Bounded Reasoning and Higher-Order Uncertainty

Willemien Kets*

September 4, 2014
First version: November 2011

Abstract

Experimental evidence suggests that individuals cannot reason about their opponents' beliefs up to arbitrarily high orders, that is, that they have a *finite depth of reasoning*. Yet, common belief can often be attained, for example, through a public announcement. This paper resolves this paradox by introducing a framework that explicitly models the beliefs of players with a finite depth. I show that players can reason about higher-order events such as the event that there is common belief if and only if the higher-order event is entailed by an event of sufficiently low order. This means that there can be common belief in an event even if high-order mutual belief in the event cannot be attained.

JEL classification: C700, C720, D800, D830
Keywords: Bounded rationality, higher-order beliefs, games, finite depth of reasoning, common belief.

*Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu. Phone: +1-505-204 8012. I am grateful to Adam Brandenburger, Yossi Feinberg, and Matthew Jackson for their guidance and support, to Adam Brandenburger, Eddie Dekel, Alfredo Di Tillio, Jeff Ely, Amanda Friedenberg, Ben Golub, Joe Halpern, Aviad Heifetz, Philippe Jehiel, Rosemarie Nagel, Antonio Penta, Marcin Peski, Tomasz Sadzik, Dov Samet, Marciano Siniscalchi, Rani Spiegler, and Jonathan Weinstein for stimulating discussions, and to numerous seminar audiences for helpful comments. Part of this research was carried out during visits to Stanford University and the NYU Stern School of Business, and I thank these institutions for their hospitality. Financial support from the Air Force Office for Scientific Research under Grant FA9550-08-1-0389 is gratefully acknowledged.

1. Introduction

Analyzing games of incomplete information requires taking into account not only the beliefs of players, but also their *higher-order beliefs*. Consider, for example, a player who needs to decide which project to invest in. The payoff associated with each choice depends on the economic fundamentals – the state of nature –, as well as the actions of other investors. The player's optimal decision thus depends on her beliefs about the actions of the other players and the state of nature, that is, her first-order belief. Because the same is true for her opponents, the player's optimal action will also depend on her belief about her opponents' beliefs about the state of nature, that is, her second-order belief. And because her opponents in turn condition their action on *their* beliefs about their opponents' beliefs about the state of nature, the player's action choice will depend on her belief about her opponents' beliefs about their opponents' beliefs about nature (her third-order belief), and so on, ad infinitum (Harsanyi, 1967–1968).

Are "real" players capable of such higher-order reasoning? The answer to this question is not so clear-cut as it may seem. A statement such as "John Dean did not know that Nixon knew that Dean knew that Nixon knew that McCord had burgled O'Brien's office in the Watergate Apartments" is inherently difficult to reason about (Clark and Marshall, 1981; Kinderman et al., 1998). At the same time, other types of higher-order reasoning seem unproblematic. If two players hear a (truthful) public announcement that the price of a widget is \$2, then clearly each of them believes that the price of the widget is \$2, believes that the other believes that, believes that the other believes that they believe that, and so on. That is, it is common belief between the players that the price of the widget is \$2.[1]

Existing models do not take into account that some higher-order events are easier to reason about than others. On the one hand, the standard game-theoretic framework models players as if they have higher-order beliefs about every possible event, at all orders, that is, as if players have an infinite depth of reasoning. On the other hand, in models in which players can have a finite depth of reasoning, such as cognitive-hierarchy models or models of level-k reasoning, it is assumed that a player with a finite depth of reasoning cannot reason about *any* event at higher orders (see Crawford et al., 2012, for a survey of the literature). Because beliefs at arbitrarily high order can have a significant impact on economic outcomes, as discussed below, it is important to carefully model which higher-order events players can have beliefs about. This paper provides a framework that does just that.

[1] I follow the recent literature in game theory in using the terms "belief" and "common belief" rather than "knowledge" and "common knowledge." The formal distinction is that knowledge is considered to be always true, while (probability-one) belief may be true or false.

The key idea is that a player may not ask herself all possible questions about the state of the world, where the state of the world describes the economic fundamentals – the state of nature – as well as players' beliefs about the state of nature, their beliefs about others' beliefs, and so on. For example, she may ask herself what the other player thinks about the state of nature, and what he thinks she thinks, but not what he thinks she thinks he thinks. This means she has a coarse perception of the state of the world: she does not distinguish among states of the world that differ only in certain details, such as the beliefs of other players at high orders.

I introduce a class of type spaces that can model such coarse perceptions. Every type is associated with a belief (probability measure) over the states of nature and the types of other players, as in the type spaces of Harsanyi (1967–1968). Unlike in Harsanyi type spaces, types can have different depths of reasoning. The depth of reasoning of a type is determined by the set of events it can assign a probability to, that is, by its σ-algebra on the other players' types. The σ-algebra represents the questions that a type asks: if it contains the event that the other player puts probability 1 on the event that the price of the widget is \$2, then a player can assign a probability to the event that her opponent believes that the widget costs \$2. Likewise, if the σ-algebra contains the set of types for the other player that believe that his opponent believes that the price of the widget is \$2, then a player can reason about the event that her opponent believes that his opponent believes that the price of the widget is \$2. And so on.[2] Generally, a type with a coarse σ-algebra can assign a probability only to certain subsets of types; and since types generate higher-order beliefs, this means that the type can form beliefs only about certain aspects of the other players' higher-order beliefs. Harsanyi type spaces are then type spaces in which types can reason about all features of the other players' higher-order beliefs. Hence, the type spaces introduced here generalize the Harsanyi framework.

This paper is the first to represents players' depth of reasoning by the set of events they can reason about, rather than as a single number, as in previous work. This makes it possible to shed light on the question whether players with a finite depth of reasoning can attain common belief. This can have important economic implications. For example, whether common belief can be attained may determine whether successful coordination is possible (Schelling,

[2]Coarse perceptions thus model small worlds, as introduced by Savage (1954) in the context of one-person decision situations. A state 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 may ignore the distinction between types for the other player that differ only in the beliefs they generate at high order, by lumping together these types into one set in her σ-algebra.

1960) and whether there is scope for speculative trade (cf. Aumann, 1976; Geanakoplos and Polemarchakis, 1982). Moreover, the set of equilibria of a game may depend sensitively on whether the payoffs are common belief, or almost-common belief (Rubinstein, 1989; Carlsson and van Damme, 1993). I show that players with a finite depth of reasoning can form a belief about a high order event if and only if it is entailed by a lower-order event. This means that players can attain common belief whenever it is induced by a low-order event, such as a public announcement. However, as I show, common belief in an event does not entail high-order mutual belief in the event, unlike in the Harsanyi case. Intuitively, even if a public event generates common belief, it need not lead players to ask themselves whether or not all players believe the event up to some finite order (but possibly not at higher orders).

These insights can in turn be used to extend standard solution concepts to settings where players have a finite depth of reasoning, making it possible to separately test hypotheses related to players' depth of reasoning and assumptions underlying the solution concepts (Kets, 2013, 2014). This is not possible with level-k models, which integrate the modeling of players' reasoning and their behavior, or with solution concepts that differ significantly from standard solution concepts (Strzalecki, 2009; Heifetz and Kets, 2013).

The idea that 'simple' events can induce common belief is not new; it is central to the conceptualization of common knowledge by the philosopher David Lewis (1969) and it underlies the formalization of common knowledge and approximate common belief in Aumann (1976) and Monderer and Samet (1989), respectively. Indeed, speaking of a belief hierarchy such as the one that describes a player's higher-order beliefs about the price of a widget, Lewis writes: "this is a chain of implications, [it does not represent] steps in anyone's actual reasoning. Therefore, there is nothing improper about its infinite length" (p. 53). One contribution of this paper is the insight that when bounds on players' reasoning are modeled by the set of events they can think about, this conclusion follows naturally; this then opens the door to extending standard game-theoretic concepts to settings where players have a finite depth of reasoning, as in Kets (2013, 2014).

The remainder of this paper is organized as follows. The next section illustrates the main results with some simple examples. The formal treatment starts in Section 3.

2. Examples

2.1. Harsanyi type spaces

There are two players, Ann (a) and Bob (b), who are uncertain about the state of nature θ. Players' higher-order beliefs can be represented in a compact way using type spaces (Harsanyi, 1967–1968). In a *Harsanyi type space*, each player $i = a, b$ is endowed with a set T_i of *types*, and associating with each type t_i a *belief* (probability measure) $\beta_i(t_i)$ about the state of nature and the other player's type. The function β_i that maps each type for i into a belief is assumed to be measurable. Each type generates a belief hierarchy, as the next example illustrates:

Example 1. There are two possible states of nature, denoted θ_1 and θ_2. Each player $i = a, b$ has four types, labeled $t_i^1, \ldots, t_i^4$. The beliefs of each type are given in Figure 1.

$\beta_a(t_a^1)$	θ_1	θ_2
t_b^1	1	0
t_b^2	0	0
t_b^3	0	0
t_b^4	0	0

$\beta_a(t_a^2)$	θ_1	θ_2
t_b^1	0	0
t_b^2	0	0
t_b^3	1	0
t_b^4	0	0

$\beta_b(t_b^1)$	θ_1	θ_2
t_a^1	1	0
t_a^2	0	0
t_a^3	0	0
t_a^4	0	0

$\beta_b(t_b^2)$	θ_1	θ_2
t_a^1	0	0
t_a^2	0	0
t_a^3	1	0
t_a^4	0	0

$\beta_a(t_a^3)$	θ_1	θ_2
t_b^1	0	0
t_b^2	0	1
t_b^3	0	0
t_b^4	0	0

$\beta_a(t_a^4)$	θ_1	θ_2
t_b^1	0	0
t_b^2	0	0
t_b^3	0	1
t_b^4	0	0

$\beta_b(t_b^3)$	θ_1	θ_2
t_a^1	0	0
t_a^2	0	1
t_a^3	0	0
t_a^4	0	0

$\beta_b(t_b^4)$	θ_1	θ_2
t_a^1	0	0
t_a^2	0	0
t_a^3	0	1
t_a^4	0	0

Figure 1: A (Harsanyi) type space. The beliefs for types for Ann on the left, and those for Bob on the right; the singleton $\{x\}$ is denoted by x.

Types and their beliefs specify players' higher-order beliefs. For example, type t_a^1 for Ann believes (with probability 1) that the state of nature is θ_1. This specifies the type's *first-order belief* $\mu_a^1(t_a^1)$; of course, the other types t_i also generate a first-order belief $\mu_i^1(t_i)$. Type t_a^1 also believes that Bob believes that the state is θ_1 (as it assigns probability 1 to type t_b^1, which puts probability 1 on $\theta = \theta_1$). This specifies the *second-order belief* $\mu_a^2(t_a^1)$ induced by t_a^1. The second-order belief $\mu_a^2(t_a^1)$ is a probability measure on the state of nature and his *first-order belief hierarchies* $H_b^1 := \{\mu_b^1(t_b) : t_b = t_b^1, \ldots, t_b^4\}$; again, the other types likewise generate a second-order belief. Type t_a^1 also induces a *third-order belief* $\mu_a^3(t_a^1)$ on the state of nature and his *second-order belief hierarchies* $H_b^2 := \{(\mu_b^1(t_b), \mu_b^2(t_b)) : t_b = t_b^1, \ldots, t_b^4\}$: the type believes that Bob believes that Ann believes that the state of nature is θ_1 (as t_b^1 assigns probability 1 to type t_a^1, which puts probability 1 on $\theta = \theta_1$).

If we continue this way, we uncover the kth-order belief $\mu_a^k(t_a^1)$ that t_a^1 generates for every k, with a kth-order belief being a probability measure on the state of nature and his $(k-1)$*th-order belief hierarchies* $H_b^{k-1} := \{(\mu_b^1(t_b), \ldots, \mu_b^{k-1}(t_b)) : t_b = t_b^1, \ldots, t_b^4\}$. This gives the *belief hierarchy* $h_a(t_a^1) = (\mu_a^1(t_a^1), \mu_a^2(t_a^1), \ldots)$ induced by (or generated by) t_a^1. $\triangleleft$

The aim is to model belief hierarchies that potentially have a finite depth of reasoning. It will, however, be useful to start with the case of an infinite depth. Say that a belief hierarchy $h_a = (\mu_a^1, \mu_a^2, \ldots)$ for Ann has an *infinite depth (of reasoning)* if for each k, the kth-order belief μ_a^k can assign a probability to each event induced by Bob's $(k-1)$th-order belief hierarchies. This is the case only if the σ-algebra on which μ_a^k is defined can distinguish between the $(k-1)$th-order belief hierarchies for Bob that differ in their $(k-1)$th-order belief. That is, the belief hierarchy $(\mu_a^1, \mu_a^2, \ldots)$ has an infinite depth of reasoning if the first-order belief μ_a^1 is a probability measure that is defined on all events involving the state of nature; the second-order belief μ_a^2 is a probability measure that is defined on all events involving the state of nature and all events in the σ-algebra $\mathcal{F}_b^1$ that describe Bob's beliefs about the state of nature; and so on.

More precisely, endow the set of state of nature with a σ-algebra $\mathcal{F}_\Theta$. Define $\mathcal{F}_b^1$ to be the coarsest σ-algebra on Bob's first-order belief hierarchies that contain the sets

$$\{\mu_b^1 : E \in \Sigma(\mu_b^1), \mu_b^1(E) \geq p\},$$

for every event $E \in \mathcal{F}_\Theta$ and every probability $p \in [0,1]$, where $\Sigma(\mu_b^1)$ is the σ-algebra on which μ_b^1 is defined. Then, the σ-algebra $\mathcal{F}_b^1$ contains all the events involving Bob's first-order beliefs. For general m, assume that for each player i, the σ-algebra $\mathcal{F}_i^{m-2}$ on player i's $(m-2)$th-order belief hierarchies has been defined. Then, let $\mathcal{F}_b^{m-1}$ be the coarsest σ-algebra on Bob's $(m-1)$th-order belief hierarchies that contains the sets

$$\big\{(\mu_b^1, \mu_b^2, \ldots, \mu_b^{m-1}) : E \in \Sigma(\mu_b^{m-1}), \mu_b^{m-1}(E) \geq p\big\} \tag{2.1}$$

for every probability $p \in [0,1]$ and every event E in $\mathcal{F}_\Theta \times \mathcal{F}_a^{m-2}$ concerning the state of nature and Ann's $(m-2)$th-order belief hierarchies; the σ-algebra $\mathcal{F}_a^{m-1}$ on Ann's $(m-1)$th-order belief hierarchies is defined similarly. The σ-algebra $\mathcal{F}_b^{m-1}$ consists of all events that involve Bob's $(m-1)$th-order belief hierarchies. The belief hierarchy $(\mu_a^1, \mu_a^2, \ldots)$ has an *infinite depth of reasoning* if the mth-order belief μ_a^m is a probability measure on the σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_b^{m-1}$ on the state of nature and Bob's $(m-1)$th-order belief hierarchies for all $m > 1$.[3]

Example 1 (cont.). In the type space in Figure 1, the belief of each type t_a for Ann is defined on the σ-algebra on Bob's type set that distinguishes each individual type for Bob. This means in particular that for any k, the kth-order belief $\mu_a^k(t_a)$ induced by Ann's type can distinguish the $(k-1)$th-order belief hierarchies induced by Bob's types that differ in their $(k-1)$th-order beliefs.[4] So, the belief hierarchy induced by the type has an infinite depth

[3]Definition 1 below is of a different form, but it is equivalent to the current one, by Lemma C.2. Taking $\mathcal{F}_b^{m-1}$ to be the *coarsest* σ-algebra that contains the sets in (2.1) is standard.

