Finite Depth of Reasoning and Equilibrium Play in Games with Incomplete Information

Willemien Kets*

July 22, 2013

Abstract

The standard framework for analyzing games with incomplete information models players as if they have an infinite depth of reasoning, which is not always consistent with experimental evidence. This paper generalizes the type spaces of Harsanyi (1967–1968) so that players can have a finite depth of reasoning. We do this restricting the set of events that a player of a finite depth can reason about. This approach allows us to extend the Bayesian-Nash equilibrium concept to environments with players with a finite depth of reasoning. We demonstrate that the standard approach of modeling beliefs with Harsanyi type spaces fails to capture the equilibrium behavior of players with a finite depth, at least in some games. Consequently, the standard approach cannot be used to describe the equilibrium behavior of players with a finite depth in general.

JEL classification: C700, C720, D800, D830

Keywords: Bounded rationality, higher-order beliefs, finite depth of reasoning, games with incomplete information, Bayesian equilibrium.

*Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu. Phone: +1-505-204 8012. This paper supersedes Kets (2009) and Kets (2010). I am grateful to Adam Brandenburger, Yossi Feinberg, and Matthew Jackson for their guidance and support, to Adam Brandenburger, Eddie Dekel, Ben Golub, Joe Halpern, Aviad Heifetz, Philippe Jehiel, Rosemarie Nagel, Antonio Penta, Marcin Peski, Tomasz Sadzik, Dov Samet, Marciano Siniscalchi, Rani Spiegler, Jonathan Weinstein, and, especially, Amanda Friedenberg 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

In games with incomplete information, it is important to not only consider players' beliefs about the state of nature, but also their beliefs about other players' beliefs. Consider, for example, a player who has to decide whether or not to invest in a given project. 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 state of nature, i.e., on her first-order belief. Because the same is true for her opponents, the player's optimal action may also depend on her belief about her opponents' first-order belief, i.e., on her second-order belief. And because her opponents in turn may condition their action on *their* beliefs about their opponents' beliefs about the state of nature, the player's optimal choice may also depend on her belief about her opponents' second-order beliefs (i.e., her third-order belief), and so on, ad infinitum.

Harsanyi (1967–1968) developed a tractable framework to analyze such games with incomplete information, and Harsanyi type spaces are widely used to study questions of economic interest. However, in the Harsanyi formalism, players are modeled as if they have an *infinite depth* of reasoning, that is, as if they can form beliefs about every possible higher-order event. Since it seems empirically plausible that players only have a *finite depth* of reasoning,[1] it is important to understand the behavior of players with a finite depth of reasoning in games with incomplete information. In particular, an important question is whether the standard Harsanyi framework can be used to model the behavior of such players.

While there is an extensive literature on the behavior of players with a finite depth in games without payoff uncertainty,[2] much less is understood about their behavior in games with incomplete information. This paper provides the first general framework that jointly models players' higher-order beliefs and their depth of reasoning, to analyze players' behavior in games with incomplete information.[3] The framework generalizes the standard Harsanyi framework to allow players to have a finite depth of reasoning, that is, to have beliefs only up

[1] For example, in experiments where subjects have to answer multiple-choice questions about a simple story that they read, most do no better than chance on questions whether it is the case that "Ann thought that Bob believed that Carol knew that Dan did not want Bob to work for him" or "Ann thought that Bob hoped that Carol would believe that Dan wanted Bob to work for him," suggesting that even if they can reason about such higher order beliefs, at least they have problems applying them effectively (e.g., Kinderman et al., 1998).

[2] Play in games with complete information by players with a finite depth is considered by, e.g., Nagel (1995), Stahl and Wilson (1995) Ho et al. (1998), Costa-Gomes et al. (2001), Strzalecki (2009), and Alaoui and Penta (2013). See Crawford et al. (2012) for a survey.

[3] Brocas et al. (2009), Crawford and Iriberri (2007), and Rogers et al. (2009) present behavioral models for games with incomplete information, but do not develop a model of beliefs independent of behavior. See Section 7 for further discussion.

to order four, say, or to think possible that an opponent has beliefs only up to order two or three.

The main innovation is that a player's depth is modeled by the set of events that a player can reason about, rather than by a simple number, as in the rest of the literature. For example, a player of depth 2, who can reason only about her opponents' first-order beliefs, can reason precisely about the events that can be described in terms of the first-order beliefs of his opponents. As we discuss in Section 7, this extends the notion of a small world of Savage (1954) to a strategic context.

This richer framework allows us to derive new strategic implications. Unlike other papers, we focus on equilibrium behavior, as this is the solution concept most commonly used in applications. Maintaining the assumption of equilibrium behavior also makes it possible to understand the effect of players' depth of reasoning in isolation. This allows us to study whether predictions obtained using the standard (equilibrium) framework remain valid when players have a finite depth of reasoning, at least in principle.[4]

In equilibrium, each player has correct beliefs about her opponent's strategy and plays a best response. This requires that the strategy of a player's opponent does not depend on his beliefs at orders she cannot reason about. Thus, players' behavior is consistent in the usual way, but players may have an imperfect understanding of their strategic environment in the sense that they may not be able to reason about the other players' beliefs at all orders. When every type has an infinite depth of reasoning, this concept reduces to Bayesian-Nash equilibrium.

A natural question is whether the equilibrium behavior of players with a finite depth of reasoning can be modeled using Harsanyi type spaces. One might hope, for instance, that for a given type space $\mathcal{T}$ in which players have a finite depth of reasoning, there is a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ that gives the same equilibria in every game as $\mathcal{T}$. If that is the case, then we do not need to be concerned with the question whether or not real players have a finite or infinite depth if we are interested in equilibrium predictions: we can simply use the Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ to model equilibrium play.

The Harsanyi type spaces typically used in applied work are natural candidates for this purpose: in these type spaces, the higher-order beliefs of a type are determined uniquely by its beliefs up to some fixed, finite order.[5] Since strategies cannot depend on beliefs at arbitrarily

[4]Of course, it is an empirical question whether players (with a finite or infinite depth) actually follow equilibrium strategies, and, if so, under what conditions.

[5]In other words, there is some $k < \infty$ such that for each type in the Harsanyi type space, the higher-order beliefs it induces are commonly known conditional on its kth-order beliefs; see, e.g., Morris et al. (1995) and Qin and Yang (2013). This class of type spaces includes Harsanyi type spaces with finite type sets, or Harsanyi type spaces in which the types are given by players' payoff types and are drawn from a common prior.

high order in such Harsanyi type spaces, one might hope that such Harsanyi type spaces can be used to model the behavior of players with a finite depth of reasoning.

We show that this is not the case in general (Proposition 5.4). Intuitively, even in "simple" type spaces like the ones used in applied work, the strategy of a player's opponent may still depend on his beliefs at an order that she cannot reason about. The key insight, discussed in more detail in Section 2, is that a player, say, Ann, who has a finite depth of reasoning k does not fully understand the beliefs of her opponent, say, Bob, meaning that some of the beliefs Bob might have can only be distinguished at order k. In equilibrium, Bob may want to condition his play on his kth-order beliefs. While in any Harsanyi type space, Ann understands such a strategy, she can only reason about strategies that depend on Bob's $(k-1)$th-order beliefs if she has depth k. This means that, at least in general games, Harsanyi type spaces are not suitable to model the equilibrium behavior of players with a finite depth of reasoning.

We would like to stress from the outset that the equilibrium concept we consider is just a straightforward extension of Bayesian-Nash equilibrium to the present setting; and, like Bayesian-Nash equilibrium itself, it is motivated more by conceptual considerations than by psychological realism. Our analysis clearly brings out the general problems with equilibrium analysis in environments where players are limited in their reasoning abilities, and more specifically, points out the issues associated with using the Harsanyi approach as an "as-if" model in such contexts. In this sense, our results complement the critique of Dekel et al. (2004), who considered learning foundations for Bayesian-Nash equilibrium.

The next section illustrates our main results with some simple examples. The formal treatment starts in Section 3.

2. Examples

2.1. Harsanyi type spaces

As shown by Harsanyi (1967–1968), players' higher-order beliefs can be represented in a compact way using type spaces. In a *Harsanyi type space*, each player i 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 θ 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. The state of nature θ can be either high (H) or low (L), and 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.

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 H, which specifies its *first-order*

$\beta_a(t_a^1)$	H	L
t_b^1	1	0
t_b^2	0	0
t_b^3	0	0
t_b^4	0	0

$\beta_a(t_a^2)$	H	L
t_b^1	0	0
t_b^2	0	0
t_b^3	1	0
t_b^4	0	0

$\beta_b(t_b^1)$	H	L
t_a^1	1	0
t_a^2	0	0
t_a^3	0	0
t_a^4	0	0

$\beta_b(t_b^2)$	H	L
t_a^1	0	0
t_a^2	0	0
t_a^3	1	0
t_a^4	0	0

$\beta_a(t_a^3)$	H	L
t_b^1	0	0
t_b^2	0	1
t_b^3	0	0
t_b^4	0	0

$\beta_a(t_a^4)$	H	L
t_b^1	0	0
t_b^2	0	0
t_b^3	0	1
t_b^4	0	0

$\beta_b(t_b^3)$	H	L
t_a^1	0	0
t_a^2	0	1
t_a^3	0	0
t_a^4	0	0

$\beta_b(t_b^4)$	H	L
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; we write x for the singleton $\{x\}$.

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 $\theta = H$ (as it assigns probability 1 to type t_b^1, which believes that $\theta = H$). This specifies the *second-order belief* $\mu_a^2(t_a^1)$ induced by t_a^1, which is a probability measure on the set of states of nature and Bob's *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 set of states of nature and Bob's *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 $\theta = H$ (as t_b^1 assigns probability 1 to type t_a^1, which puts probability 1 on $\theta = H$).

We can continue this way, uncovering the kth-order belief $\mu_a^k(t_a^1)$ that t_a^1 generates for each k, with a kth-order belief being a probability measure on the set of states of nature and Bob's $(k-1)$*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. ◁

We are interested in the depth of reasoning of belief hierarchies. We 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.

Hence, 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 on $\mathcal{F}_\Theta$; the second-order belief μ_a^2 is a probability measure on the σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_b^1$, where $\mathcal{F}_b^1$ is the σ-algebra on Bob's first-order belief hierarchies

that includes the events expressible in his beliefs about Θ; and so on.

More precisely, we define $\mathcal{F}_b^1$ to be the coarsest σ-algebra on Bob's first-order belief hierarchies that contain the sets

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

with $\Sigma(\mu_b^1)$ the σ-algebra on which μ_b^1 is defined, for any event $E \in \mathcal{F}_\Theta$ and every probability $p \in [0,1]$. 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

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

for every probability $p \in [0,1]$ and every event E in $\mathcal{F}_\Theta \times \mathcal{F}_a^{m-2}$ concerning θ and Ann's $(m-2)$th-order belief hierarchies. (Because belief hierarchies are constructed recursively, this σ-algebra also contains the sets of $(m-1)$th-order belief hierarchies that can be described in terms of Bob's belief μ_b^ℓ at lower order $\ell < m-1$; see Section 3.)

Then, 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 on $\mathcal{F}_\Theta$, and for $m > 1$, 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 Θ and Bob's $(m-1)$th-order belief hierarchies.[6]

Going back to Example 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 state (θ, t_b). 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.[7] So, the belief hierarchy induced by the type has an infinite depth of reasoning. The same is true, in fact, for any type in a Harsanyi type space (Observation 1), as should be expected.

2.2. Finite depth of reasoning

We want to extend the Harsanyi approach to allow types to induce a belief hierarchy of finite depth, where, loosely speaking, a belief hierarchy has depth $k < \infty$ if it can form a belief only about the state of nature and the other players' $(k-1)$th-order beliefs. The next example demonstrates that Ann can form a belief only about Bob's first-order beliefs (but not about his higher-order beliefs) whenever the belief of a type for Ann about Bob's types is defined

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

[7] 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 4.1 makes this connection. Also see Corollary 4.3.

on a σ-algebra that distinguishes Bob's types only when they differ in their first-order belief (but not when the beliefs induced by the types differ exclusively at higher order).

Example 2. Consider the type space in Figure 2. Each type $\tilde{t}_a$ for Ann is endowed with the σ-algebra $\Sigma_a(\tilde{t}_a)$ generated by the partition $\{\{\tilde{t}_a^1, \tilde{t}_a^2\}, \{\tilde{t}_a^3, \tilde{t}_a^4\}\}$, and likewise for the types for Bob.

$\beta_a(\tilde{t}_a^1)$	H	L
$\{\tilde{t}_b^1, \tilde{t}_b^2\}$	1	0
$\{\tilde{t}_b^3, \tilde{t}_b^4\}$	0	0

$\beta_a(\tilde{t}_a^2)$	H	L
$\{\tilde{t}_b^1, \tilde{t}_b^2\}$	0	0
$\{\tilde{t}_b^3, \tilde{t}_b^4\}$	1	0

$\beta_b(\tilde{t}_b^1)$	H	L
$\{\tilde{t}_a^1, \tilde{t}_a^2\}$	1	0
$\{\tilde{t}_a^3, \tilde{t}_a^4\}$	0	0

$\beta_b(\tilde{t}_b^2)$	H	L
$\{\tilde{t}_a^1, \tilde{t}_a^2\}$	0	0
$\{\tilde{t}_a^3, \tilde{t}_a^4\}$	1	0

$\beta_a(\tilde{t}_a^3)$	H	L
$\{\tilde{t}_b^1, \tilde{t}_b^2\}$	0	1
$\{\tilde{t}_b^3, \tilde{t}_b^4\}$	0	0

$\beta_a(\tilde{t}_a^4)$	H	L
$\{\tilde{t}_b^1, \tilde{t}_b^2\}$	0	0
$\{\tilde{t}_b^3, \tilde{t}_b^4\}$	0	1

$\beta_b(\tilde{t}_b^3)$	H	L
$\{\tilde{t}_a^1, \tilde{t}_a^2\}$	0	1
$\{\tilde{t}_a^3, \tilde{t}_a^4\}$	0	0

$\beta_b(\tilde{t}_b^4)$	H	L
$\{\tilde{t}_a^1, \tilde{t}_a^2\}$	0	0
$\{\tilde{t}_a^3, \tilde{t}_a^4\}$	0	1

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

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

More generally, to model that players can have a finite depth of reasoning (and potentially different depths), we restrict the set of events that types with a finite depth can reason about, i.e., that they can assign a probability to. Specifically, we define a type space in which a player's belief can be defined on different σ-algebras. Thus, we 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 $\tilde{t}_a^1$ can assign a probability to the event that Bob has type $\tilde{t}_b^1$ or $\tilde{t}_b^2$ (and thus to the event that Bob believes that $\theta = H$), but not to the event that Bob has type $\tilde{t}_b^1$ (and therefore not to the event that Bob believes that Ann believes that $\theta = H$).

