

# RATIONAL BELIEF HIERARCHIES<sup>1</sup>

ELIAS TSAKAS

We consider agents whose language can only express probabilistic beliefs that attach a rational number to every event. We call these probability measures rational. We introduce the notion of a rational belief hierarchy, where the first order beliefs are described by a rational measure over the fundamental space of uncertainty, the second order beliefs are described by a rational measure over the product of the fundamental space of uncertainty and the opponent's first order rational beliefs, and so on. Then, we derive the corresponding (rational) type space model, thus providing a Bayesian representation of rational belief hierarchies. Our first main result shows that this type-based representation violates our intuitive idea of an agent whose language expresses only rational beliefs, in that there are rational types associated with non-rational beliefs over the canonical state space. We rule out these types by focusing on the rational types that satisfy common certainty in the event that everybody holds rational beliefs over the canonical state space. We call these types universally rational and show that they are characterized by a bounded rationality condition which restricts the agents' computational capacity. Moreover, the universally rational types form a dense subset of the universal type space. Finally, we show that the strategies rationally played under common universally rational belief in rationality generically coincide with those satisfying correlated rationalizability.

KEYWORDS: Epistemic game theory, bounded rationality, rational numbers, belief hierarchies, type spaces, unawareness, computational capacity, common belief in rationality.

## 1. INTRODUCTION

A belief hierarchy is a description of an agent's beliefs about some fundamental space of uncertainty, beliefs about everybody else's beliefs, and so on. During the past few decades, belief hierarchies have become an integral tool of modern economic theory, often used to analyze games with incomplete information (Harsanyi, 1967-68), as well as in order to provide epistemic characterizations for several solution concepts, such as rationalizability (Brandenburger and Dekel, 1987; Tan and Werlang, 1988), Nash equilibrium (Aumann and Brandenburger, 1995), and correlated equilibrium (Aumann, 1987), just to mention a few<sup>1</sup>.

Belief hierarchies are in general very complex objects, consisting of infinite sequences of probability measures. This makes them in principle very hard to handle and sometimes even to describe, especially when it comes to high order beliefs. Having recognized this difficulty, Harsanyi (1967-68)

---

<sup>1</sup>I am indebted to...

<sup>1</sup>For an overview of the epistemic game theory literature we refer to the textbook by Perea (2012) or the review article by Brandenburger (2008).

proposed an indirect Bayesian representation of belief hierarchies, known as the type space model<sup>2</sup>. Formally, Harsanyi’s model consists of a set of *types* for each agent and a continuous function mapping each type to the corresponding *conditional beliefs* over the product of the fundamental space of uncertainty and the opponent’s type space. This structure induces a belief hierarchy for every type, thus reducing the infinite-dimensional regression of beliefs to a single-dimensional type. [Mertens and Zamir \(1985\)](#) and [Brandenburger and Dekel \(1993\)](#) independently completed the analysis by showing the existence of the universal type space, which represents all belief hierarchies satisfying some standard coherency properties.

In this paper we restrict attention to agents whose language does not contain sentences of the form “*the event  $E$  occurs with probability  $p$* ”, where  $p$  is an irrational number, e.g., Alexandra cannot express the sentence “*tomorrow it will rain with probability  $\sqrt{2}/2$* ” within the bounds of her language. This is for instance the case when she is unaware of the notion of irrational numbers ([Modica and Rustichini, 1999](#); [Heifetz et al., 2006](#)). Alternatively, she may be aware of the fact that there are irrational numbers in the interval  $[0, 1]$ , but still conceptually not understand them ([Lehtinen et al., 1997](#)). This type of beliefs are modeled by probability measures that attach a rational number to every Borel event. We call these measures *rational*.

Obviously, if an agent understands only rational numbers, it is not only her first order beliefs that are restricted but also her beliefs about everybody else’s beliefs, and so on. For instance, besides Alexandra not understanding the sentence “*tomorrow it will rain with probability  $\sqrt{2}/2$* ”, she also does not understand either the sentence “*with probability  $\sqrt{3}/3$ , Barney attaches probability  $p$  to tomorrow raining*”, or the sentence “*with probability  $q$ , Barney attaches probability  $\sqrt{5}/5$  to tomorrow raining*”. In other words, the belief hierarchy of an agent whose language is restricted to express only rational beliefs consists of a sequence of rational probability measures, where the first order (rational) beliefs are described by a rational measure over the underlying space of uncertainty, the second order (rational) beliefs are described by a rational measure over the product of the fundamental space of uncertainty and the opponent’s space of rational first order beliefs, and so on. We call this infinite regression of probability measures *rational belief hierarchy*.