[4]Of course, the formal result requires relating the σ-algebra on Bob's type set to the σ-algebra $\mathcal{F}_b^{k-1}$ on Bob's $(k-1)$th-order beliefs. The proof of Lemma 5.1 makes this connection. Also see Corollary 5.3.

of reasoning. The same is true, in fact, for any type in a Harsanyi type space, as should be expected. ◁

2.2. Finite depth of reasoning

The aim is to extend the Harsanyi approach to allow types to induce a belief hierarchy of finite depth. Loosely speaking, a belief hierarchy has depth $k < \infty$ if it can form a belief only about the state of nature and other players' $(k-1)$th-order beliefs. To capture this, I restrict the set of events that a type can assign a probability to, that is, the type's belief is defined on a coarser σ-algebra. The next example demonstrates that a player can form a belief about her opponent's first-order beliefs (but not about his higher-order beliefs) whenever her type's belief is defined on a σ-algebra that distinguishes the types of her opponent that differ in their first-order belief (but not if their beliefs differ only at higher order).

Example 2. Consider the type space in Figure 2. As before, there are two states of nature, θ_1 and θ_2. The belief of each type for Ann about Bob's type is defined on the σ-algebra generated by the partition $\{\{t_b^1, t_b^2\}, \{t_b^3, t_b^4\}\}$, and likewise for the types for Bob.

$\beta_a(t_a^1)$	θ_1	s_b^2
$\{t_b^1, t_b^2\}$	1	0
$\{t_b^3, t_b^4\}$	0	0

$\beta_a(t_a^2)$	θ_1	s_b^2
$\{t_b^1, t_b^2\}$	0	0
$\{t_b^3, t_b^4\}$	1	0

$\beta_b(t_b^1)$	θ_1	θ_2
$\{t_a^1, t_a^2\}$	1	0
$\{t_a^3, t_a^4\}$	0	0

$\beta_b(t_b^2)$	θ_1	θ_2
$\{t_a^1, t_a^2\}$	0	0
$\{t_a^3, t_a^4\}$	1	0

$\beta_a(t_a^3)$	θ_1	s_b^2
$\{t_b^1, t_b^2\}$	0	1
$\{t_b^3, t_b^4\}$	0	0

$\beta_a(t_a^4)$	θ_1	s_b^2
$\{t_b^1, t_b^2\}$	0	0
$\{t_b^3, t_b^4\}$	0	1

$\beta_b(t_b^3)$	θ_1	θ_2
$\{t_a^1, t_a^2\}$	0	1
$\{t_a^3, t_a^4\}$	0	0

$\beta_b(t_b^4)$	θ_1	θ_2
$\{t_a^1, t_a^2\}$	0	0
$\{t_a^3, t_a^4\}$	0	1

Figure 2: A type space in which types have depth 2.

As before, each type t_a for Ann generates a first-order belief $\mu_a^1(t_a)$. Type t_a^1, for example, believes that the state of nature is $\theta = \theta_1$. Each type t_a also induces a second-order belief $\mu_a^2(t_a)$. Type t_a^1, for example, assigns probability 1 to the event that Bob has type t_b^1 or t_b^2 (i.e., to $\{t_b^1, t_b^2\}$), and thus to the event that Bob believes that the state of nature is θ_1 (since both t_b^1 and t_b^2 assign probability 1 to θ_1). However, type t_a^1 cannot say whether or not Bob believes that Ann believes that the state of nature is θ_1. The reason is that t_b^1 and t_b^2 differ in their beliefs about Ann's belief about the state of nature, and t_a^1 cannot assign a probability to the individual types. The third-order belief $\mu_a^3(t_a^1)$ therefore cannot assign a probability to some events involving Bob's second-order beliefs. ◁

To model that players can have a finite depth of reasoning (and potentially different depths), I define a type space in which a player's belief can be defined on different σ-algebras.

Thus, I endow Bob's type set T_b with a collection $\mathcal{S}_b$ of σ-algebras, rather than a single one, as in Harsanyi type spaces. The belief $\beta_a(t_a)$ of a type t_a for Ann about Bob's type is defined on a σ-algebra $\Sigma_a(t_a)$ in $\mathcal{S}_b$, and similarly with the player labels interchanged. The σ-algebra $\Sigma_a(t_a)$ specifies the events that t_a can reason about: the type can assign a probability only to events in $\Sigma_a(t_a)$, but not to other events. In Example 2, type t_a^1 can assign a probability to the event that Bob has type t_b^1 or t_b^2 (and thus to the event that Bob believes that the state of nature is θ_1), but not to the event that Bob has type t_b^1 (and therefore not to the event that Bob believes that Ann believes that the state of nature is θ_1).

A belief hierarchy $h_a = (\mu_a^1, \mu_a^2, \ldots)$ has finite depth of reasoning $k < \infty$ if for any $m \leq k$, the mth-order belief μ_a^m can assign a probability to all events that are expressible in terms of Bob's $(m-1)$th-order belief hierarchies, as before, while for $m > k$, its mth-order belief can assign a probability only to those events regarding Bob's $(m-1)$th-order belief hierarchies *that are expressible in terms of his $(k-1)$th-order belief hierarchies.*

More formally, a belief hierarchy h_a has *finite depth (of reasoning) $k < \infty$* if

- for $m \leq k$, the mth-order belief μ_a^m is defined on $\mathcal{F}_\Theta \times \mathcal{F}_b^{m-1}$ on the state of nature and Bob's $(m-1)$th-order belief hierarchies, as in (2.1); and
- for $m > k$, the mth-order belief μ_a^m is defined on the σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_{b,k-1}^{m-1}$, where $\mathcal{F}_{b,k-1}^{m-1}$ is the coarsest σ-algebra that contains the events that are expressible in terms of Bob's $(k-1)$th-order belief hierarchies, that is, the events

  $$\left\{(\mu_b^1, \mu_b^2, \ldots, \mu_b^{m-1}) : E \in \Sigma(\mu_b^{k-1}), \mu_b^{k-1}(E) \geq p\right\}$$

  for $E \in \mathcal{F}_\Theta \times \mathcal{F}_a^{k-2}$ and $p \in [0,1]$, and $\mathcal{F}_{b,k-1}^{m-1}$ is a strict subset of the σ-algebra $\mathcal{F}_{b,k}^{m-1}$ that contains the events that are expressible in terms of Bob's kth-order belief hierarchies.

(The condition that the σ-algebra $\mathcal{F}_{b,k-1}^{m-1}$ is a strict subset of the σ-algebra $\mathcal{F}_{b,k}^{m-1}$ ensures that the depth of a belief hierarchy is well-defined; see Definition 1 below.) With some abuse of terminology, say that a type has depth k if it generates a belief hierarchy of depth k.

Example 2 (cont.). Refer back to the type space in Figure 2. The higher-order beliefs $\mu_a^k(t_a)$, $k \geq 2$, induced by a type t_a for Ann can assign a probability only to events that are expressible in terms of the state of nature and Bob's first-order beliefs. For example, a type for Ann can assign a probability to the event that Bob has type t_b^1 or t_b^2 (i.e., to $\{t_b^1, t_b^2\}$), and thus to the event that Bob believes that the state of nature is θ_1 (as the set of types for Bob that believe that $\theta = \theta_1$ is $\{t_b^1, t_b^2\}$). On the other hand, the types for Ann cannot assign a probability to the event that Bob has type t_b^1 (i.e., to $\{t_b^1\}$), so they cannot reason about the event that Bob believes that Ann believes that the state of nature is θ_1. It follows that Ann's types have depth 2. (And, in fact, the same is true for Bob's types.) $\triangleleft$

Determining the depth of reasoning of a type, as in the above example, requires making a connection between the σ-algebra on the other player's type set on which the type's belief $\beta_i(t_i)$ is defined on the one hand, and the σ-algebras on which the type's induced mth-order beliefs are defined on the other hand. I show that there is a slight weakening of the standard definition of type spaces such that every type has a well-defined (finite or infinite) depth (Theorem 5.2). The key is that the σ-algebras on the type sets can be chosen in such a way that a type's σ-algebra either separates the types for the other player precisely when they differ in their $(k-1)$th-order belief (but not if they differ in their higher-order beliefs), or it separates the types for the other player when they differ in their belief at any order (Lemma 5.1). The latter is the standard, Harsanyi, case; the former makes it possible to model players with finite depth k.

Example 2 (cont.). The types for Ann have a σ-algebra on Bob's type set that distinguishes the types for Bob precisely when they differ in their first-order belief: types t_b^1 and t_b^2 have the same belief about the state of nature, as do t_b^3 and t_b^4; however, types t_b^1 and t_b^2 differ in their second-order beliefs, and likewise for t_b^3 and t_b^4. In other words, the σ-algebra $\Sigma_a(t_a^1)$ separates the types for Bob if and only if they differ in their beliefs about the state of nature, but lumps them together otherwise. This means that a type for Ann can assign a probability to any event that is expressible in terms of Bob's first-order beliefs, but not events that can be expressed only in terms of his higher-order beliefs. ◁

It is now easy to see that a type of finite depth k can reason about an event if and only if the event is expressible in terms of the other player's $(k-1)$th-order beliefs (Corollary 5.3). So, a player can reason about a high order event, as long as it is entailed by an event of sufficiently low order.

2.3. Common belief

This insight can be used to understand how players with a finite depth of reasoning attain common belief in an event F, even if that involves infinitely many statements of the form "Ann believes that Bob believes... that the other player believes F."

Example 3. Consider the type space in Figure 3. There are three states of nature, denoted θ_1, θ_2 and θ_3, and each player $i = a, b$ has eight types, labeled $t_i^1, \ldots, t_i^8$. The beliefs of a type t_i for player i is defined on the σ-algebra generated by the partition $\{\{t_i^1, t_i^2\}, \{t_i^3, t_i^4\}, \{t_i^5, t_i^6\}, \{t_i^7, t_i^8\}\}$. It can be checked that every type has depth 2.

Nevertheless, it can be common belief that the state of nature is θ_1 or θ_2. Consider the types $t_i^1, \ldots, t_i^4$ for $i = a, b$. Any such type assigns probability 1 to the event F that the state of nature is θ_1 or θ_2, assigns probability 1 to types that put probability 1 on θ_1 or θ_2, and so on. Hence, whenever Ann and Bob have type t_i^1, t_i^2, t_i^3, or t_i^4, there is common belief in F.

$\beta_a(t_a^1)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	1	0	0
$\{t_b^3, t_b^4\}$	0	0	0
$\{t_b^5, t_b^6\}$	0	0	0
$\{t_b^7, t_b^8\}$	0	0	0

$\beta_a(t_a^2)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	0	0
$\{t_b^3, t_b^4\}$	1	0	0
$\{t_b^5, t_b^6\}$	0	0	0
$\{t_b^7, t_b^8\}$	0	0	0

$\beta_a(t_a^3)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	1	0
$\{t_b^3, t_b^4\}$	0	0	0
$\{t_b^5, t_b^6\}$	0	0	0
$\{t_b^7, t_b^8\}$	0	0	0

$\beta_a(t_a^4)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	0	0
$\{t_b^3, t_b^4\}$	0	1	0
$\{t_b^5, t_b^6\}$	0	0	0
$\{t_b^7, t_b^8\}$	0	0	0

$\beta_a(t_a^5)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	0	0
$\{t_b^3, t_b^4\}$	0	0	0
$\{t_b^5, t_b^6\}$	$\frac{1}{2}$	$\frac{1}{2}$	0
$\{t_b^7, t_b^8\}$	0	0	0

$\beta_a(t_a^6)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	0	0
$\{t_b^3, t_b^4\}$	0	0	0
$\{t_b^5, t_b^6\}$	0	0	0
$\{t_b^7, t_b^8\}$	$\frac{1}{2}$	$\frac{1}{2}$	0

$\beta_a(t_a^7)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	0	0
$\{t_b^3, t_b^4\}$	0	0	0
$\{t_b^5, t_b^6\}$	0	0	$\frac{1}{2}$
$\{t_b^7, t_b^8\}$	0	0	$\frac{1}{2}$

$\beta_a(t_a^8)$	θ_1	θ_2	θ_3
$\{t_b^1, t_b^2\}$	0	0	0
$\{t_b^3, t_b^4\}$	0	0	0
$\{t_b^5, t_b^6\}$	0	0	0
$\{t_b^7, t_b^8\}$	0	0	1

Figure 3: The beliefs for the types for Ann in a type space in which there is common belief that θ is θ_1 or θ_2, even if there is no third-order mutual belief in that event.

How can this be? Intuitively, the event $E := \Theta \times \{t_a^1, t_a^2, t_a^3, t_a^4\} \times \{t_b^1, t_b^2, t_b^3, t_b^4\}$ is a *public event*: whenever the players have a type that is consistent with E (e.g., $t_a = t_a^1$ and $t_b = t_b^2$), both players assign probability 1 to E. Moreover, both of them believe F; and believe that both believe F; and believe that both believe that both believe F; and so on. That is, the public event E can induce common belief in F. In practice, such public events can take the form of public announcements or rituals (Chwe, 2001). ◁

This example illustrates that common belief can be attained even if players have a finite depth of reasoning. Intuitively, there is a low-order event that types can reason about that entails all relevant high-order beliefs. If two players hear a (truthful) public announcement that the state of nature is θ_1 or θ_2, for example, then both believe that the state is θ_1 or θ_2. This low-order event implies – in fact, is equivalent to – the event that both believe that both believe that the state is θ_1 or θ_2, and to the event that both believe that both believe that both believe that, and so on.

How about mutual high-order belief in an event? In Harsanyi type spaces, there is common belief in an event F if and only if there is mutual belief in F at all orders, as is well known (e.g., Monderer and Samet (1989); also see Proposition 6.3 below). However, the next example demonstrates that this need not be the case when players have a finite depth of reasoning.

Example 3 (cont.). Refer back to the type space in Figure 3. The event $\{t_a^1, t_a^2, t_a^3, t_a^4, t_a^5\}$ that Ann believes F and that Bob believes F cannot be expressed in terms of Ann's first-order beliefs. Types t_a^1 and t_a^2, for example, have the same first-order beliefs, but differ in their second-order beliefs. This means that the types for Bob cannot assign a probability to this event; likewise, the types for Ann cannot form a belief about the event $\{t_b^1, t_b^2, t_b^3, t_b^4, t_b^5\}$ that Bob believes F and that Ann believes F. This implies that F is not third-order mutual belief. By contrast, the event $\{t_a^1, t_a^2, t_a^3, t_a^4\}$ that Ann believes F, that Bob believes F, that Bob believes that she believes F, and so on, ad infinitum, is expressible in terms of her first-order beliefs. ◁

The example demonstrates that an event can be common belief even if it cannot be mutual belief at all orders. Proposition 6.2 shows that this does not hinge on the specific details of the type space in Example 3. Intuitively, if all types have depth $k < \infty$, then an event F can be common belief if there is a public event that is expressible in players' $(k-1)$th-order beliefs that entails all higher-order beliefs. So, a single low-order event suffices. This need not be the case for high-order mutual belief: if there are types that have mth-order mutual belief in the event F, but not $(m+1)$th-order mutual belief in F, then, in order for F to be mutual belief at all orders, players need to be able to distinguish situations in which there is mth-order mutual belief in F from ones in which there $(m+1)$th-order mutual belief in F. This can only be the case if there are two distinct events, both expressible in terms of players' $(k-1)$th-order beliefs, with one indicating that there is mth-order belief in F, and one indicating that there is $(m+1)$th-order belief in F, and such events may not exist for all m (see Clark and Marshall (1981) for examples of such events).

The remainder of this paper is organized as follows. Section 3 introduces some preliminaries, and Section 4 constructs belief hierarchies of finite and infinite depth. Section 5 defines type spaces, and Section 6 presents the results on common belief and mutual belief. Section 7 discusses the related literature, and Section 8 concludes. Proofs are relegated to the appendices.

3. Preliminaries

For a set X and σ-algebra $\mathcal{F}$ on X, write $\Delta(X, \mathcal{F})$ for the set of probability measures on $\mathcal{F}$ (i.e., on the measurable space $(X, \mathcal{F})$), and endow $\Delta(X, \mathcal{F})$ with the σ-algebra $\mathcal{F}_{\Delta(X,\mathcal{F})}$ generated by 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 naturally separates beliefs (probability measures) according to the probability they assign to events; this makes it possible to talk about "beliefs about beliefs," and so on

(Heifetz and Samet, 1998). Moreover, this σ-algebra coincides with the Borel σ-algebra in the common case that $\Delta(X,\mathcal{F})$ is endowed with the weak topology, X is metrizable, and $\mathcal{F}$ is the Borel σ-algebra on X.