Types for Ann that have a different σ-algebra have a different depth of reasoning, as we will see; likewise for Bob. This means that players may be uncertain about the depth of reasoning of their opponent; see Example 7 below for an illustration.

We can now define a belief hierarchy's finite depth of reasoning. 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 expressible in terms of Bob's $(m-1)$th-order beliefs, 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 beliefs*. That is, 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 Θ 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 $\mathcal{F}_\Theta \times \mathcal{F}_{b,k-1}^{m-1}$, with $\mathcal{F}_{b,k-1}^{m-1}$ the coarsest σ-algebra on Bob's $(m-1)$th-order belief hierarchies that contains the events that are expressible in terms of Bob's $(k-1)$th-order beliefs, 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 $\mathcal{F}_{b,k-1}^{m-1} \subsetneq \mathcal{F}_{b,k}^{m-1}$, where we define $\mathcal{F}_{b,m-1}^{m-1} := \mathcal{F}_b^{m-1}$.

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

In Example 2, the higher-order beliefs $\mu_a^k(\tilde{t}_a^1)$, $k \geq 2$, induced by $\tilde{t}_a^1$ can assign a probability only to events that are expressible in terms of the state of nature and Bob's first-order beliefs: the σ-algebra $\Sigma_a(\tilde{t}_a^1)$ 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 any event in $\Sigma_a(\tilde{t}_a^1)$ can be described in terms of Bob's first-order beliefs. The same is true for the other types. It follows that every type has depth 2.

Of course, if we let the σ-algebras in $\mathcal{S}_a$ and $\mathcal{S}_b$ be arbitrary, then a type need not generate a belief hierarchy of a well-defined depth. Lemma 3.1 and Theorem 4.2 demonstrate, though, that if we relax the condition that belief maps be measurable in the definition of Harsanyi type spaces in an appropriate way, then each type generates a belief hierarchy of a well-defined depth.

2.3. Equilibrium

2.3.1. Definition

What can we say about the equilibrium play of players with a finite depth of reasoning? Roughly, a strategy profile $\sigma = (\sigma_a, \sigma_b)$ is an equilibrium if each player $i = a, b$ plays a best response $\sigma_i(t_i)$ given her type t_i to a strategy of her opponent that does not depend on his beliefs at orders she cannot reason about.

For our purposes, it is useful to distinguish in the description of a game with incomplete information between the beliefs (as given by the type space) on the one hand, and payoffs and actions on the other. We therefore take a *game* $\mathcal{G}$ to specify a set of actions for each player, as well as the players' payoffs for each action profile and state of nature. A *model* is a pair $(\mathcal{G}, \mathcal{T})$, with $\mathcal{G}$ a game, and $\mathcal{T}$ a type space. For simplicity, we restrict attention to so-called *depth-k (type) spaces* in which each type has the same depth of reasoning k. Then, a strategy profile $\sigma = (\sigma_a, \sigma_b)$ is an *equilibrium* if for each player $i = a, b$, the following hold:

- for each type t_i, every action s_i that is played with positive probability under $\sigma_i(t_i)$ is a best response to σ_j, $j \neq i$; and
- for each type t_i, the expected utility of t_i of each action s_i that i might play is well-defined if $j \neq i$ follows the strategy σ_j.

This definition of course coincides with the definition of Bayesian-Nash equilibrium when $k = \infty$. If the second condition holds, then we say that σ_j is *comprehensible* for t_j. A sufficient condition for a strategy to be comprehensible for a type is that it is measurable (with respect to the σ-algebra of the type).

In the next example, types that have the same low-order beliefs follow the same strategies in equilibrium, regardless of their depth:

Example 3. Consider the game in Figure 3, where Ann is the row player, and Bob is the column player. Suppose that players' beliefs are given by the type space in Example 2, in which the depth of reasoning of each type equals 2.

	s_b^1	s_b^2
s_a^1	1,1	1,0
s_a^2	0,1	0,0

$\theta = H$

	s_b^1	s_b^2
s_a^1	0,0	0,1
s_a^2	1,0	1,1

$\theta = L$

Figure 3: A game with dominant actions.

It is easy to see that this game has a unique equilibrium in which types $\tilde{t}_a^1$ and $\tilde{t}_a^2$ play s_a^1 (with probability 1), as s_a^1 is a dominant action for types that believe that $\theta = H$, and types $\tilde{t}_a^3$ and $\tilde{t}_a^4$ play s_a^2, as s_a^2 is dominant for types that believe that $\theta = L$; and likewise for Bob.

Now suppose players' beliefs are given by the type space in Example 1, in which each type has an infinite depth of reasoning. This type space generates the same second-order belief hierarchies as the type space in Example 2; for example, types t_a^1 and $\tilde{t}_a^1$ (in Example 1 and 2, respectively) both believe that $\theta = H$ and that Bob believes that $\theta = H$. Again, there is a unique equilibrium in which types t_a^1 and t_a^2 play s_a^1 (with probability 1), and types t_a^3 and t_a^4 play s_a^2, and similarly for Bob's types. ◁

Example 3 shows that for some depth-2 spaces and some games, there exist Harsanyi type spaces with the same set of equilibria. We want to know whether we can always find a Harsanyi type space that "mimics" the equilibrium predictions of a given finite-depth type space.

2.3.2. Strategic equivalence

To answer that question, fix a depth-k space $\mathcal{T}^k$, and let $\mathcal{T}^{\mathcal{H}}$ be a Harsanyi type space such that there is a surjective mapping φ_i for each player i that maps the Harsanyi types in $\mathcal{T}^{\mathcal{H}}$ for player i into depth-k types for i in $\mathcal{T}^k$. It is natural to consider a Harsanyi type space whose types induce belief hierarchies that extend the belief hierarchies generated by the types in $\mathcal{T}^k$ to infinite depth, with the mappings φ_i relating the depth-k types and the Harsanyi types that "extend" them, and we indeed consider such Harsanyi type spaces in the next section. However, the definitions apply to general type spaces.

We say that the type spaces $\mathcal{T}^k$ and $\mathcal{T}^{\mathcal{H}}$ are *strategically equivalent* if for each player $i = a, b$ and each game $\mathcal{G}$, the following hold:

(1) if a strategy profile $\sigma^k = (\sigma_a^k, \sigma_b^k)$ is an equilibrium of the depth-k model $(\mathcal{G}, \mathcal{T}^k)$, then the strategy profile $\sigma = (\sigma_a, \sigma_b)$, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i = a, b$, is an equilibrium of the Harsanyi model $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$; and

(2) if a strategy profile $\sigma = (\sigma_a, \sigma_b)$ is an equilibrium of the Harsanyi model $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, then the strategy profile $\sigma^k = (\sigma_a^k, \sigma_b^k)$, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i = a, b$, is an equilibrium of the depth-k model $(\mathcal{G}, \mathcal{T}^k)$.

This requires that if a Harsanyi type $t_i^{\mathcal{H}}$ for player i in $\mathcal{T}^{\mathcal{H}}$ plays the mixed action $\sigma_i(t_i^{\mathcal{H}})$, then the depth-$k$ type $t_i = \varphi_i(t_i^{\mathcal{H}})$ in $\mathcal{T}^k$ corresponding to $t_i^{\mathcal{H}}$ plays the same mixed action $\sigma_i^k(t_i) = \sigma_i(t_i^{\mathcal{H}})$, and vice versa. Note that if φ_i is not surjective for some player i, then the strategy σ_i^k is not well-defined.

The definition of strategic equivalence is easiest to understand if $\mathcal{T}^k$ and $\mathcal{T}^{\mathcal{H}}$ have the same type sets (i.e., for each player i, φ_i is the identity function). In that case, (1) requires that

for each game $\mathcal{G}$, the set of equilibria under $\mathcal{T}^k$ is a subset of the set of equilibria under $\mathcal{T}^{\mathcal{H}}$; and, conversely, (2) requires that for each game $\mathcal{G}$, the set of equilibria under $\mathcal{T}^{\mathcal{H}}$ is a subset of the set of equilibria under $\mathcal{T}^k$. Allowing arbitrary type sets strengthens our negative result (Proposition 5.4), without substantially weakening the other results.

Conditions (1) and (2) are satisfied for the game in Figure 3 by the type spaces in Example 3. The question is whether for a given depth-k type space, there is a Harsanyi type space that satisfies (1) and (2) for *every* game.

2.3.3. Harsanyi extensions

We first ask whether, for a given depth-k type space $\mathcal{T}^k$, condition (1) holds (for all games $\mathcal{G}$), that is, of there is a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ such that for any game $\mathcal{G}$, for any equilibrium σ^k of the game when beliefs are given by $\mathcal{T}^k$, there is a corresponding equilibrium σ of the game when beliefs are given by $\mathcal{T}^{\mathcal{H}}$. Not surprisingly, this does not hold for all Harsanyi type spaces:

Example 4. Consider the game in Figure 4, and consider the following type space. Each player $i = a, b$ has two types, labeled t_i^1, t_i^2. Each type t_i is endowed with the trivial σ-algebra $\{\{t_j^1, t_j^2\}, \emptyset\}$ on the type set of the other player j. Type t_a^1 assigns probability $\frac{9}{10}$ to the event that $\theta = H$ (and that Bob has a type in $\{t_b^1, t_b^2\}$), and the complementary probability to the event that $\theta = L$. Type t_a^2 assigns probability $\frac{4}{5}$ to the event that $\theta = H$, and the remaining probability to the event that $\theta = L$. The beliefs for Bob's types are defined similarly. Since the σ-algebra on which the types' beliefs are defined lumps together types that differ in their first-order beliefs, each type has depth 1.

$\theta = H$

	s_b^1	s_b^2
s_a^1	1,1	-2,0
s_a^2	0,-2	0,0

$\theta = L$

	s_b^1	s_b^2
s_a^1	-2,-2	-2,0
s_a^2	0,-2	0,0

Figure 4: A risky coordination game.

It is easy to see that this model has an equilibrium σ in which type t_i plays the risky action s_i^1, for $i = a, b$. Clearly, if beliefs are given instead by a (Harsanyi) type space in which all types have very different first-order beliefs than t_i^1 and t_i^2, $i = a, b$, – for example, assigning high probability to $\theta = L$ –, then this may no longer be an equilibrium.

Moreover, even if beliefs are given by a (Harsanyi) type space which contains types that have the same beliefs as t_i^1 and t_i^2, there may not be an equilibrium in which types that assign a high probability to $\theta = H$ play the risky action s_i^1. To wit, we can construct a Harsanyi type

space $\mathcal{T}^{\mathcal{H}}$ that for each player $i = a, b$, contains types $t_i^{\mathcal{H},1}$ and $t_i^{\mathcal{H},2}$ that generate the same first-order beliefs as t_i^1 and t_i^2, respectively, yet in *every* equilibrium of the game, types $t_i^{\mathcal{H},1}$ and $t_i^{\mathcal{H},2}$ play the safe action s_i^2. We can do this by including a type t^* in $\mathcal{T}^{\mathcal{H}}$ that believes that $\theta = L$, so that the safe action is strictly dominant for t^*, and then choose $t_i^{\mathcal{H},1}$ and $t_i^{\mathcal{H},2}$ such that these types have the same first-order beliefs as t_i^1 and t_i^2, respectively, and that believe that the other player believes... that the other player has type t^* (cf. Rubinstein, 1989; Weinstein and Yildiz, 2007). ◁

Example 4 suggest that a necessary condition for (1) to hold is that $\mathcal{T}^{\mathcal{H}}$ does not include types that have different kth-order beliefs than the types in $\mathcal{T}^k$. Proposition 5.2 shows that this condition is also sufficient.

More precisely, say that a Harsanyi type space is a *Harsanyi extension* of a depth-k type space $\mathcal{T}^k$ if it generates the same kth-order belief hierarchies as $\mathcal{T}^k$, and this is common belief. For example, the type space in Example 1 is a Harsanyi extension of the depth-2 space in Example 2. Of course, type spaces have many different Harsanyi extensions.[8] Indeed, a given type in a depth-k space can have multiple "extensions" in one of its Harsanyi extensions, in which case the Harsanyi extension has multiple types with the same kth-order belief hierarchies (for an example, see the proof of Lemma 5.1).

Proposition 5.2 says that for any depth-k space $\mathcal{T}^k$, we have that condition (1) holds (for every game $\mathcal{G}$) if and only if the Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ is a Harsanyi extension of $\mathcal{T}^k$. That (1) holds for any Harsanyi extension of $\mathcal{T}^k$ follows from the fact that the possibilities for profitable deviations do not change when we "increase" the depth of a type, while the condition that strategies be comprehensible in equilibrium becomes easier to satisfy.

Since there is a Harsanyi extension for a large class of depth-k spaces (Lemma 5.1), it follows that for any depth-k space $\mathcal{T}^k$ in that class, there is a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ with the property that for every game $\mathcal{G}$, for every equilibrium of the depth-k model $(\mathcal{G}, \mathcal{T}^k)$, there is a a corresponding equilibrium of the Harsanyi model $(\mathcal{G}, \mathcal{T}^k)$. The converse of this result is false, however, as we discuss next.

2.3.4. A negative result

In Example 3, the strategy of each player only depends on his first-order beliefs in equilibrium. The next example considers a game where the equilibrium condition that players choose a best response may lead them to play a strategy that depends on their beliefs at higher order.

[8] For example, any variant of the type space in Example 1 in which t_a^1 puts probability p on (H, t_b^1) and $1-p$ on (H, t_b^2) for some $p \in [0, 1]$ is a Harsanyi extension of the type space in Example 2.