Following [Mertens and Zamir \(1985\)](#) and [Brandenburger and Dekel \(1993\)](#), we construct a Harsanyi type space representation of rational belief hierarchies. However, as our first main result (Theorem 4.1) shows, this Bayesian representation has an odd and at the same time undesirable property. Namely, it contains rational types which are represented by *non-rational* probability measures over

---

<sup>2</sup>Later, [Aumann \(1976\)](#) introduced an alternative representation, the partitional model, which is essentially equivalent to Harsanyi’s type-based structure.

the product of the fundamental space of uncertainty and the opponent’s rational type space. In other words, there is some Borel event in the canonical state space<sup>3</sup> to which this rational type attaches an irrational conditional probability even though every order of her belief hierarchy involves only rational beliefs. Obviously, the latter contradicts our initial idea of an agent whose language cannot describe non-rational probabilistic beliefs.

In order to make our formal model consistent with our intuitive idea of agents who do not understand irrational numbers, we consider only those rational types that (i) are associated with rational conditional beliefs over the canonical state space, and (ii) satisfy common certainty in the event that everybody’s conditional beliefs are rational. We call these types *universally rational (U-rational)*, and we show the existence of a Bayesian representation of these types in the same line as [Mertens and Zamir \(1985\)](#) and [Brandenburger and Dekel \(1993\)](#). Notice that the space of *U-rational* types is derived by imposing a sequence of restrictions in the associated type space model, and therefore it is not straightforward how these conditions translate into restrictions on the space of rational belief hierarchies, which at the end of the day is our primitive concept. Our second main result (Theorem 5.1) fills this gap by characterizing the space of universally rational types in terms of a second bounded rationality condition imposed directly on the belief hierarchies. This condition restricts the agent’s computational capacity, and more specifically her ability to divide natural numbers. In our context, we say that an agent’s computational capacity is bounded by a finite subset of the natural numbers  $N \subseteq \mathbb{N}$ , whenever the agent’s belief hierarchy consists of a sequence of measures that assign probabilities that can be written as ratios with the denominator belonging to  $N$ . In other words, bounded computational capacity means that the agent can only divide with finitely many natural numbers. Then, we prove that a type is *U-rational* if and only if it (i) has bounded computational capacity, and (ii) satisfies common certainty in the event that everybody has bounded computational capacity. Notice that the previous conditions do not require the agent to believe that everybody else has the same computational capacity as herself, e.g., Alexandra’s computational capacity may be bounded by  $\{1, \dots, n-1\}$  and still believe that Barney attaches probability  $1/n$  to some event. This also implies that bounded computational capacity weaker than bounded language.

Although the space of universally rational types satisfies quite a few restrictions, we can still show that it is rather rich. Namely, the space of *U-rational* types forms a dense subset of the universal type space, implying that every belief hierarchy can be approximated by a sequence of types whose language cannot express irrational beliefs. This may be of particular interest from a practical point

---

<sup>3</sup>The canonical state space is defined as the product of the fundamental space of uncertainty and the type spaces of each agent.

of view, in that we often prefer to use “small” type spaces, but which are still sufficiently rich to approximate every belief hierarchy at an arbitrarily high order (Lipman, 2003; Dekel et al., 2006). The latter has recently attracted quite a lot of attention among authors who study the robustness of solution concepts in incomplete information games with regards to small perturbations in the belief hierarchies (Ely and Peşki, 2006; Dekel et al., 2007).

In the last part of the paper, we introduce the concept of common (universally) rational belief in rationality, as the natural analogue to common belief in rationality, which is the set of epistemic conditions that characterize correlated rationalizability (Brandenburger and Dekel, 1987; Tan and Werlang, 1988; Friedenberg and Keisler, 2011). Obviously, since the universal type space is a superset of the set of  $U$ -rational types, the strategy profiles rationally played under common  $U$ -rational belief in rationality will be a subset of the ones rationally played under common belief in rationality, which are the ones surviving iterated elimination of strictly dominated strategies. Hence, our concept is in general a refinement of correlated rationalizability. However, as we show this is the case in a very small subset of all games, as the two concepts yield the same outcomes generically (Theorem 7.1). Observe that common  $U$ -rational belief in rationality differs from other concepts in the literature that place *ex ante* restrictions on the type space (Battigalli and Siniscalchi, 2003; Battigalli and Friedenberg, 2012). The reason is that we restrict the entire belief hierarchy, rather than just the first order beliefs.

This paper belongs to a growing new strand of research within epistemic game theory, that of bounded reasoning in games. This literature is actually the product of introducing bounded rationality assumptions to the players’ belief formation and/or understanding of the game, and has developed parallelly to the emergence of related empirical evidence in the experimental economics literature. Examples include players who are not fully aware of all elements of the game (Dekel et al., 1998; Halpern, 2001; Heifetz et al., 2006; Li, 2009), players with finite depth of reasoning (Kets, 2010; Heifetz and Kets, 2011; Strzalecki, 2011), players with ambiguous beliefs (Ahn, 2007), or players whose beliefs are only finitely additive (Meier, 2006).