As is standard, the product of measurable spaces is endowed with the product σ-algebra, and a subset Y of a space X, endowed with a σ-algebra $\mathcal{F}_X$, has the relative σ-algebra, denoted by $\mathcal{F}_Y$. If μ is a probability measure on a product space $X \times Y$, then its marginal on X is denoted by $\operatorname{marg}_X \mu$.

For any family of spaces $\{X_z : z \in Z\}$, with X_z endowed with the σ-algebra $\mathcal{F}_z$, $z \in Z$, the union $X := \bigcup_{z\in Z} X_z$ is endowed with the σ-algebra $\mathcal{F}$ that contains precisely the subsets $E \subseteq X$ such that $E \cap X_z \in \mathcal{F}_z$ for all $z \in Z$. That is, $(X,\mathcal{F})$ is the sum of the measurable spaces $(X_z,\mathcal{F}_z)$, $z \in Z$. In particular, if $\mathscr{S}$ is a collection of σ-algebras on a space Y, then the space $\Delta(Y,\mathscr{S}) := \bigcup_{\mathcal{Q}\in\mathscr{S}} \Delta(Y,\mathcal{Q})$ is endowed with the σ-algebra generated by sets of the form

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

where $\Sigma(\mu)$ is the σ-algebra on which the belief μ is defined.

For any topological space X, its Borel σ-algebra is denoted by $\mathscr{B}(X)$. The set $\Delta(X,\mathscr{B}(X))$ of Borel probability measures is endowed with the topology of weak convergence; if X is Polish, then so is $\Delta(X,\mathscr{B}(X))$. As is well-known, the σ-algebra $\mathcal{F}_{\Delta(X,\mathscr{B}(X))}$ coincides with the Borel σ-algebra $\mathscr{B}(\Delta(X,\mathscr{B}(X)))$ whenever X is Polish. The union X of a family of topological spaces X_z, $z \in Z$, is endowed with the disjoint union topology, that is, with the topology whose open sets are precisely the subsets U of X such that $U \cap X_z$ is open in X_z for every $z \in Z$. Then, for any countable collection $\mathscr{S}$ of σ-algebras on a Polish space X, the union $\Delta(X,\mathscr{S})$ of spaces $\Delta(X,\mathcal{F})$, $\mathcal{F} \in \mathscr{S}$, is Polish (e.g., Kechris, 1995, Prop. 3.3).

4. Belief hierarchies

This section introduces an explicit model of players' higher-order beliefs, by constructing their belief hierarchies. I do so in a way that every belief hierarchy has a well-defined depth of reasoning. Section 5 provides an implicit description of these beliefs, by generalizing the familiar Harsanyi representation.

4.1. Construction

Players are uncertain about the state of nature $\theta \in \Theta$. The set Θ of states of nature is assumed to be finite and to contain at least two elements; it is endowed with the standard (discrete) σ-algebra $\mathcal{F}_\Theta$. For simplicity, I focus on the case of two players here; the results generalize to the case of three or more players with minor modifications, see Appendix A. I write $N = \{a, b\}$ for the player set, and if player $i = a, b$ is fixed, then the player other than i is denoted by j, i.e., $j \neq i$.

Players form beliefs about the state of nature, about their opponent's beliefs about the state, and so on. These higher-order beliefs are modeled using belief hierarchies, building on a construction of Mertens and Zamir (1985) for the Harsanyi case. The belief hierarchies are defined in such a way that each belief hierarchy has a well-defined depth of reasoning; see Section 4.2.

Generalizing Definition 2.1 of Mertens and Zamir, I define a *space of belief hierarchies* to be a sequence $\boldsymbol{C} = (\boldsymbol{C}^1, \boldsymbol{C}^2, \ldots)$, with $\boldsymbol{C}^m = \prod_{i \in N} C_i^m$ for all m, that satisfies the following conditions:

(i) For each player $i \in N$, $C_i^1 \subseteq \Delta(\Theta, \mathcal{F}_\Theta)$, and for $m = 2, 3, \ldots$,

$$C_i^m \subseteq C_i^{m-1} \times \Delta(\Theta \times C_j^{m-1}, \mathscr{S}_i^m(\boldsymbol{C}^{m-1})),$$

where the collection $\mathscr{S}_i^m(\boldsymbol{C})$ of σ-algebras is defined below;

(ii) For each player $i \in N$, $m = 1, 2, \ldots$, and $(\mu_i^1, \ldots, \mu_i^m) \in C_i^m$, the marginal $\operatorname{marg}_\Theta \mu_i^2$ equals μ_i^1, and for $m > 2$, the marginal $\operatorname{marg}_{\Theta \times C_j^{m-2}} \mu_i^m$ equals μ_i^{m-1}.

(iii) For each player $i \in N$ and $m = 1, 2, \ldots$, the projection of C_i^{m+1} into $C_i^{m-1} \times \Delta(\Theta \times C_j^{m-1}, \mathscr{S}_i^m(\boldsymbol{C}^{m-1}))$ equals C_i^m.

Condition (i) says that an *mth-order belief hierarchy* $(\mu_i^1, \ldots, \mu_i^m) \in C_i^m$ consists of an $(m-1)$th-order belief hierarchy $(\mu_i^1, \ldots, \mu_i^{m-1})$ and a belief μ_i^m about the state of nature and the other player's $(m-1)$th-order belief hierarchy. The belief μ_i^m is called the *mth-order belief* (induced by the hierarchy). I return to condition (i) below. Condition (ii) is a standard coherency condition that says that beliefs at different orders cannot contradict each other (cf. Mertens and Zamir, 1985; Brandenburger and Dekel, 1993). Condition (iii) says that every mth-order belief hierarchy can be extended to an $(m+1)$th-order belief hierarchy. It is straightforward to show that this condition can be satisfied whenever C_i^{m-1} is nonempty for every player i.

This gives a sequence $C_i^1, C_i^2, \ldots$ of spaces of finite-order belief hierarchies for each player i. A *belief hierarchy* for player i (in $\boldsymbol{C}$) is a sequence $(\mu_i^1, \mu_i^2, \ldots)$ of mth-order beliefs μ_i^m, $m \geq 1$, such that for every ℓ, the belief hierarchy $(\mu_i^1, \ldots, \mu_i^\ell)$ is in C_i^ℓ. Thus, the set of belief hierarchies for player i (in $\boldsymbol{C}$) is

$$H_i(\boldsymbol{C}) := \{(\mu_i^1, \mu_i^2, \ldots) : \text{for all } \ell,\ (\mu_i^1, \ldots, \mu_i^\ell) \in C_i^\ell\}.$$

Condition (i) generalizes a similar condition of Mertens and Zamir (1985, Definition 2.1) by allowing the beliefs of player i at a given order m to be defined on different σ-algebras, by working with a collection $\mathscr{S}_i^m(\boldsymbol{C}^{m-1})$ of σ-algebras, rather than a single one. The collection $\mathscr{S}_i^m(\boldsymbol{C}^{m-1})$ of σ-algebras on which an mth-order belief μ_i^m can be defined is given by

$$\mathscr{S}_i^m(\boldsymbol{C}^{m-1}) := \Big\{\mathcal{F}_\Theta \times \{C_j^{m-1}, \emptyset\}, \mathcal{F}_\Theta \times \mathcal{F}_{j,1}^{m-1}(\boldsymbol{C}^{m-1}), \ldots, \mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{m-1})\Big\},$$

where, for $\ell \leq m$, $\mathcal{F}_{j,\ell-1}^{m-1}(\boldsymbol{C}^{m-1})$ is the σ-algebra generated by the sets of the form

$$\big\{(\mu_j^1, \ldots, \mu_j^{m-1}) \in C_j^{m-1} : \Sigma(\mu_j^{\ell-1}) = \mathcal{F}_\Theta \times \mathcal{F}, \mu_j^{\ell-1}(E) \geq p\big\} \tag{4.1}$$

for $\mathcal{F}_\Theta \times \mathcal{F} \in \mathscr{S}_j^{\ell-1}(\boldsymbol{C}^{\ell-2})$, $E \in \mathcal{F}_\Theta \times \mathcal{F}$ and $p \in [0,1]$.[5] The σ-algebra $\mathcal{F}_{j,\ell-1}^{m-1}(\boldsymbol{C}^{m-1})$ contains precisely the subsets of $(m-1)$th-order belief hierarchies in C_j^{m-1} that can be described in terms of the first $\ell - 1$ orders of beliefs. In other words, every event $E \subseteq C_j^{m-1}$ in this σ-algebra can be characterized by some restriction on the $(\ell-1)$th-order belief hierarchies: every $(m-1)$th-order belief hierarchy in E satisfies that restriction, and, conversely, E contains every belief hierarchy in C_j^{m-1} that satisfies this restriction. Thus, the events in this σ-algebra are completely determined by player j's belief up to order $\ell - 1$. In particular, the events in $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C})$ are determined by j's belief up to order $m-1$. On the other hand, the trivial σ-algebra $\{C_j^{m-1}, \emptyset\}$ does not distinguish the belief hierarchies in any way. Thus, the σ-algebras in $\mathscr{S}_i^m(\boldsymbol{C})$ form a filtration:

$$\{C_j^{m-1}, \emptyset\} \subseteq \mathcal{F}_{j,1}^{m-1}(\boldsymbol{C}) \subseteq \ldots \subseteq \mathcal{F}_{j,m-2}^{m-1}(\boldsymbol{C}) \subseteq \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}).$$

With this selection of σ-algebras, the depth of reasoning of a belief hierarchy is well-defined, as I show next.

4.2. Depth of reasoning

The depth of reasoning of a belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots)$ is defined to be infinite if for every m, the induced mth-order belief μ_i^m can assign a probability to all events that are expressible in terms of player j's mth-order beliefs. The belief hierarchy has a finite depth of reasoning k if its induced mth-order belief can assign a probability only to events that can be expressed in terms of player j's beliefs of order at most $k-1$. Formally:

Definition 1. Let $h_i = (\mu_i^1, \mu_i^2, \ldots) \in H_i(\boldsymbol{C})$ be a belief hierarchy. Then:

[5] For $\ell = 2$, $\mathcal{F}_{j,\ell-1}^{m-1}(\boldsymbol{C}^{m-1})$ is the σ-algebra generated by the sets $\big\{(\mu_j^1, \ldots, \mu_j^{m-1}) \in C_j^{m-1} : \mu_j^{\ell-1}(E) \geq p\big\}$ for $E \in \mathcal{F}_\Theta$ and $p \in [0,1]$.

- h_i has *infinite depth*, denoted $d_i^{\boldsymbol{C}}(h_i) = \infty$, if μ_i^m is a probability measure on $\mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C})$ for all $m = 1, 2, \ldots$;
- h_i has *finite depth* $k = 1, 2, \ldots$, denoted $d_i^{\boldsymbol{C}}(h_i) = k$, if the following hold:
  - for each $m \leq k$, μ_i^m is a probability measure on $\mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C})$;
  - for each $m > k$, μ_i^m is a probability measure on $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{m-1}(\boldsymbol{C})$, and

$$\mathcal{F}_{j,k-1}^{m-1}(\boldsymbol{C}) \subsetneq \mathcal{F}_{j,k}^{m-1}(\boldsymbol{C}) \subseteq \cdots \subseteq \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}).$$

By construction, the depth of reasoning of a belief hierarchy is well-defined:

Lemma 4.1. *For any belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots) \in H_i(\boldsymbol{C})$, there is a unique $k = \infty, 1, 2, \ldots$ such that $d_i^{\boldsymbol{C}}(h_i) = k$.*

Intuitively, the σ-algebras in $\mathscr{S}_i^m(\boldsymbol{C})$, $m = 2, 3, \ldots$ are chosen in such a way that for each mth-order belief μ_i^m, there is some $k \leq m$ such that μ_i^m can assign a probability to precisely those events that can be expressed in terms of the other player's $(k-1)$th-order beliefs. The coherency condition (ii) then ensures that the depth of a belief hierarchy is well-defined: if μ_i^m can assign a probability only to order-$(k-1)$ events for $k < m$, then μ_i^{m+1} can assign a probability only to order-$(k-1)$ events.

4.3. Examples

Before turning to type spaces, it will be instructive to consider a few examples. The first example illustrates that the belief hierarchies constructed by Mertens and Zamir (1985) and others form a space of belief hierarchies in the sense defined here.

Example 4. (Mertens and Zamir, 1985) The set Θ of states of nature is taken to be Polish, and it is endowed with its Borel σ-algebra, i.e., $\mathcal{F}_\Theta = \mathscr{B}(\Theta)$. For each player $i \in N$, take $C_i^{\mathcal{U},1}$ to be the set $\Delta(\Theta, \mathcal{F}_\Theta)$ of all (Borel) probability measures on Θ. For $m = 2, 3, \ldots$, let $C_i^{\mathcal{U},m}$ be the set of mth-order belief hierarchies $(\mu_i^1, \ldots, \mu_i^m)$ (satisfying conditions (i)–(iii) above) such that μ_i^m is defined on $\mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}})$.[6] Note that $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}}) = \mathscr{B}(C_j^{\mathcal{U},m-1})$.

The resulting set $H_i^{\mathcal{U}} := H_i(\boldsymbol{C}^{\mathcal{U}})$ of belief hierarchies for player i is the set of belief hierarchies constructed by Mertens and Zamir (1985). It can be checked that every belief hierarchy has an infinite depth. Mertens and Zamir show that every type from a Harsanyi type space

[6] That is, for $m > 2$, define $C_i^{\mathcal{U},m} := \{(\mu_i^1, \ldots, \mu_i^m) \in C_i^{\mathcal{U},m-1} \times \Delta(\Theta \times C_j^{\mathcal{U},m-1}, \mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}})) : \operatorname{marg}_{\Theta \times C_j^{\mathcal{U},m-2}} \mu_i^m = \mu_i^{m-1}\}$; and for $m = 2$, define $C_i^{\mathcal{U},m} := \{(\mu_i^1, \ldots, \mu_i^m) \in C_i^{\mathcal{U},m-1} \times \Delta(\Theta \times C_j^{\mathcal{U},m-1}, \mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}})) : \operatorname{marg}_\Theta \mu_i^m = \mu_i^{m-1}\}$.

can be mapped into this space in a way that preserves beliefs. I therefore refer to the belief hierarchies $(\mu_i^1, \mu_i^2, \ldots)$ in $H_i^{\mathcal{U}}$ as *Harsanyi (belief) hierarchies.* $\triangleleft$

While the space of belief hierarchies in Example 4 contains all belief hierarchies generated by types in Harsanyi type spaces, it only includes belief hierarchies that have an infinite depth of reasoning. The following example constructs a space of belief hierarchies that contains belief hierarchies of every possible depth.

Example 5. Again, assume that Θ is a Polish space and that $\mathcal{F}_\Theta$ is its Borel σ-algebra $\mathscr{B}(\Theta)$. For each $i \in N$, let $C_i^{*,1} = \Delta(\Theta, \mathcal{F}_\Theta)$ be the set of all Borel probability measures on Θ, as before. For $m = 2, 3, \ldots$, let $C_i^{*,m}$ be the set of mth-order belief hierarchies $(\mu_i^1, \ldots, \mu_i^m)$ (satisfying conditions (i)–(iii) above) such that μ_i^m is defined on any of the σ-algebras in $\mathscr{S}_i^m(\boldsymbol{C}^*)$.[7] Again, the σ-algebra $\mathcal{F}_{j,\ell-1}^{k-1}(\boldsymbol{C}^*)$ is a proper sub-σ algebra of $\mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^*)$ for every $\ell < k$. By standard arguments, the space $C_i^{*,m}$ of mth-order belief hierarchies is nonempty and Polish. The resulting set $H_i^* := H_i(\boldsymbol{C}^*)$ of belief hierarchies contains the space $H_i^{\mathcal{U}}$ of all Harsanyi hierarchies by construction. In fact, the Harsanyi hierarchies form a belief-closed subspace of the space $\boldsymbol{C}^*$ of belief hierarchies (Mertens and Zamir, 1985) that is characterized by the event that players have an infinite depth of reasoning and that there is common belief in the event that players have an infinite depth of reasoning. The set H_i^* of belief hierarchies additionally contains belief hierarchies with a finite depth,[8] belief hierarchies that think possible that the other player has a finite depth, or think possible that the other player thinks that, and so on. $\triangleleft$

5. Type spaces

5.1. Definition

In this section, I generalize the concept of a Harsanyi type space. Each type induces a belief hierarchy, just like Harsanyi types do, except that the belief hierarchy can be of finite depth.