Example 5. Consider the game in Figure 5. When the action set is restricted to $\{s_i^1, s_i^2\}$ for each player i, then players play a game with dominant actions, or a coordination or anti-coordination game, depending on the state. Action s_i^3 gives the same payoff in every state.

	s_b^1	s_b^2	s_b^3
s_a^1	$1,1$	$1,0$	$1,0$
s_a^2	$0,1$	$0,0$	$0,0$
s_a^3	$0,1$	$0,0$	$1,1$

$\theta = \theta_1$

	s_b^1	s_b^2	s_b^3
s_a^1	$0,0$	$0,1$	$0,0$
s_a^2	$1,0$	$1,1$	$1,0$
s_a^3	$0,0$	$0,1$	$1,1$

$\theta = \theta_2$

	s_b^1	s_b^2	s_b^3
s_a^1	$1,1$	$0,0$	$0,0$
s_a^2	$0,0$	$1,1$	$0,0$
s_a^3	$0,0$	$0,0$	$1,1$

$\theta = \theta_3$

	s_b^1	s_b^2	s_b^3
s_a^1	$0,0$	$1,1$	$0,0$
s_a^2	$1,1$	$0,0$	$0,0$
s_a^3	$0,0$	$0,0$	$1,1$

$\theta = \theta_4$

Figure 5: A game for which some equilibria depend on second-order beliefs (given the type space).

Consider the type space $\mathcal{T}^2$, defined as follows. Each player $i = a, b$ has eight types, labeled $t_i^1, \ldots, t_i^8$, and each type t_i for player i is endowed with the σ-algebra $\mathcal{F}_j$ on j's type set that is generated by the pairs $\{t_j^1, t_j^2\}$, $\{t_j^3, t_j^4\}$, $\{t_j^5, t_j^6\}$, and $\{t_j^7, t_j^8\}$, where $j \neq i$. The beliefs for player $i = a, b$ are given by:

$$\begin{aligned}
\beta_i(t_i^1)(\theta_1, \{t_j^1, t_j^2\}) &= 1, & \beta_i(t_i^2)(\theta_1, \{t_j^3, t_j^4\}) &= 1;\\
\beta_i(t_i^3)(\theta_2, \{t_j^1, t_j^2\}) &= 1, & \beta_i(t_i^4)(\theta_2, \{t_j^3, t_j^4\}) &= 1;\\
\beta_i(t_i^5)(\theta_3, \{t_j^1, t_j^2\}) &= 1, & \beta_i(t_i^6)(\theta_3, \{t_j^3, t_j^4\}) &= 1;\\
\beta_i(t_i^7)(\theta_4, \{t_j^5, t_j^6\}) &= 1, & \beta_i(t_i^8)(\theta_4, \{t_j^7, t_j^8\}) &= 1,
\end{aligned}$$

where $j \neq i$. It is straightforward to verify that the σ-algebra $\mathcal{F}_j$ separates the types for j if they differ in their first-order belief, but not if they differ in their beliefs at higher order, so that each type t_i has depth 2.

Clearly, the strategy profile σ in which each type for player i plays s_i^3 (with probability 1) is an equilibrium of this model. Does this model have another equilibrium? It is natural to consider a strategy profile in which types t_a^1 and t_a^2 play s_a^1, since it is a best response given their beliefs. Likewise, it is a best response for t_a^3 and t_a^4 to play s_a^2. In that case, the unique best response for type t_b^5 for Bob is to play s_b^1, while the unique best response for t_b^6 is to play s_b^2. But what is then a best response for type t_a^7?

Since Bob's strategy depends on his second-order beliefs, and is therefore not measurable with respect to $\mathcal{F}_b$, we cannot even calculate the expected payoff of t_a^7 to each of Ann's actions; and likewise for type t_b^7. And, of course, if we cannot determine the optimal behavior of t_a^7 and t_b^7, then it is unclear what the optimal play for t_a^8 and t_b^8 is. Hence, there is no equilibrium in which for some player i, types t_i^1 and t_i^2 play s_i^1, and types t_i^3 and t_i^4 play s_i^2. ◁

This example illustrates that there can be a fundamental tension when players have a finite depth between the equilibrium conditions that players have correct beliefs and that they play a best response. This tension does not arise in Harsanyi type spaces:

Example 5 (cont.). Refer back to the game in Figure 5, but now suppose that players' beliefs are given by the Harsanyi type space $\mathcal{T}^{\mathcal{H}}$, defined as follows. Again, each player $i = a, b$ has eight types, labeled $t_i^{\mathcal{H},1}, \ldots, t_i^{\mathcal{H},8}$, and each type $t_i^{\mathcal{H}}$ for player i is endowed with the power set on j's type set, where $j \neq i$. The beliefs for player i are given by:

$$\begin{aligned}
\beta_i^{\mathcal{H}}(t_i^{\mathcal{H},1})(\theta_1, t_j^{\mathcal{H},1}) &= 1, & \beta_i^{\mathcal{H}}(t_i^{\mathcal{H},2})(\theta_1, t_j^{\mathcal{H},3}) &= 1;\\
\beta_i^{\mathcal{H}}(t_i^{\mathcal{H},3})(\theta_2, t_j^{\mathcal{H},2}) &= 1, & \beta_i^{\mathcal{H}}(t_i^{\mathcal{H},4})(\theta_2, t_j^{\mathcal{H},4}) &= 1;\\
\beta_i^{\mathcal{H}}(t_i^{\mathcal{H},5})(\theta_3, t_j^{\mathcal{H},2}) &= 1, & \beta_i^{\mathcal{H}}(t_i^{\mathcal{H},6})(\theta_3, t_j^{\mathcal{H},3}) &= 1;\\
\beta_i^{\mathcal{H}}(t_i^{\mathcal{H},7})(\theta_4, t_j^{\mathcal{H},5}) &= 1, & \beta_i^{\mathcal{H}}(t_i^{\mathcal{H},8})(\theta_4, t_j^{\mathcal{H},7}) &= 1,
\end{aligned}$$

where $j \neq i$.

It is easy to see that type $t_i^{\mathcal{H},m}$ generates the same second-order belief hierarchy as type t_i^m in the original depth-2 space $\mathcal{T}^2$. Indeed, this type space is a Harsanyi extension of $\mathcal{T}^2$. Clearly, the strategy profile in which every type for player i plays s_i^3 is still an equilibrium of this model.

But now there is another equilibrium σ in which types $t_a^{\mathcal{H},1}$ and $t_a^{\mathcal{H},2}$ play s_a^1, and types $t_a^{\mathcal{H},3}$ and $t_a^{\mathcal{H},3}$ play s_a^2. In this case, the unique best responses for types $t_b^{\mathcal{H},5}$ and $t_b^{\mathcal{H},6}$ are then s_b^1 and s_b^2, respectively. Given this, the unique best response for $t_a^{\mathcal{H},7}$ is to play s_a^2, and the unique best response for type $t_b^{\mathcal{H},8}$ is s_b^1; the best responses for the other types when $t_b^{\mathcal{H},1}$ and $t_b^{\mathcal{H},2}$ play s_b^1 and types $t_b^{\mathcal{H},3}$ and $t_b^{\mathcal{H},3}$ play s_b^2 can be determined likewise. ◁

In Example 3, the strategy profile σ is an equilibrium of the Harsanyi model, but not of the depth-2 model. The reason is that the strategy σ_a depends on Ann's second-order beliefs, so that Bob's expected payoffs cannot be calculated if his types are of depth 2, and likewise for σ_b.

Proposition 5.4 shows that this intuition holds generally: for every depth-k type space $\mathcal{T}^k$ (that satisfies a nontriviality condition), and for every Harsanyi extension $\mathcal{T}^{\mathcal{H}}$ of $\mathcal{T}^k$, there is a game $\mathcal{G}$ and an equilibrium of the Harsanyi model $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$ such that there is no corresponding equilibrium in the depth-k model $(\mathcal{G}, \mathcal{T}^k)$. The key is that if a type for Ann has finite depth k, then some of Bob's types can be distinguished only in terms of their kth-order beliefs; in Example 5, for example, types t_b^5 and t_b^6 have the same beliefs about the state of nature, and differ only in their beliefs about Ann's beliefs. In equilibrium, Bob may want to condition his strategy on his kth-order beliefs, but a type for Ann of depth k can reason only about strategies that depend on his $(k-1)$th-order beliefs.

We conclude this section by noting that there are games $\mathcal{G}$ such that a depth-k model ($\mathcal{G}$, $\mathcal{T}^k$) does not have an equilibrium, even if for any Harsanyi extension $\mathcal{T}^{\mathcal{H}}$ of $\mathcal{T}^k$, the model ($\mathcal{G}$, $\mathcal{T}^{\mathcal{H}}$) has an equilibrium; a variant of the game in Example 5 in which the action set of each player $i = a, b$ is restricted to $\{s_i^1, s_i^2\}$ is an example. While it may be desirable to have a prediction for a large class of games, we do not view the fact that certain (finite) models do not have an equilibrium as an anomaly: it simply reflects the fact that some equilibrium strategies are too complicated for players to reason about if they have limited cognitive resources. Indeed, in such a case, it may be more reasonable to assume nonequilibrium behavior; see, e.g., Brocas et al. (2009), Crawford and Iriberri (2007), Strzalecki (2009), and Heifetz and Kets (2011).[9]

2.3.5. "Simple" Harsanyi type spaces

As suggested in the introduction, a natural conjecture is that "simple" Harsanyi type spaces can be used to describe the equilibrium behavior of types with a finite depth. However, Proposition 5.4 demonstrates that this is not the case. Why this discrepancy between this natural conjecture and the formal result?

To answer this question, we first need to define what we mean by "simple" Harsanyi type spaces. Intuitively, we are interested in Harsanyi type spaces for which the belief hierarchies induced by the types are determined completely by their kth-order beliefs, for some finite k. We formally define such type spaces in Section 5, but it is easy to see that the Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ that extends the depth-2 type space $\mathcal{T}^2$ in Example 5 is of order 2. Why does this order-2 Harsanyi extension have a different set of equilibria than the depth-2 type space? The answer is simple: in a Harsanyi type space of order k, equilibrium strategies may depend on players' kth-order beliefs. But since types of depth k can reason only about $(k-1)$th-order beliefs, such strategy profiles cannot be an equilibrium in a depth-k space. Modeling the beliefs of players with a finite depth of reasoning using "simple" Harsanyi type spaces can thus be misleading if we are interested in equilibrium play: it obscures the fact that some equilibrium strategies that are comprehensible for Harsanyi types are too complicated for players of a finite depth, even if we restrict attention to "simple" Harsanyi type spaces.

[9] However, Proposition 5.4 does not directly follow from the fact that some finite-depth models do not have an equilibrium even if some models based on a Harsanyi extension of the finite-depth space do have equilibria. For this, one would have to show that for *every* depth-k type space, with arbitrary type sets, there is a game for which the depth-k model does not have an equilibrium, while the corresponding Harsanyi model does have an equilibrium for every Harsanyi extension of the depth-k space.

3. Belief hierarchies

We now begin the formal treatment. In this section, we provide an explicit model of players' higher-order beliefs about the state of nature, by constructing their belief hierarchies. We do so in a way that every belief hierarchy has a well-defined depth of reasoning. Section 4 provides an implicit description of these beliefs, by generalizing the familiar Harsanyi representation.

3.1. Preliminaries

We start with some preliminaries. For a set X and σ-algebra $\mathcal{F}$ on X, we write $\Delta(X,\mathcal{F})$ for the set of probability measures on $\mathcal{F}$ (i.e., on the measurable space $(X,\mathcal{F})$), and we 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 $(X_z,\mathcal{F}_z)$, $z\in Z$.[10] 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

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

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

[10] We implicitly assume here that the spaces X_z are disjoint. This is without loss of generality: we can replace any space X_z with an isomorphic copy if needed.

3.2. Construction

There is a set N of players, who are uncertain about the *state of nature* $\theta \in \Theta$. The set Θ of states of nature is endowed with a σ-algebra $\mathcal{F}_\Theta$, and is assumed to contain at least two elements. For simplicity, we focus on the case of two players for much of the paper and write $N = \{a, b\}$; the results generalize to the case of three or more players with minor changes, see the online appendix. Throughout this paper, if we fix a player i, then the player other than i is denoted by j, i.e., $j \neq i$.

Players form beliefs about θ, about their opponent's beliefs about θ, and so on. We define spaces of belief hierarchies to model players' higher-order beliefs, building on the construction of Mertens and Zamir (1985) for the Harsanyi case. We define the belief hierarchies in such a way that each belief hierarchy has a well-defined depth of reasoning; see Section 3.3.

Generalizing Definition 2.1 of Mertens and Zamir, we 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 $\mathscr{S}_i^m(\boldsymbol{C}^{m-1})$ of σ-algebras is defined below;

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

(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.

An example of such a space of belief hierarchies is the universal type space of Mertens and Zamir (1985); see Example 6 below.

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). We 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_n^{m-1} is nonempty for $n \in N$.

Thus, we obtain 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, we have $(\mu_i^1, \ldots, \mu_i^\ell) \in C_i^\ell$. Thus, the set of belief hierarchies for player i (in $\boldsymbol{C}$) is

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

Returning to Condition (i), we note that it 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. 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

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

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

$$\left\{(\mu_j^1, \ldots, \mu_j^{m-1}) \in C_j^{m-1} : \Sigma(\mu_j^{\ell-1}) = \mathcal{F}_\Theta \times \mathcal{F}, \mu_j^{\ell-1}(E) \geq p\right\} \tag{3.1}$$

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 $p \in [0,1]$.[11] 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 $\ell - 1 \leq m - 1$. 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 $\mathscr{S}_i^m(\boldsymbol{C}^{m-1})$ form a filtration:

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

With this selection of σ-algebras, the depth of reasoning of a belief hierarchy is well-defined, as we show in Section 3.3. From hereon, we drop the superscripts m on $\boldsymbol{C}^m$ if no confusion can result.

Before discussing the depth of reasoning of belief hierarchies, we consider a few examples. We first consider the space of belief hierarchies constructed by Mertens and Zamir (1985) and others, in which every mth-order belief μ_i^m is defined on the σ-algebra $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C})$, for every m:

Example 6. (Mertens and Zamir, 1985) We construct the space of belief hierarchies in which every mth-order belief is defined on the finest possible σ-algebra. We make some

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

topological assumptions; this will allow us to show that the space of belief hierarchies we construct defines a type space, which will be useful for showing that certain Harsanyi type spaces exist (Lemma 5.1). We assume that the set Θ of states of nature is Polish; examples of Polish spaces include finite and countable sets, and closed subsets of the real line (under their usual topologies). 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.

