_b^1) \neq 1\} \neq \emptyset,$$

a contradiction. □

References

Ahn, D. S. (2007). Hierarchies of ambiguous beliefs. *Journal of Economic Theory 136*, 286–301.

Aliprantis, C. D. and K. C. Border (2005). *Infinite Dimensional Analysis: A Hitchhiker's Guide* (3rd ed.). Berlin: Springer.

Battigalli, P. and A. Friedenberg (2009). Context-dependent forward induction reasoning. Working paper, Bocconi and Arizona State University.

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

Bogachev, V. (2007). *Measure Theory*, Volume 1. Springer.

Bourbaki, N. (1970). *Theory of Sets*. Elements of Mathematics. Springer.

Bourbaki, N. (1998). *General Topology: Chapters 5–10*. Elements of Mathematics. Springer.

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

Brandenburger, A. and H. J. Keisler (2006). An impossibility theorem on beliefs in games. *Studia Logica 84*, 211–240.

Choksi, J. R. (1958). Inverse limits of measure spaces. *Proceedings London Mathematical Society s3-8*, 321–342.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature 51*, 5–62.

Dekel, E., D. Fudenberg, and S. Morris (2007). Interim correlated rationalizability. *Theoretical Economics 2*, 15–40.

Di Tillio, A. (2008). Subjective expected utility in games. *Theoretical Economics 3*, 287–323.

Epstein, L. G. and T. Wang (1996). Beliefs about beliefs without probabilities. *Econometrica 64*, 1343–1373.

Friedenberg, A. and M. Meier (2012). On the relationship between hierarchy and type morphisms. *Economic Theory 46*, 377–399.

Friedenberg, A. and M. Meier (2016). The context of the game. *Economic Theory*. Forthcoming.

Ganguli, J., A. Heifetz, and B. S. Lee (2016). Universal interactive preferences. *Journal of Economic Theory 162*, 237–260.

Harsanyi, J. C. (1967/1968). Games of incomplete information played by Bayesian players. Parts I–III. *Management Science 14*, 159–182, 320–334, 486–502.

Heifetz, A. (1993). The Bayesian formulation of incomplete information–The non-compact case. *International Journal of Game Theory 21*, 329–338.

Heifetz, A. and W. Kets (2018). Robust multiplicity with a grain of naiveté. *Theoretical Economics 13*, 415–465.

Heifetz, A., M. Meier, and B. Schipper (2006). Interactive unawareness. *Journal of Economic Theory 130*, 78–94.

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

Heifetz, A. and D. Samet (1999). Coherent beliefs are not always types. *Journal of Mathematical Economics 32*, 475–488.

Kechris, A. S. (1995). *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Berlin: Springer-Verlag.

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

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

Lubin, A. (1974). Extensions of measures and the von Neumann selection theorem. *Proceedings of the American Mathematical Society 43*, 118–122.

Mackey, G. (1957). Borel structure in groups and their duals. *Transactions of the American Mathematical Society 85*, 134–165.

Mertens, J.-F., S. Sorin, and S. Zamir (1994). Repeated games: Part A: Background material. Discussion Paper 9420, CORE.

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

Parthasarathy, K. (2005). *Probability Measures on Metric Spaces*. AMS Chelsea Publishing.

Perea, A. and W. Kets (2016). When do types induce the same belief hierarchy? *Games 7*(28).

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons.

Siniscalchi, M. (2008). Epistemic game theory: Beliefs and types. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary Of Economics*. Palgrave Macmillan.

Strzalecki, T. (2014). Depth of reasoning and higher-order beliefs. *Journal of Economic Behavior & Organization 108*, 108–122.

Tsakas, E. (2014). Rational belief hierarchies. *Journal of Mathematical Economics 51*, 121–127.