[7] That is, for $m > 2$, define $C_i^{*,m} := \{(\mu_i^1, \ldots, \mu_i^m) \in C_i^{*,m-1} \times \Delta(\Theta \times C_j^{*,m-1}, \mathscr{S}_i^m(\boldsymbol{C}^*)) : \operatorname{marg}_{\Theta \times C_j^{*,m-2}} \mu_i^m = \mu_i^{m-1}\}$; and for $m = 2$, define $C_i^{*,m} := \{(\mu_i^1, \ldots, \mu_i^m) \in C_i^{*,m-1} \times \Delta(\Theta \times C_j^{*,m-1}, \mathscr{S}_i^m(\boldsymbol{C}^*)) : \operatorname{marg}_\Theta \mu_i^m = \mu_i^{m-1}\}$.

[8] That is, H_i^* contains belief hierarchies $(\mu_i^1, \mu_i^2, \ldots)$ such that for some $k < \infty$, for every $m \leq k$, the belief about j's $(m-1)$th-order beliefs is defined on the σ-algebra $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^*)$ that describes j's $(m-1)$th-order beliefs, while for $m > k$, it is defined on the σ-algebra $\mathcal{F}_{j,k-1}^{m-1}(\boldsymbol{C}^*)$ that describes j's $(k-1)$th-order beliefs, where $\mathcal{F}_{j,k-1}^{m-1}(\boldsymbol{C}^*)$ is a strict subset of $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^*)$.

A *type space* is a tuple

$$(T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i\in N}$$

that satisfies Assumption 1 below. For each player i, T_i is a nonempty set of *types*, and $\mathcal{S}_i$ is a nonempty collection of σ-algebras on T_i. The function Σ_i maps the types in T_i to a σ-algebra $\Sigma_i(t_i) \in \mathcal{S}_j$ on T_j, and β_i maps each type t_i into a *belief* $\beta_i(t_i) \in \Delta(\Theta \times T_j, \mathcal{F}_\Theta \times \Sigma_i(t_i))$. The function β_i is player i's *belief map*.

Assumption 1 imposes some further restrictions on the σ-algebras to ensure that each type generates a well-defined belief hierarchy. Thus, this assumption plays a similar role as the familiar condition in the definition of Harsanyi type spaces that belief maps be measurable, as discussed below. To state the assumption, some definitions will be helpful. Say that a σ-algebra $\mathcal{F}_i$ on the type set T_i of player i *dominates* a σ-algebra $\mathcal{F}_j$ on the type set T_j of player j if for every event $E \in \mathcal{F}_\Theta \times \mathcal{F}_j$ and $p \in [0,1]$,

$$\big\{t_i \in T_i : E \in \mathcal{F}_\Theta \times \Sigma_i(t_i), \beta_i(t_i)(E) \geq p\big\} \in \mathcal{F}_i.$$

If $\mathcal{F}_i$ dominates $\mathcal{F}_j$, then I write $\mathcal{F}_i \succ \mathcal{F}_j$; if $\mathcal{F}_i$ is the coarsest σ-algebra that dominates $\mathcal{F}_j$, I write $\mathcal{F}_i \succ^* \mathcal{F}_j$. (The coarsest σ-algebra that dominates $\mathcal{F}_j$ exists: it is the σ-algebra that is the intersection of all σ-algebras on T_i that dominate $\mathcal{F}_j$.) Two σ-algebras $\mathcal{F}_i$ and $\mathcal{F}_j$ on T_i and T_j, respectively, that dominate each other will be called a *mutual-dominance pair*.

Assumption 1. For every player $i \in N$ and any σ-algebra $\mathcal{F}_i \in \mathcal{S}_i$ such that $\mathcal{F}_i \neq \{T_i, \emptyset\}$, there is a σ-algebra $\mathcal{F}_j \in \mathcal{S}_j$ such that one of the following holds:

(a) $(\mathcal{F}_i, \mathcal{F}_j)$ is a mutual-dominance pair; or

(b) $\mathcal{F}_i$ is the coarsest σ-algebra that dominates $\mathcal{F}_j$, i.e., $\mathcal{F}_i \succ^* \mathcal{F}_j$.

Types that are endowed with different σ-algebras have a different depth of reasoning, as shown below. An example of a type space where players may be uncertain about players' depth can easily be constructed. Some type spaces contain types with any depth of reasoning: the space of belief hierarchies in Example 5 can be used to construct such a type space; see Appendix B for details.

Every Harsanyi type space is a type space as defined here. Recall that a *Harsanyi type space* is a tuple $\mathcal{T}^{\mathcal{H}} = (T_i^{\mathcal{H}}, \beta_i^{\mathcal{H}})_{i\in N}$, where for each player i, the type set $T_i^{\mathcal{H}}$ is endowed with some fixed σ-algebra $\mathcal{F}_i^{\mathcal{H}}$, and the belief maps $\beta_i^{\mathcal{H}}$ are measurable. This *measurability condition* is equivalent to the assumption that the σ-algebras on the type sets form a mutual-dominance pair. Hence, any Harsanyi type space $\mathcal{T}^{\mathcal{H}} = (T_i^{\mathcal{H}}, \beta_i^{\mathcal{H}})_{i\in N}$ can be viewed as a type space as defined here, and I sometimes write $\mathcal{T}^{\mathcal{H}} = (T_i^{\mathcal{H}}, \mathcal{S}_i^{\mathcal{H}}, \Sigma_i^{\mathcal{H}}, \beta_i^{\mathcal{H}})_{i\in N}$, where $\mathcal{S}_i^{\mathcal{H}} := \{\mathcal{F}_i^{\mathcal{H}}\}$ and $\Sigma_i^{\mathcal{H}}$ is the trivial mapping.

Thus, Assumption 1 relaxes the standard measurability condition for Harsanyi type spaces. Assumption 1 is clearly weaker than the measurability condition: the type space in Example 2, for example, satisfies Assumption 1, but the belief maps are not measurable (with respect to the players' σ-algebras). Assumption 1 is easy to verify: as with the measurability condition for Harsanyi type spaces, only the relation between two σ-algebras needs to be considered.

5.2. From types to belief hierarchies

Each type can be mapped into a belief hierarchy. To do so, I simultaneously construct the space of belief hierarchies generated by the type space, *and* the functions that maps each type into a belief hierarchy. Essentially, this uses the same construction as in Section 4, where the belief hierarchies were built up using arbitrary subsets C_i^m of mth-order belief hierarchies for each player $i \in N$, except that here the subsets of mth-order belief hierarchies are derived from the type space.

Fix a type space $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$. For each player $i \in N$, define a mapping $h_i^{\mathcal{T},1}$ from T_i to $\Delta(\Theta, \mathcal{F}_\Theta)$ by $h_i^{\mathcal{T},1}(t_i) := \operatorname{marg}_\Theta \beta_i(t_i)$. Clearly, $h_i^{\mathcal{T},1}(t_i) \in \Delta(\Theta, \mathcal{F}_\Theta)$. Define $C_i^{\mathcal{T},1} := h_i^{\mathcal{T},1}(T_i)$ to be the image of $h_i^{\mathcal{T},1}$, and $\boldsymbol{C}^{\mathcal{T},1} := \prod_{i \in N} C_i^{\mathcal{T},1}$. Let $\mathcal{F}_{i,1}^{\mathcal{T},1}(\boldsymbol{C}^{\mathcal{T}})$ be the relative σ-algebra on $C_i^{\mathcal{T},1}$ induced by $\mathcal{F}_{\Delta(\Theta, \mathcal{F}_\Theta)}$.

For $m > 1$, suppose that for each player $i \in N$ and for each $\ell \leq m-1$, the spaces $C_i^{\mathcal{T},\ell}$ and $\boldsymbol{C}^{\mathcal{T},\ell} = \prod_{n \in N} C_n^\ell$ have been defined, and that $\mathscr{S}_i^\ell(\boldsymbol{C}^{\mathcal{T}})$ is a collection of σ-algebras on $\Theta \times C_j^{\mathcal{T},\ell-1}$.[9] Also, assume that the functions $h_i^{\mathcal{T},\ell}$ from T_i into $C_i^{\mathcal{T},\ell}$ have been defined.[10] Define

$$\mathscr{S}_i^m(\boldsymbol{C}^{\mathcal{T}}) := \Big\{\mathcal{F}_\Theta \times \{C_j^{\mathcal{T},m-1}, \emptyset\}, \mathcal{F}_\Theta \times \mathcal{F}_{j,1}^{m-1}(\boldsymbol{C}^{\mathcal{T}}), \dots, \mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{T}})\Big\},$$

where, for $\ell \leq m$, $\mathcal{F}_{j,\ell-1}^{m-1}(\boldsymbol{C}^{\mathcal{T}})$ is generated by the sets

$$\big\{(\mu_j^1, \dots, \mu_j^{m-1}) \in C_j^{\mathcal{T},m-1} : \Sigma(\mu_j^{\ell-1}) = \mathcal{F}_\Theta \times \mathcal{F}, \mu_j^{\ell-1}(E) \geq p\big\},$$

for $\mathcal{F}_\Theta \times \mathcal{F} \in \mathscr{S}_j^{\ell-1}(\boldsymbol{C}^{\mathcal{T}})$, $E \in \mathcal{F}_\Theta \times \mathcal{F}$, and $p \in [0,1]$. Then, define the mapping $h_i^{\mathcal{T},m}$ from T_i to $C_i^{\mathcal{T},m-1} \times \Delta(\Theta \times C_j^{\mathcal{T},m-1}, \mathscr{S}_i^{\mathcal{T},m}(\boldsymbol{C}^{\mathcal{T}}))$ by:

$$h_i^{\mathcal{T},m}(t_i) := \big(h_i^{\mathcal{T},m-1}(t_i), \mu_i^k(t_i)\big),$$

where $\mu_i^k(t_i)$ is the kth-order belief induced by t_i, defined by

$$\mu_i^k(t_i)(E) := \beta_i(t_i)\left(\big\{(s_j, t_j) : (s_j, h_j^{\mathcal{T},m-1}(t_j)) \in E\big\}\right)$$

[9] For $\ell = 1$, take $\mathscr{S}_i^\ell(\boldsymbol{C}^{\mathcal{T}})$ to be the singleton $\{\mathcal{F}_\Theta\}$.

[10] This is with some abuse of notation: the range of $h_i^{\mathcal{T},\ell}$ is in fact a superset of $C_i^{\mathcal{T},\ell}$, as can be seen below; also see Lemma 5.1.

for any $E \subseteq \Theta \times C_j^{\mathcal{T},m-1}$ such that this probability is well-defined. Let $C_i^{\mathcal{T},m}$ be the image of $h_i^{\mathcal{T},m}$, and write $\boldsymbol{C}^{\mathcal{T},m} := \prod_{i \in N} C_i^{\mathcal{T},m}$.

Lemma 5.1. For every $i \in N$ and $m = 1, 2, 3, \ldots$, the functions $h_i^{\mathcal{T},m}$ are well-defined.

The key challenge in proving the lemma is relating the σ-algebra $\Sigma_i(t_i)$ on player j's type set to the σ-algebras on the space of $(m-1)$th-order hierarchies of player j. While Assumption 1 does not put much structure on the relations between σ-algebras, for example allowing cycles of σ-algebras that dominate each other, it provides just enough structure to ensure that each type can be mapped into a well-defined belief hierarchy.

We are now ready to construct the space of belief hierarchies that are generated by some type in $\mathcal{T}$. The sequence $(\boldsymbol{C}^{\mathcal{T},1}, \boldsymbol{C}^{\mathcal{T},2}, \ldots)$ defines a space of belief hierarchies, i.e., it satisfies conditions (i)–(iii) in Section 4.1. Write $\mu_i^1(t_i)$ for the first-order belief $\text{marg}_\Theta \beta_i(t_i)$ induced by t_i; then, for every type $t_i \in T_i$,

$$h_i^{\mathcal{T}}(t_i) := \big(\mu_i^1(t_i), \mu_i^2(t_i), \ldots\big)$$

is a belief hierarchy. I refer to $h_i^{\mathcal{T}}(t_i)$ as the belief hierarchy *induced* (or generated) by t_i.

This gives the following result:

Theorem 5.2. For every type space $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$, and for each player $i \in N$, the belief hierarchies in $H_i(\boldsymbol{C}^{\mathcal{T}})$ are precisely those that are generated by the types in T_i. That is,

- for each type $t_i \in T_i$, there is a belief hierarchy $(\mu_i^1, \mu_i^2, \ldots) \in H_i(\boldsymbol{C}^{\mathcal{T}})$ such that $h_i^{\mathcal{T}}(t_i) = (\mu_i^1, \mu_i^2, \ldots)$;
- for every belief hierarchy $(\mu_i^1, \mu_i^2, \ldots) \in H_i(\boldsymbol{C}^{\mathcal{T}})$, there is a type $t_i \in T_i$ such that $(\mu_i^1, \mu_i^2, \ldots) = h_i^{\mathcal{T}}(t_i)$.

The proof follows directly from Lemma 5.1 and the fact that the construction above gives a space of belief hierarchies.

It follows from Lemma 4.1 and Theorem 5.2 that each type t_i generates a belief hierarchy $h_i^{\mathcal{T}}(t_i)$ of well-defined depth. This allows us to characterize the events that a type can assign a probability to. Say that $E \subseteq \Theta \times T_j$ is a *kth-order event* for player i if it is expressible in j's $(k-1)$th-order beliefs, i.e., there is an event $B_k \in \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^{\mathcal{T}})$ such that

$$E = \{(s_j, t_j) : (s_j, h_j^{\mathcal{T},k-1}(t_j)) \in B_k\}$$

and there is no $B_{k-1} \in \mathcal{F}_\Theta \times \mathcal{F}_{j,k-2}^{k-2}(\boldsymbol{C}^{\mathcal{T}})$ for which this holds. This gives the following observation:

Corollary 5.3. The following are equivalent:

(a) Type t_i has depth $k < \infty$;

(b) for any $E \subseteq \Theta \times T_j$, the event E belongs to $\mathcal{F}_\Theta \times \Sigma_i(t_i)$ if and only if E is a kth-order event.

So, a type with depth k can form a belief about an event if and only if it is expressible in terms of the other player's $(k-1)$th-order beliefs. The proof follows directly from Lemma 5.1.

5.3. Depth of reasoning

Corollary 5.3 makes the connection between the type space and the space of belief hierarchies: a type with finite depth $k < \infty$ can reason about an event if and only if it is expressible in terms of the other player's $(k-1)$th-order beliefs. While this result can be used to understand how players can reason about higher-order events, it does not tell us how the depth of reasoning can be determined. As I show now, the depth of reasoning can be determined from the type space alone, without making reference to players' belief hierarchies. For example, types from Harsanyi type spaces have an infinite depth, as should be expected:

Observation 1. (Harsanyi type spaces) If $\mathcal{T}^{\mathcal{H}}$ is a Harsanyi type space, and t_i is a type in $\mathcal{T}^{\mathcal{H}}$, then $d_i^{\mathcal{T}^{\mathcal{H}}}(t_i) = \infty$.