We endow the set Θ of states of nature with its Borel σ-algebra, i.e., $\mathcal{F}_\Theta = \mathscr{B}(\Theta)$. For each $i \in N$, take $C_i^{\mathcal{U},1}$ to be the set $\Delta(\Theta, \mathcal{F}_\Theta)$ of all 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 $\mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}})$.[12] Note that $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}}) = \mathscr{B}(C_j^{\mathcal{U},m-1})$.

By standard arguments, the spaces $C_i^{\mathcal{U},m}$ are Polish and nonempty for every i and m. Moreover, for every m and $\ell < m$, the σ-algebra $\mathcal{F}_{j,\ell-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}})$ is a proper sub-σ algebra of $\mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^{\mathcal{U}})$. This implies that the set of mth-order beliefs that we include is a strict subset of the set of all mth-order beliefs.

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) and others. Mertens and Zamir show that every type from a Harsanyi type space can be mapped into this space in a way that preserves beliefs. We therefore refer to the belief hierarchies $(\mu_i^1, \mu_i^2, \ldots)$ in $H_i^{\mathcal{U}}$ as the *Harsanyi (belief) hierarchies*. ◁

While the space of belief hierarchies in Example 6 contains all belief hierarchies generated by types in Harsanyi type spaces, it does not contain all belief hierarchies: it does not contain belief hierarchies $(\mu_i^1, \mu_i^2, \ldots)$ for which for some m, the beliefs about the other player's mth-order beliefs are defined on one of the coarser σ-algebras $\{C_j^{\mathcal{U},m}, \emptyset\}$, $\mathcal{F}_{j,1}^m(\boldsymbol{C}^{\mathcal{U}}), \ldots, \mathcal{F}_{j,m-1}^m(\boldsymbol{C}^{\mathcal{U}})$. The following example constructs a space of belief hierarchies that does not have this restriction:

Example 7. (Kets, 2009) Again, assume that Θ is a Polish space and that $\mathcal{F}_\Theta$ is its Borel σ-algebra $\mathscr{B}(\Theta)$. We endow the union X of a family of topological spaces X_z, $z \in Z$, 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).

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

For each $i \in N$, let $C_i^{*,1} = \Delta(\Theta, \mathcal{F}_\Theta)$ be the set of all 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 $\mathscr{S}_i^m(\boldsymbol{C}^*)$.[13] 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. The set H_i^* additionally contains belief hierarchies $(\mu_i^1, \mu_i^2, \ldots)$ such that for some $k < \infty$, the mth-order belief μ_i^m is defined on $\mathcal{F}_\Theta \times \mathcal{F}_{j,m-1}^{m-1}(\boldsymbol{C}^*)$ for $m \leq k$, and on the coarser σ-algebra $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{m-1}(\boldsymbol{C}^*)$ otherwise. Thus, the set H_i^* strictly contains the set $H_i^{\mathcal{U}}$ of Harsanyi hierarchies. See the online appendix for further details. ◁

Before turning to the depth of reasoning of belief hierarchies, let us note that the typical approach in the literature is to construct a space of belief hierarchies that contains all belief hierarchies in some sense, that is, to construct a so-called universal space. Instead, we define a family of spaces of belief hierarchies, by varying $\boldsymbol{C}$. For the Harsanyi case, the two approaches are equivalent (under certain topological restrictions). For the present setting, this is not the case; see Section 6.

3.3. Depth of reasoning

We define the depth of reasoning of a belief hierarchy $h_i = (\mu_i^1, \mu_i^2, \ldots)$ 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:

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

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

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

By construction, the depth of reasoning of a belief hierarchy is well-defined:[14]

Lemma 3.1. 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.

Going back to the examples in the previous section, we see that the belief hierarchies in $H_i^{\mathcal{U}}$ in Example 6 all have an infinite depth of reasoning. By contrast, the belief hierarchies in H_i^* in Example 7 can have any depth of reasoning: for every $k = 1, 2, \ldots, \infty$, there are hierarchies that have depth k. In addition, H_i^* also contains belief hierarchies of infinite depth that assign positive probability to belief hierarchies of finite depth, and so on; see the online appendix.

4. Type spaces

4.1. Definition

While the construction of players' belief hierarchies in the previous section allows us to directly model their higher-order beliefs, including their depth of reasoning, it would be desirable to have a model that does not require us to write out the belief hierarchies explicitly, in the vein of the type spaces due to Harsanyi (1967–1968). In this section, we generalize the concept of a Harsanyi type space. As we show in Section 4.2, each type induces a belief hierarchy, just like Harsanyi types do, except that the belief hierarchy can be of finite depth.

A (Θ-based) *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

[14] While the depth of reasoning $d_i^{\boldsymbol{C}}(h_i)$ of a belief hierarchy $h_i \in H_i(\boldsymbol{C})$ is defined relative to $\boldsymbol{C}$, the depths of reasoning of belief hierarchies in different spaces are directly related: for any $\boldsymbol{C}$, there is a measurable embedding of $H_i(\boldsymbol{C})$ into the set H_i^* of belief hierarchies constructed in Example 7, and any two belief hierarchies $h_i \in H_i(\boldsymbol{C})$ and $h_i' \in H_i(\boldsymbol{C}')$ (potentially from different spaces) that are mapped into the same belief hierarchy $h_i^* \in H_i^*$ have the same depth of reasoning as h_i^*.

$\Sigma_i(t_i) \in \mathcal{S}_j$ on T_j, and β_i maps each type t_i into a *belief* $\beta_i(t_i) \in \Delta(\Theta \times T_j, \mathcal{F}_\Theta \times \Sigma_i(t_i))$. We refer to β_i as player i's *belief map*. We sometimes use the term *extended type space* to emphasize that a type space need not be a Harsanyi type space (defined below).

Assumption 1 imposes some further restrictions on the σ-algebras in $\mathcal{S}_i$, $i \in N$, 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. To state the assumption, we need some more definitions. We 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 $p \in [0,1]$,

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

If $\mathcal{F}_i$ dominates $\mathcal{F}_j$, then we write $\mathcal{F}_i \succ \mathcal{F}_j$; if $\mathcal{F}_i$ is the coarsest σ-algebra that dominates $\mathcal{F}_j$, we write $\mathcal{F}_i \succ^* \mathcal{F}_j$. 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*. We are now ready to state the condition:

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., $\mathcal{F}_i \succ^* \mathcal{F}_j$.

It follows immediately that each Harsanyi type space is an extended type space. Recall that a (Θ-based) *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.[15] Hence, any Harsanyi type space $\mathcal{T}^\mathcal{H} = (T_i^\mathcal{H}, \beta_i^\mathcal{H})_{i \in N}$ can be viewed as an extended type space, and we 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 in fact strictly 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). Note that Assumption 1 is easy to verify: as with the measurability condition for Harsanyi type spaces, we only need to consider the relation between two σ-algebras. See Section 6 for further discussion.

[15] This can be seen by noting that a function $f : X \to Y$ is measurable (with respect to the σ-algebras $\mathcal{F}_X$ and $\mathcal{F}_Y$ on X and Y, respectively) if and only if the inverse images $f^{-1}(B)$ of subsets $B \subseteq Y$ that generate $\mathcal{F}_Y$ belong to $\mathcal{F}_X$.

We next show that every type can be mapped into a belief hierarchy. We then discuss the depth of reasoning of types.

4.2. From types to belief hierarchies

To map each type in a type space into a belief hierarchy, we simultaneously construct the space of belief hierarchies generated by the type space, *and* the functions that maps each type into a belief hierarchy. Essentially, we use the same construction as in Section 3.2, where we built up belief hierarchies 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 an extended type space $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$ and a player $i \in N$. For each player $i \in N$, we define a mapping $h_i^{\mathcal{T},1}$ from T_i to $\Delta(\Theta, \mathcal{F}_\Theta)$ by $h_i^{\mathcal{T},1}(t_i) = \text{marg}_\Theta\, \beta_i(t_i)$. 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 $\boldsymbol{C}^{\mathcal{T},1} := \prod_{n \in N} C_n^{\mathcal{T},1}$. 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 $\Theta \times C_j^{\mathcal{T},\ell-1}$.[16] Also, assume that the functions $h_i^{\mathcal{T},\ell}$ from T_i into $C_i^{\mathcal{T},\ell}$ have been defined.[17] Define

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

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

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

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

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

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

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

[17] 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 4.1.

Lemma 4.1. For every $i \in N$ and $m = 1, 2, 3, \ldots$, the functions $h_i^{\mathcal{T},m}$ are well-defined.

The key to proving Lemma 4.1 is to relate the σ-algebra $\Sigma_i(t_i) \in \mathcal{S}_j$ of type t_i on T_j to the σ-algebras on the space $C_j^{\mathcal{T},m-1}$ of $(m-1)$th-order hierarchies induced by types in T_j.

We can now 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 3.2. If we 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 $t_i \in T_i$,

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

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

We thus have the following result:

Theorem 4.2. For every extended 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 4.1 and the fact that the construction above gives a space of belief hierarchies.

4.3. Depth of reasoning

By Lemma 3.1 and Theorem 4.2, each type t_i generates a belief hierarchy $h_i^{\mathcal{T}}(t_i)$ of well-defined depth. With some abuse of terminology, we refer to the depth $d_i^{\boldsymbol{C}^{\mathcal{T}}}(h_i^{\mathcal{T}}(t_i))$ of reasoning of the hierarchy induced by a type t_i as the depth of reasoning of the type, and write $d_i^{\mathcal{T}}(t_i)$ for $d_i^{\boldsymbol{C}^{\mathcal{T}}}(h_i^{\mathcal{T}}(t_i))$. As we discuss now, the depth of reasoning of a type can be determined directly from the type space.

The first step is to recognize that there is a tight connection between the σ-algebras on the type sets and the σ-algebras on the belief hierarchies. In the course of proving Lemma 4.1, we establish the following result, which we note for future reference:

Corollary 4.3. For every player $i \in N$ and type $t_i \in T_i$, if t_i has depth $k < \infty$, then

$$\begin{aligned} \Sigma_i(t_i) &= \left\{ \left\{ t_j \in T_j : h_j^{\mathcal{T},k-1}(t_j) \in B_j^{k-1} \right\} : B_j^{k-1} \in \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^{\mathcal{T}}) \right\} \\ &\subsetneq \left\{ \left\{ t_j \in T_j : h_j^{\mathcal{T},k}(t_j) \in B_j^{k} \right\} : B_j^{k} \in \mathcal{F}_{j,k}^{k}(\boldsymbol{C}^{\mathcal{T}}) \right\}; \end{aligned} \tag{4.1}$$

otherwise, if t_i has an infinite depth of reasoning, then

$$\Sigma_i(t_i) \supseteq \left\{ \left\{ t_j \in T_j : h_j^{\mathcal{T},m}(t_j) \in B_j^m \right\} : B_j^m \in \mathcal{F}_{j,m}^m(\boldsymbol{C}^{\mathcal{T}}) \right\} \tag{4.2}$$

for all m.

Thus, if type t_i has depth $k < \infty$, then its σ-algebra is generated by the function that maps its opponent's types into their $(k-1)$th-order belief hierarchies. If t_i has an infinite depth of reasoning, then its σ-algebra contains all subsets of types that can be distinguished on the basis of their finite-order belief hierarchies. Using Corollary 4.3, it is straightforward to show that the σ-algebra of a type t_i of depth k separates the types in T_j if and only if these types for j differ in their (induced) $(k-1)$th-order beliefs; similarly, if t_i has infinite depth, then its σ-algebra separates the types in T_j if the induced beliefs of these types differ at some order.[18] For future reference, we denote the σ-algebra in (4.1) by $\sigma(h_j^{\mathcal{T},k-1})$.

While the expressions in (4.1) and (4.2) refer to the hierarchy mappings $h_j^{\mathcal{T},k}$, and thus to the spaces of belief hierarchies, it is in fact possible to determine the depth of reasoning of a type directly from the σ-algebras on the type sets, without references to hierarchy mappings or 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 4.1, and is thus omitted. For example, as is well-known, the space $\boldsymbol{C}^{\mathcal{U}}$ of belief hierarchies constructed in Example 6 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 $k = 1, 2, \ldots$ 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.

[18] Recall that a σ-algebra $\mathcal{F}$ on a space X separates two (distinct) elements x, x' of X if there is a subset $B \in \mathcal{F}$ such that $x \in B$ and $x' \notin B$. As is well-known, the σ-algebra $\Sigma_i(t_i)$ of a type that has infinite depth may separate two types t_j, t_j' even if the types induce the same belief hierarchy (i.e., the type space is redundant) (e.g., Mertens and Zamir, 1985).

(b) Conversely, suppose that each type has the same finite depth. 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 4.1. An example of a type space with uniform finite depth is the type space in Figure 2: every type for player $i = a, b$ is endowed with the σ-algebra $\mathcal{F}_j^*$, and $\mathcal{F}_a^*$ does not dominate $\mathcal{F}_b^*$ or vice versa. Moreover, the σ-algebras $\mathcal{F}_a^*$ and $\mathcal{F}_b^*$ dominate only the trivial σ-algebra (on Bob's and Ann's type set, respectively), so 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, there are also type spaces in which types can have different depths of reasoning, so that there can be uncertainty about a player's depth. For example, the online appendix shows that the space $\boldsymbol{C}^*$ of belief hierarchies in Example 7 defines a type space such that for every k, finite or infinite, there exists a type for each player with depth k. Also for the general case, it is possible to determine a type's depth directly from the type space, so that we do not have to write out its belief hierarchy to know its depth, essentially by counting the number of σ-algebras that the type's σ-algebra dominates; see Kets (2010) for details.

5. Equilibrium

5.1. Definition

We start with some preliminaries. As in much of the paper, we focus here on the case of two players. The results extend immediately to the general case. In the remainder of the paper, we assume that the set Θ of states of nature is finite, to avoid technicalities, and we endow Θ with its usual (discrete) σ-algebra $\mathcal{F}_\Theta$.

A (Θ-based) *game* is a tuple $G = (S_i, u_i)_{i \in N}$, where for each player i, S_i is a (nonempty) finite set of actions, endowed with its standard σ-algebra $\mathcal{F}_{S_i}$, and $u_i : S \times \Theta \to \mathbb{R}$ is a payoff function (where $S := \prod_i S_i$). A (Θ-based) *(Bayesian) model* is a pair $(G, \mathcal{T})$, where G is a game, and $\mathcal{T}$ is an (extended) type space.

For simplicity, we write $\Delta(S_i)$ and $\Delta(\Theta)$ for $\Delta(S_i, \mathcal{F}_{S_i})$ and $\Delta(\Theta, \mathcal{F}_\Theta)$, respectively. Also, if μ is a probability measure on a product space $X \times Y$, and E is a measurable subset of X, we sometimes write $\mu(E)$ for $\text{marg}_X \mu(E)$.

Fix a game $G = (S_i, u_i)_{i \in N}$ and a type space $\mathcal{T} = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$. A *strategy* for player $i \in N$ is a mapping $\sigma_i : T_i \to \Delta(S_i)$, with $\sigma_i(t_i)(s_i)$ the probability that type t_i plays action s_i. The (interim) expected utility of type t_i of action $s_i \in S_i$ given strategy σ_j of the other

player is given by

$$U_i(s_i, \sigma_j; t_i) := \int_{\Theta \times T_j} u_i(s_i, s_j, \theta)\sigma_j(t_j)(s_j)d\beta_i(t_i)$$

whenever this expression is well-defined; in that case, we say that σ_j is *comprehensible for* t_i. A sufficient condition for σ_j to be comprehensible for t_i is that σ_j is measurable with respect to $\Sigma_i(t_i)$ and the usual σ-algebra on $\Delta(S_j)$.[19]

Definition 2. Let $\mathcal{G}$ be a game, and let $\mathcal{T}$ be a type space. A strategy $\sigma = (\sigma_i)_{i \in N}$ is a *(Bayesian-Nash) equilibrium* of the model $(\mathcal{G}, \mathcal{T})$ if for each player $i \in N$ and type $t_i \in T_i$, the following hold:

- the strategy σ_j is comprehensible for t_i; and
- for each action $s_i \in S_i$ such that $\sigma_i(t_i)(s_i) > 0$,

$$U_i(s_i, \sigma_j; t_i) \geq U_i(s_i', \sigma_j; t_i)$$

  for every action $s_i' \in S_i$.

This definition reduces to the standard definition of Bayesian-Nash equilibrium when $\mathcal{T}$ is a Harsanyi type space. The second condition in Definition 2 is just the familiar best-reply condition; the first condition is needed to ensure that each type can calculate its expected payoffs if the other player follows the equilibrium strategy. This condition is standard,[20] but, as we observed in Example 5, it can have more "bite" for finite-depth types than for Harsanyi types.

For simplicity, we focus primarily on type spaces in which every type has the same depth; the results extend directly to the general case, see Remark 2 below. We refer to a type space in which every player has depth $k < \infty$ as a *depth-k (type) space*; also, recall that a type space in which every type has infinite depth is a Harsanyi type space. To abstract from the issue that the set of equilibria can depend on the presence of redundant types –an issue that is orthogonal to the focus of the present paper –, we restrict attention throughout this section to type spaces that are *nonredundant* in the sense that no two types generate the same belief hierarchy (cf. Mertens and Zamir, 1985). Thus, a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ is nonredundant if for each player i, the hierarchy mapping $h_i^{\mathcal{T}^{\mathcal{H}}}$ is injective, and a depth-k space $\mathcal{T}^k$ is (kth-order) nonredundant if the kth-order hierarchy mapping $h_i^{\mathcal{T}^k, k}$ is injective.

[19] The condition that a strategy be measurable with respect to a type's σ-algebra is not necessary: a strategy σ_j is comprehensible for t_i even if it is not measurable with respect to $\Sigma_i(t_i)$ if (1) σ_j is measurable only on a support of $\beta_i(t_i)$; or (2) for each mixed action $\alpha_i \in \Delta(S_i)$, the function $u_i(\alpha_i, \sigma_j(\cdot), \cdot) : T_j \times \Theta \to \mathbb{R}$ (where u_i is extended to mixed actions in the usual way) is measurable with respect to $\Sigma_i(t_i) \times \mathcal{F}_\Theta$.

[20] For example, it is automatically satisfied for Harsanyi type spaces with countable type sets.

5.2. Strategic equivalence

Our aim is to understand whether the equilibrium behavior of types with a finite depth can be described by the standard equilibrium concept applied to Harsanyi type spaces. Formally, fix a depth-k space $\mathcal{T}^k = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i\in N}$, and let $\mathcal{T}^{\mathcal{H}} = (T_i^{\mathcal{H}}, \beta_i^{\mathcal{H}})_{i\in N}$ be a Harsanyi type space such that there is a surjective mapping φ_i from $T_i^{\mathcal{H}}$ to T_i for each player $i \in N$.

Definition 3. The type spaces $\mathcal{T}^{\mathcal{H}}$ and $\mathcal{T}^k$ are *strategically equivalent* if for each game $\mathcal{G}$, the following hold:

(1) for every equilibrium $\sigma^k = (\sigma_i^k)_{i\in N}$ of $(\mathcal{G}, \mathcal{T}^k)$, there is a corresponding equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, that is, the strategy profile σ, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i \in N$, is an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$; and

(2) for every equilibrium $\sigma = (\sigma_i)_{i\in N}$ of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, there is a corresponding equilibrium of $(\mathcal{G}, \mathcal{T}^k)$, that is, the strategy profile $\sigma^k = (\sigma_i^k)_{i\in N}$, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i \in N$, is an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$.

If (1) holds for every game $\mathcal{G}$, we say that $\mathcal{T}^{\mathcal{H}}$ *contains* the equilibria of $\mathcal{T}^k$; and, conversely, if (2) holds for every game $\mathcal{G}$, we say that $\mathcal{T}^k$ contains the equilibria of $\mathcal{T}^{\mathcal{H}}$.

5.3. Results

We first consider the question whether given a depth-k space $\mathcal{T}^k$, there is a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ that contains the equilibria of $\mathcal{T}^k$. As illustrated by Example 4, this does not hold for arbitrary Harsanyi type spaces. We characterize the subclass of Harsanyi type spaces that contain the equilibria of $\mathcal{T}^k$.

Let $k = 1, 2, \ldots$ and fix a depth-k space $\mathcal{T}^k = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i\in N}$; note that there is a unique σ-algebra $\mathcal{F}_i^k \in \mathcal{S}_i$ such that $\Sigma_j(t_j) = \mathcal{F}_i^k$ for all $t_j \in T_j$ (Observation 2). We define a family of Harsanyi type spaces $\mathcal{T}^{\mathcal{H}}$ that extend $\mathcal{T}^k$ in the sense that the beliefs of each type in $\mathcal{T}^{\mathcal{H}}$ up to order k are consistent with the kth-order beliefs of a type in $\mathcal{T}^k$.

Formally, a Harsanyi type space $\mathcal{T}^{\mathcal{H}} = (T_i^{\mathcal{H}}, \{\mathcal{F}_i^{\mathcal{H}}\}, \Sigma_i^{\mathcal{H}}, \beta_i^{\mathcal{H}})_{i\in N}$ is a *Harsanyi extension* of the depth-k space $\mathcal{T}^k$ if for each player $i \in N$, there is a surjective mapping $\varphi_i : T_i^{\mathcal{H}} \to T_i$ such that:

- φ_i is measurable (with respect to $\mathcal{F}_i^k$ and $\mathcal{F}_i^{\mathcal{H}}$); and
- for each $t_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$ and $E \in \mathcal{F}_\Theta \times \mathcal{F}_j^k$, we have

$$\beta_i(\varphi_i(t_i^{\mathcal{H}}))(E) = \beta_i^{\mathcal{H}}(t_i^{\mathcal{H}})\big(\{(\theta, t_j^{\mathcal{H}}) : (\theta, \varphi_j(t_j^{\mathcal{H}})) \in E\}\big).$$

Thus, the mappings φ_i, $i \in N$, preserve the belief structure of $\mathcal{T}^k$ in a similar way as so-called type morphisms in the context of Harsanyi type spaces (Mertens and Zamir, 1985). We therefore refer to $\varphi := (\varphi_i)_{i \in N}$ as an *(extended) type morphism* (from $\mathcal{T}^{\mathcal{H}}$ to $\mathcal{T}^k$), and, with some abuse of terminology, we sometimes refer to the pair $(\mathcal{T}^{\mathcal{H}}, \varphi)$ as a Harsanyi extension of $\mathcal{T}^k$.[21] For any depth-k type $t_i \in T_i$, a type $t_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$ is said to be an *extension* of t_i if $\varphi_i(t_i^{\mathcal{H}}) = t_i$. Note that a type in $\mathcal{T}^k$ can have multiple extensions in $\mathcal{T}^{\mathcal{H}}$. It can be shown that the Harsanyi extensions of $\mathcal{T}^k$ are precisely the Harsanyi type spaces in which the kth-order belief hierarchies are given by those in $\mathcal{T}^k$, and there is common belief in that event. Hence, the current definition coincides with the definition given in Section 2.

The next result shows that a Harsanyi extension exists for broad class of type spaces.

Lemma 5.1. Suppose that $\mathcal{T}^k$ is kth-order nonredundant, and that for each $i \in N$, the type set T_i is Polish and that the Borel σ-algebra $\mathscr{B}(T_i)$ is generated by the kth-order hierarchy mapping $h_i^{\mathcal{T}^k,k}$, i.e., $\mathscr{B}(T_i) = \sigma(h_i^{\mathcal{T}^k,k})$. Then $\mathcal{T}^k$ has a Harsanyi extension.

The proof is relegated to the online appendix. The next result states that a Harsanyi type space contains the equilibria of a depth-k space if and only if it is a Harsanyi extension of the depth-k space:

Proposition 5.2. Let $\mathcal{T}^k$ be a depth-k type space and let $\mathcal{T}^{\mathcal{H}}$ be a Harsanyi type space such that there is a surjective mapping φ_i for each player i from i's type set in $\mathcal{T}^{\mathcal{H}}$ to her type set in $\mathcal{T}^k$. The following are equivalent:

- For every game $\mathcal{G}$ and every equilibrium σ^k of $(\mathcal{G}, \mathcal{T}^k)$, the strategy profile σ, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i \in N$, is an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$;
- $\mathcal{T}^{\mathcal{H}}$ is a Harsanyi extension of $\mathcal{T}^k$.

Proof. The proof that $\mathcal{T}^{\mathcal{H}}$ contains the equilibria of $\mathcal{T}^k$ whenever $\mathcal{T}^{\mathcal{H}}$ is a Harsanyi extension of $\mathcal{T}^k$ uses standard techniques and is therefore relegated to the appendix. To prove the converse, fix a depth-k space $\mathcal{T}^k = (T_i, \mathcal{S}_i, \Sigma_i, \beta_i)_{i \in N}$, and recall that for each $i \in N$ and $t_i \in T_i$, we have that $\Sigma_i(t_i) = \mathcal{F}_j^k$ for some σ-algebra $\mathcal{F}_j^k$ on T_j. Let $\mathcal{T}^{\mathcal{H}} = (T_i^{\mathcal{H}}, \{\mathcal{F}_i^{\mathcal{H}}\}, \Sigma_i^{\mathcal{H}}, \beta_i^{\mathcal{H}})_{i \in N}$ be a Harsanyi type space such that for each player $i \in N$, there is a surjective mapping φ_i from $T_i^{\mathcal{H}}$ to T_i.

Suppose that $\mathcal{T}^{\mathcal{H}}$ is not a Harsanyi extension of $\mathcal{T}^k$. We claim that there is a game $\mathcal{G}$ and an equilibrium σ^k of $(\mathcal{G}, \mathcal{T}^k)$ such that the strategy profile σ, with $\sigma_i = \sigma_i^k \circ \varphi_i$ is not an

[21] The extended type morphisms as defined here do not generalize the type morphism of Mertens and Zamir (1985), as the type morphism of Mertens and Zamir need not be surjective. It is straightforward to define a concept that generalizes both the type morphisms of Mertens and Zamir and the extended type morphisms defined here, but we do not need such a concept for our purposes.

equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$. Since $\mathcal{T}^{\mathcal{H}}$ is not a Harsanyi extension of $\mathcal{T}^k$, there is a player $i \in N$, a type $t_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$, $\theta \in \Theta$, and $B \in \mathcal{F}_j^k$ such that

$$\beta_i^{\mathcal{H}}(t_i^{\mathcal{H}})(\{(\theta, t_j^{\mathcal{H}}) : \varphi_j(t_j^{\mathcal{H}}) \in B\}) \neq \beta_i(t_i)(\theta, B) \tag{5.1}$$

where we have defined $t_i := \varphi_i(t_i^{\mathcal{H}})$. Also, it is without loss of generality to assume that $\beta_i(t_i)(\theta, B) > 0$. We consider the case where $\beta_i^{\mathcal{H}}(t_i^{\mathcal{H}})(\theta) > 0$; we treat the complementary case in the appendix.

Define the game $\mathcal{G}_y = (S_n, u_n)_{n \in N}$ as follows. For each player $n \in N$, let $S_n := \{s_n^1, s_n^2\}$. For each state $\theta' \neq \theta$ and every action profile $s \in S$, let $u_n(s, \theta') := 0$. The payoffs in state θ are given by (where i is the row player):

	s_j^1	s_j^2
s_i^1	$y, 0$	0,0
s_i^2	1,0	1,0

where

$$y := 1 + \frac{\beta_i(t_i)(\theta, T_j \setminus B)}{\beta_i(t_i)(\theta, B)}.$$

Consider the strategy σ_j^k for player j, defined by:

$$\sigma_j^k(t_j) := \begin{cases} s_j^1 & \text{if } t_j \in B; \\ s_j^2 & \text{otherwise.} \end{cases}$$

Then, the strategy profile σ_j^k is comprehensible for every type $t_i' \in T_i$, and type t_i is indifferent between s_i^1 and s_i^2. Define the strategies σ_i^k and $\tilde{\sigma}_i^k$ as follows. Let $\sigma_i^k(t_i)$ and $\tilde{\sigma}_i^k(t_i)$ assign probability 1 to s_i^1 and s_i^2, respectively, and for $t_i' \neq t_i$, let $\sigma_i^k(t_i')$ and $\tilde{\sigma}_i^k(t_i')$ assign probability 1 to s_i^1 if

$$\beta_i(t_i')(\theta, B)(y - 1) - \beta_i(t_i')(\theta, T_j \setminus B) \geq 0,$$

and probability 1 to s_i^2 otherwise. Then all types choose a best response under σ_j^k, σ_i^k, and $\tilde{\sigma}_i^k$, and each of these strategies is comprehensible for the types of the other player. Hence, the strategy profiles (σ_i^k, σ_j^k) and $(\tilde{\sigma}_i^k, \sigma_j^k)$ are equilibria of $(\mathcal{G}, \mathcal{T}^k)$.