The proof follows directly from Lemma 5.1, and is thus omitted. For example, as is well-known, the space $\boldsymbol{C}^{\mathcal{U}}$ of belief hierarchies constructed in Example 4 defines a type space, the so-called universal (Harsanyi) type space (e.g., Mertens and Zamir, 1985), and every type in this type space has an infinite depth of reasoning.

As a second example, it is easy to characterize the type spaces in which all types have the same finite depth k:

Observation 2. (Uniform finite depth) Fix a type space $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$.

(a) Suppose that for each player $i \in N$, every type $t_i \in T_i$ is endowed with the same σ-algebra $\mathcal{F}_j \in \mathcal{S}_j$, and that $\mathcal{F}_a$ does not dominate $\mathcal{F}_b$ or vice versa. Then there is $k < \infty$ such that for each $i \in N$, the σ-algebra $\mathcal{F}_i$ dominates exactly $k-1$ σ-algebras in $\mathcal{S}_j$, and the depth of each type equals k.

(b) Conversely, suppose that each type has the depth $k < \infty$. Then for each player $j \in N$, there is a σ-algebra $\mathcal{F}_j \in \mathcal{S}_j$ such that every type $t_i \in T_i$ is endowed with the σ-algebra $\Sigma_i(t_i) = \mathcal{F}_j$, and $\mathcal{F}_a$ does not dominate $\mathcal{F}_b$ or vice versa.

Again, the proof follows directly from Lemma 5.1. An example of a type space with uniform finite depth is the type space in Figure 2: the σ-algebra with which the types are endowed do not dominate each other, and each σ-algebra dominates only the trivial σ-algebra. It follows that every type has depth 2.

While the observations above apply to type spaces in which every type has the same (finite or infinite) depth, it is also possible to determine the depth of reasoning of a type from a type space when players can be uncertain about each other's depth, essentially by counting the number of σ-algebras that the type's σ-algebra dominates (details available upon request). Therefore, as in the Harsanyi case, the properties of the belief hierarchies can be determined from the type space alone.

6. Common belief

I now illustrate how the framework can be used to gain insight into the question how common belief can be attained. Fix a type space $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$. Every tuple $(\theta, t_a, t_b) \in \Theta \times T_a \times T_b$ is a *state*. For every event $E \subseteq \Theta \times T_a \times T_b$, player i and type $t_i \in T_i$, let

$$E_{t_i} := \{(\theta, t_j) : (\theta, t_a, t_b) \in E\}.$$

Then, define

$$\mathcal{B}_a(E) := \{(\theta, t_a, t_b) \in \Theta \times T_a \times T_b : E_{t_a} \in \mathcal{F}_\Theta \times \Sigma_a(t_a), \beta_a(t_a)(E_{t_a}) = 1\}$$

to be the event that Ann *believes* E (with probability 1), and let the event $\mathcal{B}_b(E)$ that Bob believes E be defined analogously. Thus, a type believes an event if it can reason about it (i.e., E is a measurable event for the type), and it assigns probability 1 to the event. Clearly, the belief operator $\mathcal{B}_i$ coincides with the standard one if $\mathcal{T}$ is a Harsanyi type space (cf. Brandenburger and Dekel, 1987; Monderer and Samet, 1989); see Appendix D.1.1 for a discussion of its properties. Let

$$\mathcal{B}(E) := \mathcal{B}_a(E) \cap \mathcal{B}_b(E)$$

be the event that both players believe the event E, that is, $\mathcal{B}(E)$ is the event that E is *mutual belief*.

The set of states at which E is *mth-order mutual belief* is $\mathcal{B}^m(E)$, where $\mathcal{B}^1(E) := \mathcal{B}(E)$, and $\mathcal{B}^m(E) := \mathcal{B}(\mathcal{B}^{m-1}(E))$. Then, the event that E is *mutual belief at all orders* is

$$\mathcal{MB}(E) := \bigcap_{m=1}^{\infty} \mathcal{B}^m(E).$$

There is *common belief* in a measurable event F at state (θ, t_a, t_b) if there is a set $E \subseteq \Theta \times T_a \times T_b$ such that $(\theta, t_a, t_b) \in E$ and there is a subset $E' \subseteq E$ with the property that for each player i,

$$E \subseteq \mathcal{B}_i(E'), \text{ and}$$
$$E \subseteq \mathcal{B}_i(F).$$

The set of states at which there is common belief in F is denoted by $\mathcal{C}(F)$. If E' can be taken to be equal to E in the above definition (for every state $(\theta, t_a, t_b) \in \mathcal{C}(F)$), then both players believe E whenever it occurs (i.e., $E \subseteq \mathcal{B}_i(E)$ for $i = a, b$).[11]

The next few results demonstrate that mutual belief in an event at all orders implies common belief in the event, while the converse need not hold when players have a finite depth of reasoning, even if their depth is arbitrarily high. I also characterize a condition under which both notions coincide.

Proposition 6.1. For any measurable event F and state (θ, t_a, t_b) of the world, if F is mutual belief at all orders in state (θ, t_a, t_b), then F is common belief at (θ, t_a, t_b).

Proposition 6.1 says that mutual belief in an event F at all orders implies that there is common belief in the event. However, as Example 3 demonstrates, if players have a finite depth of reasoning, there may be common belief in an event F even if there is no mutual belief in the event at some order. The next result shows that this true even if players have an arbitrarily high (but finite) depth of reasoning.[12]

Proposition 6.2. Suppose that the set Θ of states of nature has at least three elements, and that $\mathcal{F}_\Theta$ contains the singletons. For any $k = 2, 3, \ldots$, there is a type space in which all players have depth k and an event F such that there is common belief in F at some state, while there is no state in which F is mutual belief at all orders.

[11] To see why E' can be a strict subset of E in the definition of common belief, consider the type space that differs from the one in Example 3 only in that type t_a^5 for Ann assigns probability $\frac{1}{2}$ to $(\theta_1, \{t_b^1, t_b^2\})$ and probability $\frac{1}{2}$ to $(\theta_2, \{t_b^1, t_b^2\})$, and similarly for type t_b^5 for Bob. It seems intuitive that there is common belief in $F = \{\theta_1, \theta_2\} \times T_a \times T_b$ at $(\theta, t_a^5, t_b^5\}$: both types believe F, believe that the other player believes F, and so on, and indeed there is common belief at that state according to the current definition (use $E = \Theta \times \{t_a^1, \ldots, t_a^5\} \times \{t_b^1, \ldots, t_b^5\}$ and $E' = \Theta \times \{t_a^1, \ldots, t_a^4\} \times \{t_b^1, \ldots, t_b^4\}$). However, the event E itself is not expressible in terms of players' first-order beliefs, so if a stronger definition of common belief were used, under which an event F is common belief at (θ, t_a, t_b) if and only if there is a event E such that $(\theta, t_a, t_b) \in E$ and for each player i, $E \subseteq \mathcal{B}_i(E)$ and $E \subseteq \mathcal{B}_i(F)$, then F is not common belief at (θ, t_a^5, t_b^5).

[12] The condition in Proposition 6.2 that the σ-algebra $\mathcal{F}_\Theta$ contains the singletons is satisfied in all standard cases, e.g., when Θ is finite or Polish and $\mathcal{F}_\Theta$ is the Borel σ-algebra. In fact, all what is needed for the proof is that the σ-algebra $\mathcal{F}_\Theta$ contains three disjoint, nonempty events.

The intuition behind Proposition 6.2 is evident from Example 3: that an event E' is a kth-order event does not imply that the event that she believes E' is a kth-order event. This means it may not possible to attain high-order mutual belief in an event F even if there is common belief in F. An analogous statement does not hold for Harsanyi type spaces, as the next result shows.

Proposition 6.3. If $\mathcal{T}$ is a Harsanyi type space, then, for every event $F \subseteq \Theta \times T_a \times T_b$, there is common belief in F if and only if there is mutual belief in F at all orders, that is, $\mathcal{C}(F) = \mathcal{MB}(F)$.

Monderer and Samet (1989, Prop. 3) prove this result for a slightly different setting. The proof of Proposition 6.3 follows directly from the Tarski-Kantorovitch Theorem (e.g., Ok, 2014), and is therefore omitted. Intuitively, types in a Harsanyi type space that can reason about an event F can reason about the event that all players believe F (that is, if F is measurable for the types, then so is $\mathcal{B}(F)$); applying this repeatedly gives that players can reason about the event that all players believe F, believe that all players believe F, and so on, for any number of iterations. This property is not satisfied when players have a finite depth.

7. Related literature

Finite depth of reasoning. This paper is the first to model players' depth of reasoning by the set of events that they can reason about. This makes it possible to consider the conditions under which players can attain common belief. Kets (2013, 2014) uses the framework developed here to study the equilibria and rationalizable strategies of games, respectively, when players have a finite depth of reasoning. Also in these papers, it is critical that the events that players can reason about are modeled explicitly. Other work either does not model the higher-order beliefs of players explicitly, as in the level-k literature (Crawford et al., 2012), or a player's depth of reasoning is represented by a single number, rather than by the sets of events she can reason about (Strzalecki, 2009; Heifetz and Kets, 2013). In neither setting is it possible to study common belief, rationalizability, or equilibrium.[13]

Alternative ways of modeling that players are somehow bounded in their reasoning about the higher-order beliefs of other players include ambiguous beliefs, incomplete preferences, and unawareness. Ahn (2007) defines type spaces with ambiguous beliefs, but does not apply his framework to model players' depth of reasoning. Di Tillio (2008) considers a class of

[13] Under the solution concepts studied by Strzalecki (2009) and Heifetz and Kets (2013), players can choose an action that is not consistent with rationality and common belief in rationality, unlike in Kets (2013, 2014). That makes these solution concepts quite different from the standard ones.

type structures that allow for incomplete preferences. However, he does not consider type spaces in which the incompleteness of a type's preferences reflect its depth of reasoning. An advantage of the present approach is that it is possible to compute players' expected utility in the standard way, unlike in the settings studied by Ahn and Di Tillio. This is particularly useful in game-theoretic settings. Heifetz and Kets (2013) define type spaces in which players can have a finite depth of reasoning that bears some similarity with models of unawareness (e.g., Heifetz et al., 2006). As noted earlier, standard game-theoretic concepts like equilibrium or common belief cannot be extended to that framework.

Common belief The question whether common belief can be attained has been the focus of a number of papers in logic, linguistics, and game theory (e.g., Lewis, 1969; Aumann, 1976; Clark and Marshall, 1981; Barwise, 1988; Monderer and Samet, 1989; Halpern and Moses, 1990). In Harsanyi type spaces and related belief structures, an event is common belief if and only if it is mutual belief at all orders. This need not be true in more general belief structures. However, while the existing literature has shown that an event need not be common belief even if it is mutual belief at all orders (Halpern and Moses, 1990; Lismont and Mongin, 1995; Heifetz, 1999), this paper is the first to show the opposite can also be true: there are type spaces in which an event can be common belief, even if it is not mutual belief at all orders.

Measurable structures on type sets. One insight of the present paper is that, by choosing the measurable sets on which a type's belief is defined, it is possible to define types that can reason about only finitely many orders of beliefs. Indeed, the technical contribution of this paper is to formulate conditions on the type space that guarantee that the σ-algebra of a type with a finite depth k lumps together precisely the types that induce belief hierarchies that coincide up to order $k-1$ (Assumption 1). The idea that a type's σ-algebra can determine its depth of reasoning fits in with a broader literature that studies how the measurable structure associated with types in Harsanyi type spaces can implicitly impose restrictions on reasoning, that is, on belief hierarchies (e.g., Brandenburger and Keisler, 2006; Friedenberg and Meier, 2012); see Friedenberg and Keisler (2011) for a detailed discussion and further references.

8. Concluding remarks

This paper introduces a novel framework for analyzing situations in which players can have a finite depth of reasoning. Using a richer framework to model the bounds on players' reasoning than the existing literature makes it possible to understand how players with a finite depth of reasoning can attain common belief. In turn, this makes it possible to extend standard game-theoretic concepts, like rationality and common belief in rationality and equilibrium, to settings where players may have a finite depth of reasoning (Kets, 2013, 2014). On the one hand, this suggests that standard game-theoretic concepts need not be abandoned in the face of experimental evidence that players have a finite depth of reasoning. On the other hand, it turns out that there are important differences in terms of predictions; for example, Bayesian-Nash equilibrium in a finite-depth type space corresponds to a refinement of Bayesian-Nash equilibrium in Harsanyi type spaces (Kets, 2013).

An important step for future research is to formalize under which conditions standard game-theoretic concepts can be used, and under which conditions non-equilibrium concepts such as level-k models are more appropriate. The richness of the present framework makes it possible to model the reasoning of players with limited reasoning abilities explicitly, and understand when their thinking is essentially the same as the reasoning of an unbounded reasoner, and when it is not.

Appendix A Multiple players

The results readily extend to the case of more than two players. The construction of the belief hierarchies in Section 4 extends directly. The results involving type spaces requires some care, as the following example illustrates.

Example 6. The set of states of nature is $\Theta = \{\theta_1, \theta_2\}$. There are three players, Ann (a), Bob (b), and Carol (c), each with four types, labeled $t_i^1, t_i^2, t_i^3, t_i^4$, $i = a, b, c$. The σ-algebras in $\mathcal{S}_i$ are the trivial σ-algebra $\mathcal{F}_i^0$ and the σ-algebra $\mathcal{F}_i^1$ generated by the pairs $\{t_i^1, t_i^2\}$, $\{t_i^3, t_i^4\}$. Each type for Ann is now endowed with a product σ-algebra that describes the beliefs it can have about the types for Bob and Carol, and similarly for the types for Bob and Carol. For each player $i = a, b, c$, the types t_i^1 and t_i^2 believe that the state of nature is θ_1; the other types believe that the state of nature is θ_2. Suppose that type t_a^1 has the product σ-algebra $\Sigma_a(t_a^1) = \mathcal{F}_b^1 \times \mathcal{F}_c^0$, and that it assigns probability 1 to the event $\{\theta_1\} \times \{t_b^1, t_b^2\} \times \{t_c^1, t_c^2, t_c^3, t_c^4\}$. Then, the type believes that Bob believes that the state of nature is θ_1 (as both t_b^1 and t_b^2 assign probability 1 to that event), but it does not have a well-articulated belief about Carol's belief about nature. That is, Ann can reason about Bob's first-order beliefs, but not about Carol's. $\triangleleft$

In this example, Ann's "depth" of reasoning depends on the identity of the player she reasons about. To rule this out, I impose a condition not on the σ-algebras on individual type sets (cf. Assumption 1), but on the combinations of σ-algebras, that is, on the collection of product σ-algebras. The results then extend to the multi-player case in a straightforward way.

That is, a type space (on Θ) is a tuple

$$\mathcal{T} = ((T_i, \Sigma_i, \beta_i)_{i \in N}, \Pi),$$

that satisfies Assumption 2 below. As before, the set T_i is a nonempty set of types for player $i \in N$, and Π is a set of product σ-algebras $\boldsymbol{F}$ on the set $T := \prod_{i \in N} T_i$ of type profiles. Given the set Π and a player $i \in N$, define

$$\mathcal{S}_{-i} := \{\boldsymbol{F}_{-i} : \text{ there is } \mathcal{F}_i \text{ such that } \mathcal{F}_i \times \boldsymbol{F}_{-i} \in \Pi\}$$

to be the set of σ-algebras on T_{-i} that are induced by one of the elements of Π. The function Σ_i maps each type $t_i \in T_i$ into a (product) σ-algebra $\Sigma_i(t_i)$ in $\mathcal{S}_{-i}$. The function β_i maps each type t_i into a probability measure $\beta_i(t_i)$ on the product σ-algebra $\mathcal{F}_\Theta \times \Sigma_i(t_i)$ on $\Theta \times T_{-i}$, where $T_{-i} := \prod_{j \neq i} T_j$.