Now consider the case where beliefs are given by $\mathcal{T}^{\mathcal{H}}$. If $\{(\theta, t_j^{\mathcal{H}}) : \varphi_j(t_j^{\mathcal{H}}) \in B\} \notin \mathcal{F}_j^{\mathcal{H}}$, then $\sigma_j = \sigma_j^k \circ \varphi_j$ is not comprehensible for $t_i^{\mathcal{H}}$, so the strategy profiles in $\mathcal{T}^{\mathcal{H}}$ corresponding to (σ_i^k, σ_j^k) and $(\tilde{\sigma}_i^k, \sigma_j^k)$ are not an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$. Otherwise, if $\beta_i^{\mathcal{H}}(t_i^{\mathcal{H}})(\{(\theta, t_j^{\mathcal{H}}) : \varphi_j(t_j^{\mathcal{H}}) \in B\}) > \beta_i(t_i)(\theta, B)$, then $(\tilde{\sigma}_i^k, \sigma_j^k)$ is not an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$; else, the strategy profile (σ_i^k, σ_j^k) is not an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$. □

We next ask whether for a given depth-k space $\mathcal{T}^k$, there is a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ such that $\mathcal{T}^k$ contains the equilibria of $\mathcal{T}^{\mathcal{H}}$. Given that we are interested in finding Harsanyi type spaces that are strategically equivalent to $\mathcal{T}^k$, Proposition 5.2 allows us to restrict attention to Harsanyi extensions of $\mathcal{T}^k$. (However, our results do not depend on this restriction.)

We first show that it is without loss of generality to restrict attention to order-k extensions for our purposes, where an order-k extension is a Harsanyi extension in which the higher-order beliefs of each type are completely determined by its kth-order beliefs. Formally, a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$ is of *order* k if the induced belief hierarchies are determined completely by the kth-order beliefs, that is, for each player $i \in N$ and type $t_i^{\mathcal{H}}$ in $\mathcal{T}^{\mathcal{H}}$, we have that $h_i^{\mathcal{T}^{\mathcal{H}},k}(\tilde{t}_i^{\mathcal{H}}) = h_i^{\mathcal{T}^{\mathcal{H}},k}(t_i^{\mathcal{H}})$ implies $h_i^{\mathcal{T}^{\mathcal{H}}}(\tilde{t}_i^{\mathcal{H}}) = h_i^{\mathcal{T}^{\mathcal{H}}}(t_i^{\mathcal{H}})$ for every type $\tilde{t}_i^{\mathcal{H}}$ in $\mathcal{T}^{\mathcal{H}}$. An *order-$k$ (Harsanyi) extension* of a depth-k space $\mathcal{T}^k$ is a Harsanyi extension of $\mathcal{T}^k$ that is of order k.[22]

Lemma 5.3. Let $\mathcal{T}^k$ be a depth-k space that is kth-order nonredundant, and let $\mathcal{T}^{\mathcal{H}}$ be a Harsanyi extension of $\mathcal{T}^k$. If for every game $\mathcal{G}$, for every equilibrium σ of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, the strategy profile σ^k, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i \in N$, is an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$, then $\mathcal{T}^{\mathcal{H}}$ is an order-k extension of $\mathcal{T}^k$. Moreover, if $\mathcal{T}^{\mathcal{H}}$ is nonredundant, then it is without loss of generality to take the type sets in $\mathcal{T}^{\mathcal{H}}$ to be the same as those in $\mathcal{T}^k$, i.e., $T_i^{\mathcal{H}} = T_i$ for $i \in N$.

By the second claim, the requirement that a depth-k space and one of its Harsanyi extensions are strategically equivalent reduces to the condition that for every game $\mathcal{G}$, the set of equilibria of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$ coincides with the set of equilibria of $(\mathcal{G}, \mathcal{T}^k)$, given that we restrict attention to nonredundant type spaces.

The next result shows that a converse of Proposition 5.2 does not hold, at least for depth-k type spaces that are nontrivial in the sense that for some player and one of her types, the support of the type's beliefs contains a subset of types of her opponent that can be described only in terms of his $(k-1)$th-order beliefs, but cannot describe his kth-order beliefs. Formally, a depth-k space $\mathcal{T}^k$ is *nontrivial* if there is a player i, a type t_i for i, and an event $E \subseteq T_j$ such that $E \subseteq \operatorname{supp} \beta_i(t_i)$ and E contains types t_j, t_j' such that $h_j^{\mathcal{T}^k,k-1}(t_j) = h_j^{\mathcal{T}^k,k-1}(t_j')$ and $h_j^{\mathcal{T}^k,k}(t_j) \neq h_j^{\mathcal{T}^k,k}(t_j')$. If a depth-$k$ space does not satisfy the nontriviality condition, then each type assigns probability one to events that are not refined further by describing the other player's kth-order beliefs, that is, only each player's $(k-1)$th-order beliefs can be strategically relevant.

[22] For depth-k spaces with countable type sets, an order-k extension can easily be shown to exist, but existence cannot be shown in general. (While every belief of a depth-k type can be extended to a belief on an appropriately finer σ-algebra (under certain topological conditions), as in the proof of Lemma 5.1, the resulting belief map need not be measurable.)

Proposition 5.4. Let $\mathcal{T}^k$ be a nontrivial depth-k type space and let $\mathcal{T}^{\mathcal{H}}$ be a Harsanyi extension of $\mathcal{T}^k$. Then there is a game $\mathcal{G}$ and an equilibrium σ of the Harsanyi model $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$ such that σ^k, with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i \in N$, is not an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$.

Proof. Since we restrict attention to nonredundant type spaces, by Lemma 5.3, it is without loss of generality to consider order-k extensions $\mathcal{T}^{\mathcal{H}}$ of $\mathcal{T}^k$ in which every player has the same type set as in $\mathcal{T}^k$. We show that there is a game $\mathcal{G}$ and a strategy profile σ such that for every such Harsanyi extension $\mathcal{T}^{\mathcal{H}}$, the strategy profile σ is an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, but it is not an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$.

We focus on the case $k \geq 2$; the proof for the case $k = 1$ is similar and can be found in the online appendix. Let $i \in N$. By Corollary 4.3, the σ-algebra of each type $t_i \in T_i$ is generated by the $(k-1)$th-order hierarchy mapping $h_j^{\mathcal{T}^k, k-1}$, i.e., $\Sigma_i(t_i) = \sigma(h_j^{\mathcal{T}^k, k-1})$ (Eq. (4.1)). Moreover, the σ-algebra $\sigma(h_j^{\mathcal{T}^k, k-1})$ is a proper sub-σ algebra of $\sigma(h_j^{\mathcal{T}^k, k})$. Hence, there are types $t_i, t_i' \in T_i$ that differ only in their kth-order beliefs, that is, $h_i^{\mathcal{T}^k, k-1}(t_i) = h_i^{\mathcal{T}^k, k-1}(t_i')$ and $h_i^{\mathcal{T}^k, k}(t_i) \neq h_i^{\mathcal{T}^k, k}(t_i')$. As $\mathcal{T}^k$ is nontrivial, it is without loss of generality to assume that t_i and t_i' belong to the support of the belief of some type for j.

It follows that there exist $\theta \in \Theta$ and $B \in \sigma(h_j^{\mathcal{T}^k, k-1}) \setminus \sigma(h_j^{\mathcal{T}^k, k-2})$ such that

$$\beta_i(t_i)(\theta, B) > 0, \qquad \beta_i(t_i')(\theta, B) \neq \beta_i(t_i)(\theta, B).$$

Without loss of generality, assume that $\beta_i(t_i)(\theta, B) - \beta_i(t_i')(\theta, B) = \varepsilon$ for some $\varepsilon > 0$. Note that for any Harsanyi extension $\mathcal{T}^{\mathcal{H}}$ (with $T_i^{\mathcal{H}} = T_i$ and φ_i the identity function for every $i \in N$), we have that $\beta_i^{\mathcal{H}}(t_i)(\theta, B) = \beta_i(t_i)(\theta, B)$, and likewise for t_i'.

Consider the following game, denoted $\mathcal{G}_x$. Each player n has two actions, denoted by s_n^1 and s_n^2. Payoffs are given by:

	s_j^1	s_j^2
s_i^1	$x, 0$	$0, 0$
s_i^2	$1, 1$	$1, 1$

θ

	s_j^1	s_j^2
s_i^1	$0, 0$	$0, 0$
s_i^2	$1, 1$	$1, 1$

$\theta' \neq \theta$

where

$$x = \frac{1}{\beta_i(t_i)(\theta, B)} + \varepsilon.$$

Clearly, the model $(\mathcal{G}_x, \mathcal{T}^k)$ has an equilibrium in which every type t_i of player i plays s_i^2 (with probability 1), and every type t_j of player j plays s_j^2; by Proposition 5.2, any model $(\mathcal{G}_x, \mathcal{T}^{\mathcal{H}})$ such that $\mathcal{T}^{\mathcal{H}}$ is a Harsanyi extension of $\mathcal{T}^k$ has a corresponding equilibrium.

We show that there is another strategy profile $\sigma = (\sigma_i, \sigma_j)$ that is an equilibrium of $(\mathcal{G}_x, \mathcal{T}^{\mathcal{H}})$ for any Harsanyi extension $\mathcal{T}^{\mathcal{H}}$ (with $T_i^{\mathcal{H}} = T_i$ for every $i \in N$) which is not an equilibrium of the depth-k model $(\mathcal{G}_x, \mathcal{T}^k)$. Define the strategy σ_j for player j by

$$\sigma_j(t_j)(s_j^1) = 1 \qquad \text{if } t_j \in B;$$
$$\sigma_j(t_j)(s_j^2) = 1 \qquad \text{if } t_j \notin B.$$

Note that σ_j is comprehensible for each type of player i. Then the difference in expected payoffs for t_i between s_i^1 and s_i^2 is

$$U_i(s_i^1, \sigma_j; t_i) - U_i(s_i^2, \sigma_j; t_i) = \beta_i(t_i)(\theta, B)x - 1.$$

Likewise, the difference in expected payoffs for t_i' between s_i^1 and s_i^2 is

$$U_i(s_i^1, \sigma_j; t_i') - U_i(s_i^2, \sigma_j; t_i') = \beta_i(t_i')(\theta, B)x - 1.$$

It can be verified that type t_i strictly prefers s_i^1, and type t_i' strictly prefers s_i^2. If we set $\sigma_i(t_i)(s_i^1) = 1$, $\sigma_i(t_i')(s_i^2) = 1$, and for $\tilde{t}_i \neq t_i, t_i'$, we define $\sigma_i(\tilde{t}_i)(s_i^1) = 1$ if $U_i(s_i^1, \sigma_j; \tilde{t}_i) - U_i(s_i^2, \sigma_j; \tilde{t}_i) \geq 0$, and $\sigma_i(\tilde{t}_i)(s_i^2) = 1$ otherwise, then $\sigma = (\sigma_i, \sigma_j)$ is an equilibrium of $(\mathcal{G}_x, \mathcal{T}^{\mathcal{H}})$ for any Harsanyi extension $\mathcal{T}^{\mathcal{H}}$ (with $T_i^{\mathcal{H}} = T_i$ for every $i \in N$). But σ is not an equilibrium of $(\mathcal{G}_x, \mathcal{T}^k)$, as t_i and t_i' differ only in their kth-order beliefs and play different actions. Thus, in $\mathcal{T}^k$, the strategy σ_i is not comprehensible for the types for j (regardless of the strategies chosen by the types $\tilde{t}_i \neq t_i, t_i'$), and indeed the expected utility is not well-defined for some types for j in $\mathcal{T}^k$. □

The intuition behind Proposition 5.4 is that there is a tension between the equilibrium conditions that players choose a best response given their type, and that beliefs are correct. Because there are types in a depth-k space that have different kth-order beliefs (cf. Corollary 4.3), these types may take different actions in games in which best responses depend on kth-order beliefs, thus violating the condition that equilibrium strategies be comprehensible. Together, Propositions 5.2 and 5.4 imply that there is no depth-k space and Harsanyi type space that are strategically equivalent.

Remark 1. The proofs of Propositions 5.2 and 5.4 make use of games in which some players are indifferent among their actions. This makes it possible to clearly bring out the intuition behind the results. It is possible to prove the results using games in which players are not indifferent, such as the report-your-beliefs games of Dekel et al. (2006), at the expense of a loss of transparency.[23] ◁

[23] In particular, such alternative proofs can be used to prove analogous results for approximate Bayesian-Nash equilibrium.

Remark 2. While we have restricted attention to depth-k spaces for simplicity, Propositions 5.2 and 5.4 directly extend to arbitrary type spaces, where players can be uncertain about the depth of reasoning of their opponent (cf. Example 7). The type for a player that has the lowest depth of reasoning can then determine the maximal "complexity" of the equilibrium strategies (at least on the support of its belief), through the condition that the equilibrium strategies be comprehensible for each type. Otherwise, the results remain the same. ◁

6. Discussion and extensions

Rationalizability. Results analogous to Propositions 5.2 and 5.4 can be derived for a finite-depth version of (interim correlated) rationalizability (Dekel et al., 2007). As in the infinite-depth case discussed by Dekel et al., this concept can be defined in two equivalent ways.[24] The first is as a fixed point of the best-response correspondence. Fix a depth-k space $\mathcal{T}^k$ and a game $\mathcal{G}$. Then, for any collection $F = (F_{t_i})_{i \in N, t_i \in T_i}$ of subsets $F_{t_i} \subseteq S_i$ of actions, define the set of best responses of a type $t_i \in T_i$ of player i against actions in F by

$$BR_i^{\mathcal{G},\mathcal{T}^k}(t_i; F) := \left\{ s_i \in S_i : \begin{array}{c} \text{there is } \sigma_j : \Theta \times T_j \to \Delta(S_j) \text{ s.t.} \\ (1) \quad \sigma_j \text{ is measurable w.r.t. } \Sigma_i(t_i); \\ (2) \quad \sigma_j(\theta, t_j)(s_j) > 0 \text{ implies that } s_j \in F_{t_j}; \\ (3) \quad s_i \in \arg\max_{s_i' \in S_i} \sum_{\theta, s_j} \int u_i(s_i', s_j, \theta) \sigma_j(\theta, t_j)(s_j) \beta_i(t_i)(\theta, dt_j). \end{array} \right\}.$$

The condition that the conjecture σ_j be measurable for t_i is ensures that the type can calculate its expected payoffs. The greatest fixed point of the best-response correspondence then gives the set $R_i^{\mathcal{G},\mathcal{T}^k}(t_i)$ of finite-order rationalizable actions for type t_i. Equivalently, we can define this set as follows: for each player $i \in N$ and type $t_i \in T_i$, define $R_i^{\mathcal{G},\mathcal{T}^k,0}(t_i) := S_i$, and for $m > 0$, define

$$R_i^{\mathcal{G},\mathcal{T}^k,m}(t_i) := BR_i^{\mathcal{G},\mathcal{T}^k}(t_i; (R_n^{\mathcal{G},\mathcal{T}^k,m-1}(t_n))_{n \in N, t_n \in T_n})$$

to be the set of best replies for t_i to the $(m-1)$th-order rationalizable actions. The difference with the standard definition here is again that the conjectures of a type are required to be measurable. It is then possible to show that the set $R_i^{\mathcal{G},\mathcal{T}^k}(t_i)$ of finite-order rationalizable actions for type t_i is equal to the intersection $\bigcap_m R_i^{\mathcal{G},\mathcal{T}^k,m}(t_i)$, as in the infinite-depth case.

[24] While a full epistemic treatment of rationalizability as in Battigalli et al. (2011) is beyond the scope of the present paper, we note that the iterative definition has the following nice interpretation: a player with a finite depth k has an imperfect understanding of the higher-order beliefs of the other player (and can therefore only form conjectures of limited complexity), but can apply the best-reply operator as many times as she wishes (since this does not require complex reasoning, just bookkeeping). This is in line with a view of rationalizability as an "algorithmic" solution concept.

Though the proofs for the analogues of Propositions 5.2 and 5.4 are slightly different than the ones for the equilibrium case, the intuition behind the results given in Section 2 carries over: while any actions that can be rationalized by a type in a finite-depth model can be rationalized by some extension of the type, the converse need not hold, because some actions can be rationalized only by conjectures that depend on high-order beliefs.

Characterizing "simple" equilibria in Harsanyi models. Our negative result (Proposition 5.4) raises the question whether it is possible to characterize the strategy profiles in some Harsanyi type space that form an equilibrium in a corresponding finite-depth type space for a given game. More precisely, fix a Harsanyi type space $\mathcal{T}^{\mathcal{H}}$, and suppose that $\mathcal{T}^{\mathcal{H}}$ is a Harsanyi extension of some finite-depth space $\mathcal{T}$, that is, a type space in which all types have a finite depth of reasoning; without loss of generality, we can take the type sets in $\mathcal{T}$ to be the same as in $\mathcal{T}^{\mathcal{H}}$.[25] Fix a game $\mathcal{G}$, and suppose for simplicity that $\mathcal{T}^{\mathcal{H}}$ and $\mathcal{T}$ are nonredundant, and that a strategy that is comprehensible for a type in $\mathcal{T}$ is measurable with respect to the σ-algebra of that type. Then, it is easy to show that a strategy profile σ is an equilibrium of the finite-depth model $(\mathcal{G}, \mathcal{T})$ if and only if it is a (Bayesian-Nash) equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$ and for each player i, the strategy σ_i is measurable with respect to i's $(k_j - 1)$th-order beliefs, where k_j is the minimum depth of a type for player j in $\mathcal{T}$.[26] That is, the equilibrium behavior of players with a finite depth of reasoning can be described with a refinement of Bayesian-Nash equilibrium that rules out equilibria that depend on players' beliefs at high order.

Thus, we have a refinement of Bayesian-Nash equilibrium for the Harsanyi model $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$ that precisely captures the equilibrium behavior of the finite-depth types in $(\mathcal{G},\mathcal{T})$. The strategies that are measurable with respect to a player's $(k-1)$th-order beliefs in $\mathcal{T}^{\mathcal{H}}$ can be found using the notion of dominance (applied to $\mathcal{T}^{\mathcal{H}}$) introduced in Section 4, so that these strategies can be identified without having to write out players' belief hierarchies. This gives a simple method to study the equilibrium behavior of players with a finite depth of reasoning that uses only Harsanyi type spaces, and that does not require modeling players' depth of reasoning or their belief hierarchies explicitly.

Assumption 1 revisited. As noted in Section 4, Assumption 1 relaxes the condition in the definition of Harsanyi type spaces that belief maps be measurable. Assumption 1 plays an important role in the characterization of the σ-algebra of each type in terms of the hierarchy

[25] It can be shown that every nonredundant Harsanyi type space whose type sets contain at least two elements is a Harsanyi extension of some finite-depth space (using a straightforward generalization of the definition of a Harsanyi extension that allows the types in the finite-depth space to have different depths of reasoning).

[26] The reason that the measurability condition is linked to the minimum depth of the types in $\mathcal{T}$ is that the comprehensibility condition is strongest for the types of the lowest depth; also see Remark 2.

mappings in the proof of Lemma 4.1, which plays a central role in the paper. To see how, note that every σ-algebra $\mathcal{F}_i^1$ on T_i that dominates the trivial σ-algebra $\{T_j, \emptyset\}$ contains the subsets of i's types that can be distinguished on the basis of their beliefs about θ, i.e., in terms of their first-order belief hierarchies; in turn, every σ-algebra $\mathcal{F}_j^2$ on T_j that dominates $\mathcal{F}_i^1$ contains the subsets of types for j that are can be distinguished on the basis of their beliefs about player i's first-order belief hierarchies, i.e., in terms of their second-order belief hierarchies, and so on. Assumption 1 ensures that for each type t_i, either its σ-algebra $\Sigma_i(t_i)$ is part of a finite chain

$$\Sigma_i(t_i) \succ^* \mathcal{F}_i^\ell \succ^* \mathcal{F}_j^{\ell-1} \succ^* \cdots \succ^* \mathcal{F}_m^1 \succ^* \{T_n, \emptyset\}$$

of σ-algebras that are the coarsest σ-algebras that dominate each other (where $n \in N$ and $m \neq n$), so that t_i has a finite depth; or $\Sigma_i(t_i)$ dominates all σ-algebras (on T_i) that are part of such chains, in which case t_i has an infinite depth.

The requirement that the σ-algebra of each type is either part of a finite chain as the one above, or dominates all such chains is essentially equivalent to Assumption 1 (in the sense that the same spaces of belief hierarchies can be modeled under both conditions). Since Assumption 1 requires one to consider only the relation between two σ-algebras at the time, just like the measurability condition, instead of requiring one to construct chains of σ-algebras of arbitrary length, we have chosen the present formulation.

Finally, we note that it is possible to relax the requirement that the σ-algebras in the chains be the coarsest to dominate each other, and just require that they dominate each other. Under this weaker condition, σ-algebras corresponding to a finite depth may include events that are unrelated to a type's depth of reasoning (i.e., events that cannot be described in terms of the hierarchy mappings). In that case, it may not be possible to compare two types of a given player that have the same depth (in a given type space) in terms of the events that they can reason about. Further study of such phenomena and their strategic implications is beyond the scope of the present paper.

A universal space? For the present purpose of studying equilibrium behavior, it is not necessary to construct a universal type space for our class of type spaces, as Mertens and Zamir (1985) and others have done for Harsanyi type spaces. One might nevertheless wonder whether such a type space, which embeds all other type spaces (in the sense that there is a unique type morphism from each type space into the space), can be constructed for the class of type spaces that we consider. One observation is that for any space $\boldsymbol{C}$ of belief hierarchies, there is a measurable mapping from each type space into the type space in Example 7. However, the images of such mappings in $\prod_i H_i^*$ will generally not form a belief-closed subset, as is the case for Harsanyi type spaces (Mertens and Zamir, 1985). We leave a full exploration of such

issues for future research.

7. Related literature

Level-k and cognitive hierarchy models. An important and thriving literature in experimental and behavioral economics studies the behavior of players with a finite depth of reasoning, where it is assumed that a player of depth $k < \infty$ applies the best-response operator k times: a level-0 player is non-strategic and follows some exogenously specified strategy, while for $k > 0$, a level-k player chooses a best response to a belief that his opponents are of a lower level; see footnote 2 for references, and see Crawford et al. (2012) for an excellent survey. We depart from this literature in two ways.

First, while much of the literature in experimental and behavioral economics has focused on games with complete information about payoffs,[27] we study games with incomplete information. This is of interest because in games with incomplete information, limitations on a players' depth of reasoning can affect behavior even beyond the initial periods of play. While in games with complete information, there is no remaining higher-order uncertainty once play has converged to equilibrium (as it often does in experiments), so that the behavior of players with a finite depth is indistinguishable from the play of infinitely sophisticated players, higher-order uncertainty continues to play a role in games with incomplete information, so that the behavior of players with a finite depth of reasoning may differ from that of players with an infinite depth of reasoning even if players have a lot of experience with the game.

Second, while the existing literature has focused on nonequilibrium concepts, we consider an extension of Bayesian-Nash equilibrium. This allows us to reach the strong conclusion that the behavior of players with a finite depth can differ significantly from that of players with an infinite depth of reasoning, even if one considers equilibrium play. We make no claim that equilibrium play provides a good model of behavior, let alone that it is a better model than the models studied in the literature. Rather, the motivation is to extend the Bayesian-Nash equilibrium concept to the present setting, so as to isolate the effect of the

[27] Notable exceptions include Brocas et al. (2009), Crawford and Iriberri (2007), and Rogers et al. (2009), who present behavioral models for specific classes of games with incomplete information and test these models experimentally. However, these authors do not develop a theoretical framework to analyze behavior in games beyond the specific applications that they consider, like we do in the present paper. Strzalecki (2009) introduces type spaces to model uncertainty about other players' depth of reasoning in games with complete information. While his framework can be used to analyze games with incomplete information in the agent-normal form (when type sets are countable), his type spaces only allow for uncertainty about players' depth of reasoning, not about payoffs. Heifetz and Kets (2011) develop a framework based on the framework presented here to study robustness questions.

assumption that players have a finite depth of reasoning from the effect of the assumption of nonequilibrium play, and to investigate whether we can use Harsanyi type spaces and Bayesian-Nash equilibrium to describe the equilibrium behavior of players with a finite depth of reasoning.

The innovation of our approach is that we model a player's depth by the set of events she can reason about. This extends the notion of a small world, introduced by Savage (1954) in the context of single-person decision problems, to a strategic setting. A state (of the world) in a small world describes the possible uncertainties a decision-maker faces in less detail than a state in a larger world, by neglecting certain distinctions between states. This means that "a state of the smaller world corresponds not to one state of the larger, but to a *set* of states" (Savage, 1954, p. 9, emphasis added). In the present framework, a player 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. A player with a lower depth of reasoning makes fewer distinctions between states than a player with a higher depth of reasoning, and thus has a smaller world.

This approach allows us to extend the Bayesian-Nash concept to a setting with players with an infinite depth of reasoning, which does not seem straightforward in models where a player's depth is simply given by a number, as in Strzalecki (2009) and Heifetz and Kets (2011). In addition, the present approach makes it possible to study the question whether players with a finite depth of reasoning can attain common belief (Kets, 2013).

Robustness of predictions. While it is well-known that game-theoretic predictions can be sensitive to small changes in higher-order beliefs (see, e.g., Rubinstein, 1989; Weinstein and Yildiz, 2007, among many others), this paper is the first to point out how bounds on players' depth of reasoning can affect equilibrium play. Our result that the equilibrium behavior of types with a finite depth of reasoning may differ from the equilibrium play of players with an infinite depth of reasoning, even if the beliefs of the latter are described by "simple" Harsanyi type spaces (Proposition 5.4), does *not* rely on small perturbations of players' beliefs at high order. Rather, the key insight is that if a type for player i has finite depth k, then player $j \neq i$ has types that generate belief hierarchies that differ only at order k (Corollary 4.3). Thus, it may be optimal for j to follow a strategy that depends on his kth-order beliefs, but such a strategy is too complex for i's types of depth k.

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, we can get types that can reason about only finitely many orders of beliefs. Indeed, a technical contribution of this paper is to formulate a condition on the type space (Assumption 1) that guarantees 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 $k-1$ (Lemma 4.1 and Corollary 4.3). 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, i.e., on belief hierarchies (e.g., Brandenburger and Keisler, 2006; Friedenberg and Meier, 2012); see Friedenberg and Keisler (2011) for an excellent discussion and further references.

Appendix A Proofs for Sections 3 and 4

A.1 Proof of Lemma 3.1

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 $\operatorname{marg}_{\Theta\times C_j^{k-2}} \mu_i^k = \mu_i^{k-1}$) 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 $\mathcal{F}_\Theta \times \mathcal{F}_{j,k-2}^{k-1}(\boldsymbol{C})$. □

A.2 Proof of Lemma 4.1

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 A.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 we show now, in an important class of cases, $\sigma(\mathcal{A})$ and $\sigma(\mathcal{A}')$ coincide:

Lemma A.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. We first 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, $F' \in \sigma(\mathcal{A})$.

We next show 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 $\underline{\mathcal{F}}$, we can label the σ-algebras in $\mathscr{S}$ as

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

Then,

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

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

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

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, we have $\Delta(X, \mathcal{F}) \in \sigma(\mathcal{A}')$. □

We can now prove Lemma 4.1. We will prove the result by induction. As part of the proof, we construct an inductive structure, in the following way. For each $k = 1, 2, \ldots$, we 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.) We 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 us 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.[28]

To prove the result, we 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,

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

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 generated by the function $h_i^{\mathcal{T},0}$ is simply the trivial σ-algebra $\{T_i, \emptyset\}$.

Lemmas A.3–A.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 A.3 is an auxiliary result that gives a useful characterization of $\sigma(h_i^{\mathcal{T},1})$.

Lemma A.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., $\sigma(h_i^{\mathcal{T},1}) \succ^* \sigma(h_j^{\mathcal{T},0})$.