To state Assumption 2, some more notation is needed. Fix product σ-algebras $\boldsymbol{F} = \bigtimes_{j \in N} \mathcal{F}_j$ and $\boldsymbol{F}' = \bigtimes_{j \in N} \mathcal{F}'_j$ on the set T of type profiles, and let $i \in N$. Then, the σ-algebra

$\mathcal{F}_i$ *i-dominates* the product σ-algebra $\boldsymbol{F}'$ if for each event $E \in \mathcal{F}_\Theta \times \boldsymbol{F}'_{-i}$, and $p \in [0,1]$,

$$\{t_i \in T_i : E \in \Sigma_i(t_i), \beta_i(t_i)(E) \geq p\} \in \mathcal{F}_i.$$

If $\mathcal{F}_i$ i-dominates $\boldsymbol{F}'$ for each player $i \in N$, then we say that $\boldsymbol{F}$ *N-dominates* $\boldsymbol{F}'$. If the product σ-algebra $\boldsymbol{F}$ N-dominates itself, that is, $\boldsymbol{F} \succ \boldsymbol{F}$, then the product σ-algebra $\boldsymbol{F}$ is a *self-dominating profile*.

Assumption 2. For every profile $\boldsymbol{F} = \bigtimes_{j \in N} \mathcal{F}_j \in \Pi$ such that $\mathcal{F}_i \neq \{T_i, \emptyset\}$ for some $i \in N$, one of the following holds:

(a) $\boldsymbol{F}$ is self-dominating; or

(b) there is a profile $\boldsymbol{F}' \in \Pi$ such that for each player $j \in N$, the σ-algebra $\mathcal{F}_j$ is the coarsest σ-algebra that j-dominates $\boldsymbol{F}'$.

This condition, which puts restrictions on the set of product σ-algebras, is somewhat stronger than Assumption 1 for the case of two players. However, for every type space for two players that satisfies the weaker condition (Assumption 1), there exists a type space that satisfies Assumption 2 such that types generate the same belief hierarchies in both type spaces. In that sense, there is no loss of generality in adopting the stronger version also for the two-player case.

The results for the two-player case extend if product σ-algebras are substituted for σ-algebras on individual type sets, self-dominating profiles for mutual-dominance pairs, and N-dominance of product σ-algebras for dominance of individual σ-algebras.

Appendix B Details of Example 5

Consider the space $\boldsymbol{C}^*$ of belief hierarchies constructed in Example 5, and recall that the set of belief hierarchies for player i is denoted by H_i^*. By a simple modification of the argument for the case of Harsanyi hierarchies (Example 4),[14] it can be shown that for each player $i \in N$, there is a homeomorphism β_i^* from H_i^* to the set $\Delta(\Theta \times H_j^*, \mathscr{S}_i^*)$ of beliefs about $\Theta \times H_j^*$, where the collection of σ-algebras on $\Theta \times H_j^{\mathcal{U}}$ is given by

$$\mathscr{S}_i^* := \{\mathscr{B}(\Theta) \times \{H_j^*, \emptyset\}, \mathscr{B}(\Theta) \times \mathscr{B}(H_j^*), \mathscr{B}(\Theta) \times \mathcal{F}_j^{*,1}, \mathscr{B}(\Theta) \times \mathcal{F}_j^{*,2}, \ldots\},$$

[14] See, e.g., Mertens et al. (1994) for the Harsanyi case. The only modification relative to the Harsanyi case is that the mth-order beliefs of a player can now be defined on different σ-algebras. However, since a disjoint union of countably many Polish spaces is Polish (e.g., Kechris, 1995), the argument is essentially the same as in the Harsanyi case.

with

$$\mathcal{F}_j^{*,m} := \big\{\{(\mu_j^1, \mu_j^2, \ldots) \in H_j^* : (\mu_j^1, \ldots, \mu_j^m) \in B\} : B \in \mathscr{B}(C_j^{*,m})\big\}$$

the σ-algebra generated by the mth-order belief hierarchies. Hence, define the function Σ_i^* for each player $i \in N$ that assigns to each belief hierarchy $h_i \in H_i^*$ its σ-algebra in the following way: if $\beta_i^*(h_i)$ is defined on $\mathscr{B}(\Theta) \times \mathcal{F}_j \in \mathscr{S}_i^*$, then set $\Sigma_i^*(h_i) := \mathcal{F}_j$. Also, define

$$\mathcal{S}_i^* := \big\{\{H_j^*, \emptyset\}, \mathcal{F}_j^{*,1}, \mathcal{F}_j^{*,2}, \cdots, \mathscr{B}(H_j^*)\big\}.$$

It is easy to show that the σ-algebras in $\mathcal{S}_i^*$ form a proper filtration, i.e.,

$$\{H_j^*, \emptyset\} \subsetneq \mathcal{F}_j^{*,1} \subsetneq \mathcal{F}_j^{*,2} \subsetneq \cdots \subsetneq \mathscr{B}(H_j^*).$$

Moreover, it is easy to check that the σ-algebras in $\mathcal{S}_a$ and $\mathcal{S}_b$ satisfy Assumption 1. In particular, for each $i \in N$,

$$\cdots \succ^* \mathcal{F}_i^{*,m} \succ^* \mathcal{F}_j^{*,m-1} \succ^* \cdots \succ^* \mathcal{F}_i^{*,1} \succ^* \{H_j^*, \emptyset\},$$

while $\mathscr{B}(H_a^*)$ and $\mathscr{B}(H_b^*)$ form a mutual-dominance pair. Thus, $\mathcal{T}^* := (H_i^*, \mathcal{S}_i^*, \Sigma_i^*, \beta_i^*)_{i\in N}$ is a type space.

Using the proof of Lemma 5.1, it can be shown that types $h_i \in H_i^*$ that are endowed with the σ-algebra $\Sigma_i^*(h_i) = \{H_j^*, \emptyset\}$ have depth of reasoning equal to 1, while types h_i with σ-algebra $\Sigma_i^*(h_i) = \mathcal{F}_j^{*,m-1}$ have depth of reasoning m. Types t_i that are endowed with $\Sigma_i^*(h_i) = \mathscr{B}(H_j^*)$ have infinite depth.

Since β_i^* is a homeomorphism, for every $\mu_i \in \Delta(\Theta \times H_j^*, \mathscr{S}_i^*)$, there is a type $h_i \in H_i^*$ such that $\beta_i^*(h_i) = \mu_i$. In particular, for each of the σ-algebras in $\mathcal{S}_i^*$, there is a type with that σ-algebra; and for every $k = \infty, 1, 2, \ldots$, there is a type of depth k.

Appendix C Proofs for Sections 4–5

C.1 Proof of Lemma 4.1

The result follows directly from the coherency condition (ii). If μ_i^{k-1} is defined on the σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_{j,\ell-2}^{k-2}(\boldsymbol{C})$ for $\ell < k$, then any probability measure μ_i^k in $\Delta(\Theta \times C_j^{k-1}, \mathscr{S}_i^k(\boldsymbol{C}))$ that satisfies (ii) (i.e., is such that $\mathrm{marg}_{\Theta\times C_j^{k-2}}\, \mu_i^k = \mu_i^{k-1}$) is defined on the σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_{j,\ell-2}^{k-1}(\boldsymbol{C})$. Similarly, if μ_i^{k-1} is defined on $\mathcal{F}_\Theta \times \{C_j^{k-2}, \emptyset\}$, then any probability measure μ_i^k in $\Delta(\Theta \times C_j^{k-1}, \mathscr{S}_i^k(\boldsymbol{C}))$ that satisfies the coherency condition is defined on $\mathcal{F}_\Theta \times \{C_j^{k-1}, \emptyset\}$. Finally, if μ_i^{k-1} is defined on the σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-2}^{k-2}(\boldsymbol{C})$, then a probability measure μ_i^k in $\Delta(\Theta \times C_j^{k-1}, \mathscr{S}_i^k(\boldsymbol{C}))$ is coherent with μ_i^{k-1} only if it is defined on $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C})$ or on $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-2}^{k-1}(\boldsymbol{C})$. □

C.2 Proof of Lemma 5.1

It will be useful to introduce some notation and state some preliminary results. For any nonempty set X and any nonempty collection $\mathcal{E}$ of subsets of X, let $\sigma(\mathcal{E})$ be the coarsest σ-algebra on X that contains the sets in $\mathcal{E}$, that is, $\sigma(\mathcal{E})$ is the σ-algebra generated by $\mathcal{E}$.

The following preliminary result says that taking inverse images preserves σ-algebras:

Lemma C.1. Let $f : X \to Y$ be a function from X into Y, and let $\mathcal{E}$ be a nonempty collection of subsets of Y. Then,

$$\sigma\big(\{f^{-1}(E) : E \in \mathcal{E}\}\big) = \{f^{-1}(E) : E \in \sigma(\mathcal{E})\}.$$

The proof is standard, and thus omitted. To state the second preliminary result, let X be some nonempty set, and let $\mathscr{S}$ be a nonempty collection of σ-algebras on X. As before, $\Delta(X, \mathscr{S})$ is 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

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

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

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

and let $\sigma(\mathcal{A})$ and $\sigma(\mathcal{A}')$ be the σ-algebras on $\Delta(X, \mathscr{S})$ generated by $\mathcal{A}$ and $\mathcal{A}'$, respectively. In general, these two σ-algebras can be different. However, as I show now, in an important class of cases, $\sigma(\mathcal{A})$ and $\sigma(\mathcal{A}')$ coincide:

Lemma C.2. Suppose $\mathscr{S}$ is countable and forms a filtration, and suppose there is $\underline{\mathcal{F}} \in \mathscr{S}$ such that $\underline{\mathcal{F}} \subseteq \mathcal{F}$ for all $\mathcal{F} \in \mathscr{S}$. Then $\sigma(\mathcal{A}) = \sigma(\mathcal{A}')$.

Proof. The first step is to show that $\sigma(\mathcal{A}') \subseteq \sigma(\mathcal{A})$. It suffices to show that $\mathcal{A}' \subseteq \sigma(\mathcal{A})$. 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, $F' \in \sigma(\mathcal{A})$.

The next step is to demonstrate that $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{A}')$. 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 $E_2 \in \mathcal{F}_2 \backslash \mathcal{F}_1$, 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,

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

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 $\sigma(\mathcal{A}')$. □

We are now ready to prove Lemma 5.1. The proof is by induction. As part of the proof, I construct an inductive structure, in the following way. For each $k = 1, 2, \ldots$, it is possible to define a σ-algebra $\mathcal{Q}_i^k$ on T_i for each player $i \in N$. (The σ-algebras $\mathcal{Q}_i^k$ will be the σ-algebras $\sigma(h_i^{\mathcal{T},k})$, defined below.) I then show that the σ-algebra of each type is either coincides with $\mathcal{Q}_i^m$ for some $m < k$, or is a superset of $\mathcal{Q}_i^k$. This gives the inductive structure needed to prove the result. Note that Assumption 1 does not, by itself, provide such an order. In particular, it does not imply that $\mathcal{S}_i$ is a (countable) filtration.[15]

To prove the result, define $\sigma(h_i^{\mathcal{T},1})$ to be the σ-algebra on T_i that is generated by the function $h_i^{\mathcal{T},1}$, that is,

$$\sigma(h_i^{\mathcal{T},1}) := \{\{t_i \in T_i : h_i^{\mathcal{T},1}(t_i) \in B\} : B \in \mathcal{F}_{i,1}^1(\boldsymbol{C}^{\mathcal{T}})\}.$$

[15]Indeed, it is possible to have $\mathcal{F}_i, \mathcal{F}_i' \in \mathcal{S}_i$ such that $\mathcal{F}_i \not\subseteq \mathcal{F}_i'$ and vice versa, or to have $\mathcal{F}_i^1, \mathcal{F}_i^2, \ldots, \mathcal{F}_i^{-1}, \mathcal{F}_i^{-2}, \ldots \in \mathcal{S}_i$ such that $\cdots \succ^* \mathcal{F}_i^{-1} \succ^* \mathcal{F}_i \succ^* \mathcal{F}_i^1 \succ^* \mathcal{F}_i^2 \succ^* \cdots$. It follows from the proof that for any such σ-algebra $\mathcal{F}_i$, the σ-algebra $\mathcal{F}_i$ contains $\mathcal{Q}_i^m$ for all m, so that such σ-algebras do not affect the inductive structure.

It will be notationally convenient to introduce the function $h_i^{\mathcal{T},0} : T_i \to \{x\}$, where x is an arbitrary singleton, defined in the obvious way; thus, the σ-algebra $\sigma(h_i^{\mathcal{T},0})$ on T_i that is generated by the function $h_i^{\mathcal{T},0}$ is simply the trivial σ-algebra $\{T_i, \emptyset\}$.

Lemmas C.3–C.5 help order the σ-algebras with which types are endowed, using the σ-algebras $\sigma(h_i^{\mathcal{T},0})$ and $\sigma(h_i^{\mathcal{T},1})$. Lemma C.3 is an auxiliary result that gives a useful characterization of $\sigma(h_i^{\mathcal{T},1})$.

Lemma C.3. The σ-algebra $\sigma(h_i^{\mathcal{T},1})$ is the coarsest σ-algebra on T_i that dominates $\sigma(h_j^{\mathcal{T},0})$, i.e., $\sigma(h_i^{\mathcal{T},1}) \succcurlyeq^* \sigma(h_j^{\mathcal{T},0})$.

Proof. Note that

$$\begin{aligned} \sigma\big(h_i^{\mathcal{T},1}\big) &= \Big\{\big\{t_i \in T_i : \mathrm{marg}_\Theta \beta_i(t_i) \in B\big\} \colon B \in \mathcal{F}^1_{i,1}(\boldsymbol{C}^{\mathcal{T}})\Big\} \\ &= \sigma\Big(\Big\{\big\{t_i \in T_i : \mathrm{marg}_\Theta \beta_i(t_i)(E) \geq p\big\} : E \in \mathcal{F}_\Theta, p \in [0,1]\Big\}\Big) \\ &= \sigma\Big(\Big\{\big\{t_i \in T_i : E' \in \mathcal{F}_\Theta \times \Sigma_i(t_i), \beta_i(t_i)(E') \geq p\big\} : E' \in \mathcal{F}_\Theta \times \sigma(h_j^{\mathcal{T},0}), p \in [0,1]\Big\}\Big), \end{aligned}$$

where the second equality uses Lemma C.1. $\square$

Lemma C.4. For each type $t_i \in T_i$, the σ-algebra $\Sigma_i(t_i)$ is a strict subset of $\sigma(h_j^{\mathcal{T},1})$ (i.e., $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},1})$) or it is a superset (i.e., $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},1})$).

Proof. If $\Sigma_i(t_i) = \{T_j, \emptyset\}$, then clearly, $\Sigma_i(t_i) \subseteq \sigma(h_j^{\mathcal{T},1})$. If $\Sigma_i(t_i) \neq \{T_j, \emptyset\}$, then, by Assumption 1, there is $\mathcal{F}_i \in \mathcal{S}_i$ such that $\Sigma_i(t_i)$ dominates $\mathcal{F}_i$. Since any σ-algebra $\mathcal{F}_i \in \mathcal{S}_i$ is at least as fine as the trivial σ-algebra $\{T_i, \emptyset\}$, i.e., $\mathcal{F}_i \supseteq \{T_i, \emptyset\}$, $\Sigma_i(t_i)$ dominates $\{T_i, \emptyset\}$. But, by Lemma C.3, the σ-algebra $\sigma(h_j^{\mathcal{T},1})$ is the coarsest σ-algebra that dominates $\{T_i, \emptyset\}$. Hence, $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},1})$. $\square$

Lemma C.5. For each $t_i \in T_i$, if $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},1})$, then $\Sigma_i(t_i) = \sigma(h_j^{\mathcal{T},0})$.