Proof. Note that

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

where the second equality uses Lemma A.1. □

Lemma A.4. For each type $t_i \in T_i$, we have $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},1})$ or $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},1})$.

[28] 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 $\mathcal{F}_i$, we have $\mathcal{F}_i \supseteq \mathcal{Q}_i^m$ for all m, so that such σ-algebras do not affect the inductive structure.

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 A.3, the σ-algebra $\sigma(h_j^{\mathcal{T},1})$ is the coarsest σ-algebra that dominates $\{T_i, \emptyset\}$. Hence, $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},1})$. □

Lemma A.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 A.4 gives that $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},1})$, a contradiction. □

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

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

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 $t_i \in T_i$, we have $\Sigma_i(t_i) \subsetneq \sigma(h_j^{\mathcal{T},\ell})$ or $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},\ell})$;
- 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 A.6. For each type $t_i \in T_i$, we have $h_i^{\mathcal{T},k}(t_i) \in C_i^{\mathcal{T},k-1} \times \Delta(\Theta \times C_j^{\mathcal{T},k-1}, \mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}}))$.

Proof. By the induction hypothesis, we have that 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 $E \in \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^{\mathcal{T}})$, we have $(\mathrm{Id}_\Theta, h_j^{\mathcal{T},k-1})^{-1}(E) \in \mathcal{F}_\Theta \times \Sigma_i(t_i)$, so that $\mu_i^k(t_i)$ 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 $\mathscr{S}_i^k(\boldsymbol{C}^\mathcal{T})$. □

By Lemma A.6, we can define 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 $h_i^{\mathcal{T},k}$. We next establish the analogues of Lemmas A.3–A.5 for general k, to order the σ-algebras on the type sets.

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

Proof. By Lemma A.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

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

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 $\ell \leq k-1$,

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

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

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

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 $\mathcal{F} \in \mathscr{S}_j^k(\boldsymbol{C}^\mathcal{T})$, we have $\mathcal{F} \subseteq \mathcal{F}_\Theta \times \mathcal{F}_{j,k-1}^{k-1}(\boldsymbol{C}^\mathcal{T})$, we have that $\sigma(h_i^{\mathcal{T},k})$ is generated by the sets

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

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

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

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

Lemma A.8. For each $t_i \in T_i$, we have $\Sigma_i(t_i) \subseteq \sigma(h_j^{\mathcal{T},k})$ or $\Sigma_i(t_i) \supseteq \sigma(h_j^{\mathcal{T},k})$.

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

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

if m is odd, and

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

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

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

We claim that if (a) or (c) is the case, then $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},k})$. We 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 A.7 that $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},2})$. Repeating this argument gives the desired result.

It remains to consider (b). We 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 A.7, we have $\mathcal{F}_i = \sigma(h_i^{\mathcal{T},m}) \subseteq \sigma(h_i^{\mathcal{T},k})$. If $m > k$, then we have $\mathcal{F}_j^{m-k} = \sigma(h_j^{\mathcal{T},k})$ or $\mathcal{F}_i^{m-k} = \sigma(h_i^{\mathcal{T},k})$, depending on whether k is odd or even. We treat 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 A.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 A.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, we have $\mathcal{F}_i \supseteq \sigma(h_i^{\mathcal{T},k-1})$. If $\mathcal{F}_i = \sigma(h_i^{\mathcal{T},k-1})$, then we are done. So suppose $\mathcal{F}_i \supsetneq \sigma(h_i^{\mathcal{T},k-1})$. The proof is complete if we 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 A.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. It follows that for each player $i \in N$ and $k = 1, 2, \ldots$, we have that $h_i^{\mathcal{T},k} : T_i \to C_i^{\mathcal{T},k-1} \times \Delta(\Theta \times C_j^{\mathcal{T},k-1}, \mathscr{S}_i^k(\boldsymbol{C}^{\mathcal{T}}))$ 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 B Proofs for Sections 5

B.1 Proof of Proposition 5.2 (continued)

We first complete the proof of the claim that if $\mathcal{T}^{\mathcal{H}}$ is not a Harsanyi extension of $\mathcal{T}^k$, then we can find a game and an equilibrium of the depth-k model such that the corresponding strategy profile is not an equilibrium of the Harsanyi model, by considering the case in which $\beta_i^{\mathcal{H}}(t_i^{\mathcal{H}})(\theta) = 0$ (with $t_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$ and $\theta \in \Theta$ as defined in the proof in the main text). Note that $\beta_i(t_i)(\theta) > 0$ by assumption, where we recall that $t_i = \varphi_i(t_i^{\mathcal{H}})$. Fix $z \geq (10 - 10\beta_i(t_i)(\theta))/\beta_i(t_i)(\theta)$, and consider the game $\mathcal{G}$ with action sets $S_i = \{s_i^1, s_i^2\}$ and $S_j = \{s_j\}$, and payoffs given by:

	s_j
s_i^1	z,0
s_i^2	0,0

θ

	s_j
s_i^1	-10,0
s_i^2	0,0

$\theta' \neq \theta$

It is easy to see that the depth-k model $(\mathcal{G}, \mathcal{T}^k)$ has an equilibrium in which t_i plays s_i^1 with positive probability, while in any equilibrium of the Harsanyi model $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, type $t_i^{\mathcal{H}}$ plays s_i^2 with probability 1.

We next prove the converse. Specifically, we show that for any game $\mathcal{G}$, and any Harsanyi extension $(\mathcal{T}^{\mathcal{H}}, \varphi)$ of $\mathcal{T}^k$, if $\sigma^k = (\sigma_i^k)_{i \in N}$ is an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$, then $\sigma = (\sigma_i)_{i \in N}$, with $\sigma = \sigma_i^k \circ \varphi_i$ for $i \in N$, is an equilibrium of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$. Fix a Harsanyi extension $(\mathcal{T}^{\mathcal{H}}, \varphi)$ of $\mathcal{T}^k$. Let $\mathcal{G}$ be a game, and suppose σ^k is an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$

Since σ^k is an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$, the strategy σ_i^k is comprehensible for each type for j in $\mathcal{T}^k$. It is straightforward to show that the strategy $\sigma_i^k \circ \varphi_i$ is comprehensible for each type for j in $\mathcal{T}^{\mathcal{H}}$. Also, for each player $i \in N$, Harsanyi type $\tilde{t}_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$, and action $s_i \in S_i$ such that

$\sigma_i^k(\varphi_i(\tilde{t}_i^{\mathcal{H}}))(s_i) > 0$, we have that for all $b_i \in S_i$,

$$\begin{aligned}\int_{\Theta\times T_j^{\mathcal{H}}} u_i(s_i,s_j,\theta)\sigma_j^k(\varphi_j(t_j^{\mathcal{H}}))(s_j)d\beta_i^{\mathcal{H}}(\tilde{t}_i^{\mathcal{H}}) &= \int_{\Theta\times T_j} u_i(s_i,s_j,\theta)\sigma_j^k(t_j)(s_j)d\beta_i(\tilde{t}_i)\\ &\geq \int_{\Theta\times T_j} u_i(b_i,s_j,\theta)\sigma_j^k(t_j)(s_j)d\beta_i(\tilde{t}_i)\\ &= \int_{\Theta\times T_j^{\mathcal{H}}} u_i(b_i,s_j,\theta)\sigma_j^k(\varphi_j(t_j^{\mathcal{H}}))(s_j)d\beta_i^{\mathcal{H}}(\tilde{t}_i^{\mathcal{H}}),\end{aligned}$$

where $\tilde{t}_i := \varphi_i(\tilde{t}_i^{\mathcal{H}})$. The first and third lines use the standard change-of-variables result and that φ is a type morphism, and the second line uses the best-response property. □

B.2 Proof of Lemma 5.3

We start with an auxiliary result.

Lemma B.1. Let $\mathcal{T}^k$ be a depth-k space that is kth-order nonredundant, and let $(\mathcal{T}^{\mathcal{H}},\varphi)$ be a Harsanyi extension of $\mathcal{T}^k$. For each $i \in N$ and $t_i^{\mathcal{H}}, \tilde{t}_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$, we have that $\varphi_i(t_i^{\mathcal{H}}) = \varphi_i(\tilde{t}_i^{\mathcal{H}})$ if and only if $h_i^{\mathcal{T}^{\mathcal{H}},k}(t_i^{\mathcal{H}}) = h_i^{\mathcal{T}^{\mathcal{H}},k}(\tilde{t}_i^{\mathcal{H}})$.

The proof is standard, and is included in the online appendix. We are now ready to prove Lemma 5.3. Let $\mathcal{T}^k$ be a depth-k space that is kth-order nonredundant, and let $(\mathcal{T}^{\mathcal{H}},\varphi)$ be a Harsanyi extension of $\mathcal{T}^k$. Suppose that for every game $\mathcal{G}$, for every Bayesian-Nash equilibrium $\sigma = (\sigma_i)_{i\in N}$ of $(\mathcal{G}, \mathcal{T}^{\mathcal{H}})$, there is a corresponding equilibrium of $(\mathcal{G}, \mathcal{T}^k)$. Fix a game $\mathcal{G}$, and let σ be a Bayesian-Nash equilibrium of $(\mathcal{G},\mathcal{T}^{\mathcal{H}})$. Since there is a corresponding equilibrium of $(\mathcal{G}, \mathcal{T}^k)$, that is, the strategy profile σ^k with $\sigma_i = \sigma_i^k \circ \varphi_i$ for $i \in N$ is an equilibrium of $(\mathcal{G}, \mathcal{T}^k)$, we have that for any $t_i^{\mathcal{H}}, \tilde{t}_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$,

$$\varphi_i(t_i^{\mathcal{H}}) = \varphi_i(\tilde{t}_i^{\mathcal{H}}) \quad \Longrightarrow \quad \sigma_i(t_i^{\mathcal{H}}) = \sigma_i(\tilde{t}_i^{\mathcal{H}}).$$

By Lemma B.1, therefore, it follows that for any $t_i^{\mathcal{H}}, \tilde{t}_i^{\mathcal{H}} \in T_i^{\mathcal{H}}$,

$$h_i^{\mathcal{T}^{\mathcal{H}},k}(t_i^{\mathcal{H}}) = h_i^{\mathcal{T}^{\mathcal{H}},k}(\tilde{t}_i^{\mathcal{H}}) \quad \Longrightarrow \quad \sigma_i(t_i^{\mathcal{H}}) = \sigma_i(\tilde{t}_i^{\mathcal{H}}). \tag{B.1}$$

Since for each Harsanyi type space $\widetilde{\mathcal{T}}^{\mathcal{H}}$, there is a game $\widetilde{\mathcal{G}}$ and a Bayesian-Nash equilibrium $\widetilde{\sigma}$ of $(\widetilde{\mathcal{G}}, \widetilde{\mathcal{T}}^{\mathcal{H}})$ such that (B.1) does not hold if $h_i^{\mathcal{T}^{\mathcal{H}},m}(t_i^{\mathcal{H}}) \neq h_i^{\mathcal{T}^{\mathcal{H}},m}(\tilde{t}_i^{\mathcal{H}})$ for some $m \geq k$,[29] we must have that

$$h_i^{\mathcal{T}^{\mathcal{H}},k}(t_i^{\mathcal{H}}) = h_i^{\mathcal{T}^{\mathcal{H}},k}(\tilde{t}_i^{\mathcal{H}}) \quad \Longrightarrow \quad h_i^{\mathcal{T}^{\mathcal{H}},m}(t_i^{\mathcal{H}}) = h_i^{\mathcal{T}^{\mathcal{H}},m}(\tilde{t}_i^{\mathcal{H}})$$

for all $m \geq k$. That is, $\mathcal{T}^{\mathcal{H}}$ is an order-k extension of $\mathcal{T}^k$. If $\mathcal{T}^{\mathcal{H}}$ is nonredundant, it is thus without loss of generality to take $T_i^{\mathcal{H}} = T_i$ for each $i \in N$. □

[29] For instance, take $\widetilde{\mathcal{G}}$ to be the game in the proof of Proposition 5.4.

References

Alaoui, L. and A. Penta (2013). Strategic thinking and the incentives to reason. Working paper, UW Madison and U Pompeu Fabra.

Battigalli, P., A. Di Tillio, E. Grillo, and A. Penta (2011). Interactive epistemology and solution concepts for games with asymmetric information. *The B.E. Journal of Theoretical Economics 11*(1 (Advances)), Article 6.

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.

Brocas, I., C. Camerer, J. Carrillo, and S. Wang (2009). Measuring attention and strategic behavior in games with private information. Working paper, Caltech.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica 69*, 1193–1235.

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.

Crawford, V. P. and N. Iriberri (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winners curse and overbidding in private-value auctions? *Econometrica 75*, 1721–1770.

Dekel, E., D. Fudenberg, and D. Levine (2004). Learning to play Bayesian games. *Games and Economic Behavior 46*, 282–303.

Dekel, E., D. Fudenberg, and S. Morris (2006). Topologies on types. *Theoretical Economics 1*, 275–309.

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

Ershov, M. P. (1974). Extension of measures and stochastic equations. *Theory of Probability and its Applications XIX*, 431–444.

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

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

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. and W. Kets (2011). All types naive and canny. Working paper, Northwestern University.

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

Ho, T.-H., C. Camerer, and K. Weigelt (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review 88*, 947–969.

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

Kets, W. (2009). Do you think about what I think you think? Finite belief hierarchies in games. Working paper, Stanford University.

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

Kets, W. (2013). Common belief with 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.

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.

Morris, S., A. Postlewaite, and H. Shin (1995). Depth of knowledge and the effect of higher order uncertainty. *Economic Theory 6*, 453–467.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review 85*, 1313–1326.

Purves, R. (1966). Bimeasurable functions. *Fundamenta Mathematicae 58*, 149–157.

Qin, C. and C. Yang (2013). An explicit approach to modeling finite-order type spaces and applications. *Journal of Economic Theory*. Forthcoming.

Rogers, B. W., T. R. Palfrey, and C. F. Camerer (2009). Heterogeneous quantal response equilibrium and cognitive hierarchies. *Journal of Economic Theory 144*, 1440–1467.

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.

Stahl, D. O. and P. W. Wilson (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior 10*, 218–254.

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

Weinstein, J. and M. Yildiz (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica 75*, 365–400.