Proof. Suppose $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},1})$. Then, by Assumption 1, either $\Sigma_i(t_i) = \{T_j, \emptyset\} = \sigma(h_j^{\mathcal{T},0})$, or there is a σ-algebra $\mathcal{F}_i \in \mathcal{S}_i$ such that $\Sigma_i(t_i)$ dominates $\mathcal{F}_i$. If there is such a σ-algebra $\mathcal{F}_i \in \mathcal{S}_i$, then an argument similar to the one in the proof of Lemma C.4 gives that $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},1})$, a contradiction. $\square$

For $k > 1$, assume inductively that for any $\ell \leq k-1$ and $i \in N$, the set $C_i^{\mathcal{T},\ell}$ has been defined and that the functions $h_i^{\mathcal{T},\ell}$ are well-defined. Let

$$\sigma(h_i^{\mathcal{T},\ell}) = \{\{t_i \in T_i : h_i^{\mathcal{T},\ell}(t_i) \in B\} : B \in \mathcal{F}^\ell_{i,\ell}(\boldsymbol{C}^{\mathcal{T}})\}$$

be the σ-algebra on T_i that is generated by the function $h_i^{\mathcal{T},\ell}$. Also, assume that the following hold:

- the σ-algebra $\sigma(h_i^{\mathcal{T},\ell})$ is the coarsest σ-algebra on T_i that dominates the σ-algebra $\sigma(h_j^{\mathcal{T},\ell-1})$;
- for each type $t_i \in T_i$, the σ-algebra $\Sigma_i(t_i)$ is a strict subset of $\sigma(h_j^{\mathcal{T},\ell})$ (i.e., $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},\ell})$) or it is a superset (i.e., $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},\ell})$).
- for each type $t_i \in T_i$, if $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},\ell})$, then there is $m < \ell$ such that $\Sigma_i(t_i) = \sigma(h_j^{\mathcal{T},m})$.

The next result shows that the function $h_i^{\mathcal{T},k}$ is well-defined:

Lemma C.6. For each type $t_i \in T_i$, the kth-order belief hierarchy $h_i^{\mathcal{T},k}(t_i)$ is in $C_i^{\mathcal{T},k-1} \times \Delta(\Theta \times C_j^{\mathcal{T},k-1}, \mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}}))$.

Proof. By the induction hypothesis, the claim holds if and only if $\mu_i^k(t_i) = \beta_i(t_i) \circ (\mathrm{Id}_\Theta, h_j^{\mathcal{T},k-1})^{-1}$ is a probability measure in $\Delta(\Theta \times C_j^{\mathcal{T},k-1}, \mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}}))$, where Id_Θ is the identity function on Θ. By the induction hypothesis, $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},k-1})$ or $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},k-1})$. First suppose $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},k-1})$. Then, for each $E \in \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^{\mathcal{T}})$, it follows that $(\mathrm{Id}_\Theta, h_j^{\mathcal{T},k-1})^{-1}(E) \in \mathcal{F}_\Theta \times \Sigma_i(t_i)$. Hence, $\mu_i^k(t_i)$ is a probability measure on $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^{\mathcal{T}}) \in \mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}})$. Next suppose that $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},k-1})$. By the induction hypothesis, there is $m < k-1$ such that $\Sigma_i(t_i) = \sigma(h_j^{\mathcal{T},m})$; let m' be the largest $m' < k-1$ for which this holds. By a similar argument as before, it follows that $\mu_i^k(t_i)$ is a probability measure on $\mathcal{F}_\Theta \times \mathcal{F}_{j,m'}^{k-1}(\boldsymbol{C}^{\mathcal{T}})$, and this σ-algebra belongs to $\mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}})$. $\square$

By Lemma C.6, the σ-algebra

$$\sigma(h_i^{\mathcal{T},k}) := \{\{t_i \in T_i : h_i^{\mathcal{T},k}(t_i) \in B\} : B \in \mathcal{F}_{i,k}^k(\boldsymbol{C}^{\mathcal{T}})\}$$

on T_i that is generated by the function $h_i^{\mathcal{T},k}$ is well-defined. The next step is to establish the analogues of Lemmas C.3–C.5 for general k, to order the σ-algebras on the type sets.

Lemma C.7. The σ-algebra $\sigma(h_i^{\mathcal{T},k})$ is the coarsest σ-algebra on T_i that dominates $\sigma(h_j^{\mathcal{T},k-1})$, that is, $\sigma(h_i^{\mathcal{T},k}) \succeq^* \sigma(h_j^{\mathcal{T},k-1})$.

Proof. By Lemma C.1, $\sigma(h_i^{\mathcal{T},k})$ is the coarsest σ-algebra that contains the sets in $\sigma(h_i^{\mathcal{T},k-1})$ as well as the sets

$$\left\{t_i \in T_i : \Sigma(\mu_i^k(t_i)) = \mathcal{F}, \mu_i^k(t_i)(E) \geq p\right\} \tag{C.1}$$

for $\mathcal{F} \in \mathscr{S}_i^k(\boldsymbol{C}^\mathcal{T})$, $E \in \mathcal{F}$, and $p \in [0,1]$. Since beliefs are coherent, that is, for all $\ell \leq k-1$,

$$\operatorname{marg}_{\Theta \times C_j^{\mathcal{T},\ell-1}} \mu_i^k(t_i) = \mu_i^\ell(t_i),$$

the σ-algebra $\sigma(h_i^{\mathcal{T},k})$ is the σ-algebra generated by the sets in (C.1). Since $\mathscr{S}_i^k(\boldsymbol{C}^\mathcal{T})$ is a countable filtration with a minimal element, it follows from Lemma C.2 that $\sigma(h_i^{\mathcal{T},k})$ is generated by the sets

$$\{t_i \in T_i : E \in \Sigma(\mu_i^k(t_i)), \mu_i^k(t_i)(E) \geq p\}$$

for $\mathcal{F} \in \mathscr{S}_i^k(\boldsymbol{C}^\mathcal{T})$, $E \in \mathcal{F}$, and $p \in [0,1]$. Using that for each $\mathcal{F} \in \mathscr{S}_i^k(\boldsymbol{C}^\mathcal{T})$, $\mathcal{F} \subseteq \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^\mathcal{T})$, it follows that $\sigma(h_i^{\mathcal{T},k})$ is generated by the sets

$$\{t_i \in T_i : E \in \Sigma(\mu_i^k(t_i)), \mu_i^k(t_i)(E) \geq p\}$$

for $E \in \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^\mathcal{T})$, and $p \in [0,1]$, or, equivalently, the sets

$$\{t_i \in T_i : E' \in \mathcal{F}_\Theta \times \Sigma_i(t_i), \beta_i(t_i)(E') \geq p\}$$

for $E \in \mathcal{F}_\Theta \times \sigma(h_j^{\mathcal{T},k-1})$, and $p \in [0,1]$. Hence, $\sigma(h_i^{\mathcal{T},k}) \succeq^* \sigma(h_j^{\mathcal{T},k-1})$. $\square$

Lemma C.8. For each type $t_i \in T_i$, the σ-algebra $\Sigma_i(t_i)$ is a strict subset of $\sigma(h_j^{\mathcal{T},k})$ (i.e., $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},k})$) or it is a superset (i.e., $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},k})$).

Proof. If $\mathcal{F}_i = \{T_i, \emptyset\}$, then clearly $\mathcal{F}_i \subseteq \sigma(h_i^{\mathcal{T},k})$. So suppose $\mathcal{F}_i \neq \{T_i, \emptyset\}$. By Assumption 1, one of the following holds:

(a) $\mathcal{F}_i$ is part of a mutual-dominance pair, that is, there is $\mathcal{F}_j \in \mathcal{S}_j$ such that $\mathcal{F}_i \succ \mathcal{F}_j$ and vice versa; or

(b) $\mathcal{F}_i$ is part of a finite chain, that is, there exist $m < \infty$ and (distinct) σ-algebras $\mathcal{F}_j^1, \mathcal{F}_j^3, \ldots, \mathcal{F}_j^m \in \mathcal{S}_j$ and $\mathcal{F}_i^2, \mathcal{F}_i^4 \ldots, \mathcal{F}_i^m \in \mathcal{S}_i$ such that

$$\mathcal{F}_i \succeq^* \mathcal{F}_j^1 \succeq^* \mathcal{F}_i^2 \succeq^* \cdots \succeq^* \mathcal{F}_j^m = \{T_j, \emptyset\}$$

if m is odd, and

$$\mathcal{F}_i \succeq^* \mathcal{F}_j^1 \succeq^* \mathcal{F}_i^2 \succeq^* \cdots \succeq^* \mathcal{F}_i^m = \{T_i, \emptyset\}$$

if m is even; or

(c) $\mathcal{F}_i$ is part of a cycle or infinite chain, that is, there exist σ-algebras $\mathcal{F}_j^1, \mathcal{F}_j^3, \ldots \in \mathcal{S}_j$ and $\mathcal{F}_i^2, \mathcal{F}_i^4, \ldots \in \mathcal{S}_i$ (where $\mathcal{F}_n^\ell, \mathcal{F}_n^m$ are not necessarily distinct, $n \in N$) such that

$$\mathcal{F}_i \succeq^* \mathcal{F}_j^1 \succeq^* \mathcal{F}_i^2 \succeq^* \mathcal{F}_j^3 \succeq^* \cdots.$$

I claim that if (a) or (c) is the case, then $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},k})$. I present the argument for case (c); the argument for (a) is similar and thus omitted. Note that $\mathcal{F}_j^1 \supseteq \sigma(h_j^{\mathcal{T},0}) = \{T_j, \emptyset\}$. By the induction hypothesis, therefore, $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},1})$. By a similar argument, $\mathcal{F}_j^1 \supseteq \sigma(h_j^{\mathcal{T},1})$. Since $\mathcal{F}_i$ dominates $\mathcal{F}_j^1$, it follows from the induction hypothesis and Lemma C.7 that $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},2})$. Repeating this argument gives the desired result.

It remains to consider (b). Consider the case that m is odd; the argument for the case that m is even is similar. If $m \leq k$, then by the induction hypotheses and Lemma C.7, it follows that $\mathcal{F}_i = \sigma(h_i^{\mathcal{T},m}) \subseteq \sigma(h_i^{\mathcal{T},k})$. If $m > k$, then it is the case that $\mathcal{F}_j^{m-k} = \sigma(h_j^{\mathcal{T},k})$ or $\mathcal{F}_i^{m-k} = \sigma(h_i^{\mathcal{T},k})$, depending on whether k is odd or even. I discuss the case that $\mathcal{F}_j^{m-k} = \sigma(h_j^{\mathcal{T},k})$; the argument for the case $\mathcal{F}_i^{m-k} = \sigma(h_i^{\mathcal{T},k})$ is similar. Since $\mathcal{F}_i^{m-k-1}$ dominates $\mathcal{F}_j^{m-k} \supseteq \sigma(h_j^{\mathcal{T},k-1})$, it follows from Lemma C.7 that $\mathcal{F}_i^{m-k-1} \supseteq \sigma(h_i^{\mathcal{T},k}) \supseteq \sigma(h_i^{\mathcal{T},k-1})$. By a similar argument, $\mathcal{F}_j^{m-k-2} \supseteq \sigma(h_j^{\mathcal{T},k}) \supseteq \sigma(h_j^{\mathcal{T},k-1})$. Repeating this argument gives $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},k})$. □

Lemma C.9. For each $t_i \in T_i$, if $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},k})$, then there is $m < k$ such that $\Sigma_i(t_i) = \sigma(h_j^{\mathcal{T},m})$.

Proof. Suppose $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},k})$. If $\mathcal{F}_i \subsetneq \sigma(h_i^{\mathcal{T},k-1})$, then the result follows from the induction hypothesis. So suppose $\mathcal{F}_i$ is not a strict subset of $\sigma(h_i^{\mathcal{T},k-1})$. By the induction hypothesis, $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},k-1})$. If $\mathcal{F}_i = \sigma(h_i^{\mathcal{T},k-1})$, then we are done. So suppose $\mathcal{F}_i \supsetneq \sigma(h_i^{\mathcal{T},k-1})$. It suffices to show that the joint hypothesis $\mathcal{F}_i \subsetneq \sigma(h_i^{\mathcal{T},k})$ and $\mathcal{F}_i \supsetneq \sigma(h_i^{\mathcal{T},k-1})$ leads to a contradiction. To derive a contradiction, use an argument similar to the one in the proof of Lemma C.8 to show that $\mathcal{F}_i \subsetneq \sigma(h_i^{\mathcal{T},k})$ implies that $\mathcal{F}_i$ is not part of a mutual-dominance pair, cycle or infinite chain. It then follows from Assumption 1 that $\mathcal{F}_i$ is part of a finite chain. But then $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},k})$, or $\mathcal{F}_i = \sigma(h_i^{\mathcal{T},m})$ for some $m \leq k-1$, a contradiction. □

This completes the induction. For each player $i \in N$ and $k = 1, 2, \ldots$, the function $h_i^{\mathcal{T},k} : T_i \to C_i^{\mathcal{T},k-1} \times \Delta(\Theta \times C_j^{\mathcal{T},k-1}, \mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}}))$ is well-defined. Also, note that for each $t_i \in T_i$, either $\Sigma_i(t_i) = \sigma(h_j^{\mathcal{T},k-1}) \subsetneq \sigma(h_j^{\mathcal{T},k})$, or $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},m})$ for all m. □

Appendix D Proofs for Section 6

D.1 Preliminary results

It will be useful to start with some auxiliary results. I first discuss the properties of the belief operator. I then provide an alternative characterization of the set of states at which there is common belief in an event.

D.1.1 Properties of the belief operator

The belief operator satisfies the standard properties (Monderer and Samet, 1989). Let $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i=a,b}$ be a type space. To state the properties, it will be useful to define the negation of the belief operators $\mathcal{B}_i$: for each player i and event $G \subseteq \Theta \times T_a \times T_b$, write

$$\mathcal{NB}_i(G) := \{(\theta, t_a, t_b) : G_{t_i} \notin \mathcal{F}_\Theta \times \Sigma_i(t_i), \text{ or } \beta_i(t_i)(G_{t_i}) \neq 1\}$$

for the complement $(\Theta \times T_a \times T_b) \setminus \mathcal{B}_i(G)$ of $\mathcal{B}_i(G)$.

Lemma D.1. (Positive introspection) For each event $G \subseteq \Theta \times T_a \times T_b$ and player i, $\mathcal{B}_i(G) = \mathcal{B}_i(\mathcal{B}_i(G))$.

Lemma D.2. (Negative introspection) For each event $G \subseteq \Theta \times T$ and player i, $\mathcal{NB}_i(G) \subseteq \mathcal{B}_i(\mathcal{NB}_i(G))$.

Lemma D.3. (Monotonicity) Let $G, G' \subseteq \Theta \times T_a \times T_b$ such that $G \subseteq G'$. Suppose that for each player $i \in N$ and type $t_i \in T_i$ such that $(\theta, t_i, t_j) \in \mathcal{B}_i(G)$ for some $(\theta, t_j) \in \Theta \times T_j$, $G'_{t_i} \in \mathcal{F}_\Theta \times \Sigma_i(t_i)$. Then, for each player i, $\mathcal{B}_i(G) \subseteq \mathcal{B}_i(G')$.

Lemma D.4. (Conjunction) Let $E_1, E_2, \ldots \subseteq \Theta \times T_a \times T_b$ and fix $i = a, b$. Then, $\bigcap_{k=1}^{\infty} \mathcal{B}_i(E_k) \subseteq \mathcal{B}_i\big(\bigcap_{k=1}^{\infty} E_k\big)$. If $\mathcal{T}$ is a Harsanyi type space, with each type t_i endowed with the σ-algebra $\Sigma_i(t_i) = \mathcal{F}_j^{\mathcal{H}}$ on T_j, and $E_1, E_2, \ldots$ are events in $\mathcal{F}_\Theta \times \mathcal{F}_a^{\mathcal{H}} \times \mathcal{F}_b^{\mathcal{H}}$, then the reverse inclusion also holds, that is, $\mathcal{B}_i\big(\bigcap_{k=1}^{\infty} E_k\big) \subseteq \bigcap_{k=1}^{\infty} \mathcal{B}_i(E_k)$.

The proofs are standard and thus omitted.

D.1.2 Common belief as a fixed point

Common belief in an event can be characterized as the greatest fixed point of a mapping. For every event $E \subseteq \Theta \times T_a \times T_b$, define the function f_E from the power set on $\Theta \times T_a \times T_b$ into itself by:

$$f_E(A) := \{(\theta, t_a, t_b) \in \Theta \times T_a \times T_b : \text{ there is } Q \subseteq E \cap A \text{ s.t. } Q_{t_i} \in \mathcal{F}_\Theta \times \Sigma_i(t_i), \beta_i(t_i)(Q_{t_i}) = 1 \text{ for } i = a, b\}.$$

Clearly, the empty set is a fixed point of f_E, so the function f_E has a fixed point. It follows from Tarski's fixed point theorem that f_E has a unique greatest fixed point, given by

$$gfp(E) = \bigcup \{X : X = f_E(X)\}.$$

The greatest fixed point of f_E is in fact given by the event that E is common belief:

Lemma D.5. The set $\mathcal{C}(E)$ is the greatest fixed point of f_E.

Proof. Suppose $X \subseteq \Theta \times T_a \times T_b$ is such that $X = f_E(X)$. I claim that $X \subseteq \mathcal{C}(E)$. Let $(\theta, t_a, t_b) \in X$. Since $X = f_E(X)$, there is $Q \subseteq E \cap X$ such that for all $i = a, b$, $Q_{t_i} \in \mathcal{F}_\Theta \times \Sigma_i(t_i)$ and $\beta_i(t_i)(Q_{t_i}) = 1$. Define $F := X$, so that $(\theta, t_a, t_b) \in F$. Also, let $F' := Q \subseteq F$. It follows from the monotonicity of the belief operator (Lemma D.3) that $\beta_i(t_i)(E_{t_i}) = 1$. It is then immediate that $(\theta, t_a, t_b) \in \mathcal{C}(E)$.

To show the reverse inclusion, let $(\theta, t_a, t_b) \in \mathcal{C}(E)$. Then there is $F \subseteq \Theta \times T_a \times T_b$ and $F' \subseteq F$ such that for each player $i = a, b$, $F \subseteq \mathcal{B}_i(E)$ and $F \subseteq \mathcal{B}_i(F')$. Without loss of generality, assume that F is a product event, i.e., $F = \Theta \times G_a \times G_b$ for $G_i \subseteq T_i$ for $i = a, b$.[16] By construction, for $i = a, b$, $[E \cap F']_{t_i}$ is in $\mathcal{F}_\Theta \times \Sigma_i(t_i)$ and $\beta_i(t_i)([E \cap F']_{t_i}) = 1$. It follows that $F \subseteq f_E(F)$ (set $Q := E \cap F'$ in the definition of f_E). To show the other direction, suppose $(\theta, t_a, t_b) \in f_E(F)$. Then there is $Q \subseteq E \cap F$ such that for each player $i = a, b$, Q_{t_i} is in $\mathcal{F}_\Theta \times \Sigma_i(t_i)$ and $\beta_i(t_i)(Q_{t_i}) = 1$. It follows that $Q_{t_i} \neq \emptyset$ for $i = a, b$, and thus $(\theta, t_a, t_b) \in Q \subseteq F$ (recall that F is a product event). $\square$

D.2 Proof of Proposition 6.1

Suppose $(\theta, t_a, t_b) \in \mathcal{MB}(F) = \bigcap_m \mathcal{B}^m(F)$, and let $E := \mathcal{MB}(F)$. Hence, by definition, $(\theta, t_a, t_b) \in E$. Then, for $i = a, b$,

$$E = \bigcap_{m=1}^{\infty} \mathcal{B}^m(F) \subseteq \mathcal{B}(F) \subseteq \mathcal{B}_i(F).$$

Also,

$$E = \bigcap_{m=1}^{\infty} \mathcal{B}^m(F) \subseteq \bigcap_{m=1}^{\infty} \mathcal{B}_i(\mathcal{B}^{m-1}(F)) \subseteq \mathcal{B}_i\big(\bigcap_{m=1}^{\infty} \mathcal{B}_i^{m-1}(F)\big) \subseteq \mathcal{B}_i\big(\bigcap_{m=1}^{\infty} \mathcal{B}_i^{m}(F)\big) = \mathcal{B}_i(E),$$

where $\mathcal{B}^0(F) := F$, and where the second and third inclusion use the conjunction and monotonicity property (Appendix D.1.1) of the belief operator, respectively. $\square$

D.3 Proof of Proposition 6.2

Fix $k \geq 2$. I construct a type space $\mathcal{T}^k$ in which all types have depth k, and then show that it has the desired properties.

[16] This is without loss of generality in the sense that if there is a non-product event F such that $(\theta, t_a, t_b) \in F$ and for each player $i = a, b$, $F \subseteq \mathcal{B}_i(E)$ and $F \subseteq \mathcal{B}_i(F')$, then it is possible to construct a product event with the same properties.

By assumption, the set Θ of states of nature has at least three elements; fix two distinct elements $\theta_1, \theta_2 \in \Theta$. I construct a type set from a set of belief hierarchies.

For $i = a, b$, let $T_i^1 := \Delta(\Theta, \mathcal{F}_\Theta)$. For $k = 1$, define $\mathcal{F}_{i,0}^1 := \{T_i^1, \emptyset\}$ to be the trivial σ-algebra on T_i^1; and for $k > 1$, define $\mathcal{F}_{i,k-1}^1$ to be the usual σ-algebra on T_i^1 (Section 3). Define

$$\Delta_i^1 := \{\mu \in T_i^1 : \mu(\{\theta_1, \theta_2\}) = 1\}$$

to be the set of probability measures that assign probability 1 to $\{\theta_1, \theta_2\}$; note that Δ_i^1 is measurable, that is, $\Delta_i^1 \in \mathcal{F}_{i,1}^1$. Let C_i^1, C_i^2 be nonempty measurable (in $\mathcal{F}_{i,1}^1$) subsets of Δ_i^1 such that $C_i^1 \cap C_i^2 = \emptyset$ and $C_i^1 \cup C_i^2 = \Delta_i^1$. That is, $\{C_i^1, C_i^2\}$ is a measurable partition of Δ_i^1. (For example, take $C_i^1 := \{\mu \in T_i^1 : \mu(\theta_1) = 1\}$ and $C_i^2 := \{\mu \in T_i^1 : \mu(\theta_2) = 1\}$.) Then, define

$$\begin{aligned} T_i^2 := \big\{(\mu_i^1, \mu_i^2) \in T_i^1 \times \Delta(\Theta \times T_j^1, \mathcal{F}_\Theta \times \mathcal{F}_{j,1}^1) : \mathrm{marg}_\Theta \mu_i^2 = \mu_i^1 \text{ and} \\ \mu_i^1 \in C_i^1 \Longrightarrow \mu_i^2(C_j^1) = 1; \mu_i^1 \notin C_i^1 \Longrightarrow \mu_i^2(T_j^1 \setminus C_j^1) = 1\big\}. \end{aligned}$$

to be the set of second-order belief hierarchies such that the belief hierarchies with first-order belief in C_i^1 believe that the other player has a first-order belief in C_j^1, and where the other belief hierarchies believe that the other player's first-order belief is not in C_j^1. As before, the condition that the marginal of μ_i^2 on Θ coincides with μ_i^1 is a standard coherency condition that requires that beliefs of different orders do not contradict each other (Mertens and Zamir, 1985; Brandenburger and Dekel, 1993). If $k \geq 3$, let $\mathcal{F}_{i,k-1}^2$ be the σ-algebra on T_i^2 induced by the product σ-algebra on $T_i^1 \times \Delta(\Theta \times T_j^1)$. Otherwise, if $k = 2$, define

$$\mathcal{F}_{i,k-1}^2 := \big\{\{(\mu_i^1, \mu_i^2) \in T_i^2 : \mu_i^1 \in B\} : B \in \mathcal{F}_{i,1}^1\big\},$$

and if $k = 1$, let

$$\mathcal{F}_{i,k-1}^2 := \{T_i^2, \emptyset\}$$

be the trivial σ-algebra.

For $m > 2$, suppose that for each player i, the set T_i^{m-1} and σ-algebra $\mathcal{F}_{i,k-1}^{m-1}$ on T_i^{m-1} have been defined. For each player $i \in N$, define, inductively,

$$T_i^m := \big\{(\mu_i^1, \dots \mu_i^m) \in T_i^{m-1} \times \Delta(\Theta \times T_j^{m-1}, \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{m-1}) : \mathrm{marg}_{\Theta \times T_j^{m-2}} \mu_i^m = \mu_i^{m-1}\big\}$$

to be the set of mth-order belief hierarchies, where the condition on the marginal is again a standard coherency condition. Define the σ-algebra $\mathcal{F}_{i,k-1}^m$ on T_i^m as follows. If $k \geq m+1$, the σ-algebra $\mathcal{F}_{i,k-1}^m$ is the relative σ-algebra on T_i^m induced by the product σ-algebra. If $k = 1$, $\mathcal{F}_{i,k-1}^m$ is the trivial σ-algebra on T_i^m. Otherwise, if $k = 2, \dots, m$,

$$\mathcal{F}_{i,k-1}^m := \big\{\{(\mu_i^1, \dots, \mu_i^m) \in T_i^m : (\mu_i^1, \dots, \mu_i^{k-1}) \in B\} : B \in \mathcal{F}_{i,k-1}^{k-1}\big\}.$$

In words, if $k \leq m$, $\mathcal{F}^m_{i,k-1}$ is the σ-algebra on the mth-order belief hierarchies for i generated by events that can be expressed in terms of her $(k-1)$th-order belief hierarchies.

Define

$$T_i := \big\{(\mu_i^1, \mu_i^2, \dots) : (\mu_i^1, \dots, \mu_i^m) \in T_i^m \text{ for all } m\big\}.$$

By standard arguments, T_i is nonempty. Also, define

$$\begin{aligned} \mathcal{F}^*_{i,0} =& \{T_i, \emptyset\} \\ \mathcal{F}^*_{i,1} =& \big\{\{(\mu_i^1, \mu_i^2, \dots) \in T_i : \mu_i^1 \in B\} : B \in \mathcal{F}^1_{i,1}\big\} \\ & \dots \\ \mathcal{F}^*_{i,k-1} =& \big\{\{(\mu_i^1, \mu_i^2, \dots) \in T_i : (\mu_i^1, \dots, \mu_i^{k-1}) \in B\} : B \in \mathcal{F}^{k-1}_{i,k-1}\big\}. \end{aligned}$$

That is, $\mathcal{F}^*_{i,k-1}$ is the σ-algebra on T_i that is generated by the events that are expressible in i's $(k-1)$th-order belief hierarchies. For $t_i = (\mu_i^1, \mu_i^2, \dots) \in T_i$, let $\Sigma_i(t_i) := \mathcal{F}^*_{j,k-1}$, and define the probability measure $\beta_i(t_i)$ on $\mathcal{F}_\Theta \times \Sigma_i(t_i)$ by

$$\beta_i(t_i)(G) := \mu_i^k(\{(\theta, \mu_j^1, \dots, \mu_j^{k-1}) : (\theta, \mu_j^1, \dots, \mu_j^{k-1}, \mu_j^k, \dots) \in G \text{ for some } \mu_j^k, \mu_j^{k+1}, \dots\})$$

for $G \in \mathcal{F}_\Theta \times \Sigma_i(t_i)$. Finally, define

$$\mathcal{S}_i := \big\{\mathcal{F}^*_{i,0}, \mathcal{F}^*_{i,1}, \dots, \mathcal{F}^*_{i,k-1}\big\}.$$

It is straightforward to verify that $\mathcal{T}^k := (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$ is a type space in which every type has depth k.

Let $E := \{\theta_1, \theta_2\} \times T_a \times T_b$ be the event that the state of nature is θ_1 or θ_2. For $i = a, b$, $m = 2, 3, \dots$, define

$$\Delta_i^m := \{(\mu_i^1, \dots, \mu_i^m) \in T_i^m : (\mu_i^1, \dots, \mu_i^{m-1}) \in \Delta_i^{m-1}, \mu_i^m(\Delta_j^{m-1}) = 1\}.$$

It can be shown that for $m \leq k$, the set Δ_j^{m-1} is nonempty and belongs to $\mathcal{F}^{m-1}_{j,k-1}$. Moreover, $\Delta_j^k \notin \mathcal{F}^{k-1}_{j,k-1}$. Therefore, for $m = 1, 2, \dots$,

$$\mathcal{B}^m(E) = \big\{(\theta, (\mu_a^1, \mu_a^2, \dots), (\mu_b^1, \mu_b^2, \dots)) \in \Theta \times T_a \times T_b : (\mu_i^1, \dots, \mu_i^m) \in \Delta_i^m \text{ for } i = a, b\big\}$$

is nonempty if and only if $m \leq k$. Hence, $\mathcal{MB}(E)$ is empty.

It remains to show that $\mathcal{C}(E)$ is nonempty. Let

$$F := \big\{(\theta, (\mu_a^1, \mu_a^2, \dots), (\mu_b^1, \mu_b^2, \dots)) \in \Theta \times T_a \times T_b : \mu_i^1 \in C_i^1 \text{ for } i = a, b\big\}.$$

Because the projection function is measurable (with respect to $\mathcal{F}^*_{i,k-1}$ and $\mathcal{F}^1_{i,1}$), and C_i^1 is measurable (in $\mathcal{F}^1_{i,1}$), it follows that $F \in \mathcal{F}_\Theta \times \mathcal{F}^*_{a,k-1} \times \mathcal{F}^*_{b,k-1}$. Also, F is nonempty. It is easy to check that for each $i = a, b$, $F \subseteq \mathcal{B}_i(E)$ and $F \subseteq \mathcal{B}_i(F)$. Hence, $\mathcal{C}(E) \supseteq F$ is nonempty. $\square$.

References

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

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics 4*(6), 1236–1239.

Barwise, J. (1988). Three views of common knowledge. In *TARK '88: Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*, San Francisco, CA, pp. 365–379. Morgan Kaufmann Publishers Inc.

Brandenburger, A. and E. Dekel (1987). Common knowledge with probability 1. *Journal of Mathematical Economics 16*, 237–245.

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.

Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica 61*, 989–1018.

Chwe, M. S.-Y. (2001). *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, NJ: Princeton University Press.

Clark, H. H. and C. Marshall (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, and I. A. Sag (Eds.), *Elements of Discourse Understanding*, Chapter 1. Cambridge, UK: Cambridge University Press.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2012). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*. Forthcoming.

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

Friedenberg, A. and H. J. Keisler (2011). Iterated dominance revisited. Working paper, Arizona State University and University of Wisconsin-Madison.

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

Geanakoplos, J. D. and H. M. Polemarchakis (1982). We can't disagree forever. *Journal of Economic Theory 28*, 192–200.

Halpern, J. Y. and Y. Moses (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM 37*, 549–587.

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

Heifetz, A. (1999). Iterative and fixed point common belief. *Journal of Philosophical Logic 28*, 61–79.

Heifetz, A. and W. Kets (2013). All types naive and canny. Working paper, Northwestern University.

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.

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.

Kets, W. (2014). Rationalizable strategic behavior and finite depth of reasoning. Working paper, Northwestern University.

Kinderman, P., R. Dunbar, and R. P. Bentall (1998). Theory-of-Mind deficits and casual attributions. *British Journal of Psychology 89*, 191–204.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Lismont, L. and P. Mongin (1995). Belief closure: A semantics of common knowledge for modal propositional logic. *Mathematical Social Sciences 30*, 127–153.

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.

Monderer, D. and D. Samet (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior 1*, 170–190.

Ok, E. A. (2014). Elements of order theory. Available at https://files.nyu.edu/eo1 /public/books.html.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge". *American Economic Review 79*, 385–391.

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

Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press.

Strzalecki, T. (2009). Depth of reasoning and higher order beliefs. Working paper, Harvard University.