The paper is structured as follows: In Section 2 we introduce rational probability measures and we prove some of their properties; Section 3 extends this framework to an interactive setting by introducing rational belief hierarchies; In Section 4 we construct the rational type space and show the existence of a type associated with an non-rational measure over the canonical state space; In Section 5 we introduce the universally rational types, provide their corresponding Bayesian representation, as well as their foundation in terms of a second bounded rationality assumption; Section 6 contains some topological properties of rational type spaces; In Section 7, we show that the outcomes

rationally played under common universally rational belief in rationality generically coincide with those surviving iterated elimination of strictly dominated strategies; Section 8 concludes.

## 2. RATIONAL PROBABILITY MEASURES

We begin with some definitions and the basic notation. Let  $X$  be a Polish<sup>4</sup> space, together with the Borel  $\sigma$ -algebra,  $\mathcal{B}$ . As usual,  $\Delta(X)$  denotes the space of probability measures on  $(X, \mathcal{B})$ , endowed with the topology of weak convergence<sup>5</sup>. For each  $\mu \in \Delta(X)$ , let  $\text{supp}(\mu)$  denote the support, i.e., the smallest closed subset<sup>6</sup> of  $X$  that receives probability 1 by  $\mu$ .

Consider the Borel probability measures that assign to every Borel event a rational number.

DEFINITION 1 We define the set of *rational probability measures* by

$$(2.1) \quad \Delta^{\mathbb{Q}}(X) := \left\{ \mu \in \Delta(X) : \mu(B) \in \mathbb{Q}, \forall B \in \mathcal{B} \right\}.$$

We use rational probability measures to model the beliefs of an agent whose language does not contain sentences of the form “ $E$  occurs with probability  $p$ ”, where  $p$  is an irrational number. This idea is strongly related to one underlying the notion of awareness, in that an agent is aware of an event if and only if this event can be described by a sentence within her language (Modica and Rustichini, 1999; Heifetz et al., 2006). We come back to the relationship of our work to the literature on unawareness later in the paper.

Below, we provide some results on rational probability measures, which we will use later in the paper. Throughout this section, unless stated otherwise, we assume that  $X$  is separable and metrizable.

PROPOSITION 2.1 *Every  $\mu \in \Delta^{\mathbb{Q}}(X)$  has a finite support.*

The previous, quite surprising result rules out all probability measures with countably infinite support, even if each point in the support receives a rational probability. The following example illustrates such a case.

---

<sup>4</sup>A topological space is called Polish whenever it is separable and completely metrizable. Examples of Polish spaces include countable sets endowed with the discrete topology and  $\mathbb{R}^n$  together with the usual topology. Closed subsets of Polish spaces endowed with the relative topology are Polish. The countable product of Polish spaces, together with the product topology, is also Polish.

<sup>5</sup>The topology of weak convergence, which is usually denoted by  $w^*$ , is the coarsest topology that makes the mapping  $\mu \mapsto \int f d\mu$  continuous, for every bounded and continuous real-valued function,  $f$ . If  $X$  is Polish, then  $\Delta(X)$  endowed with the topology of weak convergence is also Polish. For further properties of  $w^*$ , we refer to Aliprantis and Border (1994, Ch. 15).

<sup>6</sup>If  $X$  is separable and metrizable, the support is unique (Parthasarathy, 1967, Thm. 2.1).

EXAMPLE 2.1 Let  $X = \{x_1, x_2, \dots\}$  and suppose that  $\mu \in \Delta(X)$  assigns probability  $2^{-k}$  to each  $x_k$ . It is straightforward verifying that  $\mu$  is a probability measure. Now, consider an arbitrary  $\xi \in (0, 1)$ , and construct the Borel subset  $B_\xi \subseteq X$  so that  $x_k \in B_\xi$  if and only if  $\xi \in [\frac{1}{2^k}, \frac{2}{2^k}) \cup \dots \cup [\frac{2^k-1}{2^k}, 1)$ . Observe that  $\xi = \mu(B_\xi)$ , implying that for every irrational  $\xi$  there is a Borel event receiving an irrational probability, and therefore  $\mu$  is not a rational measure.  $\triangleleft$

The following result proves that every Borel probability measure can be approximated by a sequence of rational probability measures (in the topology of weak convergence).

PROPOSITION 2.2  $\Delta^{\mathbb{Q}}(X)$  is dense in  $\Delta(X)$ .

Before moving forward, let us first introduce the concept of *N-rational probability measures*. For some finite  $N \subseteq \mathbb{N}$ , consider the subset of the rational numbers,

$$\mathbb{Q}_N := \{m/n : m = 0, \dots, n ; n \in N\},$$

and define the set of *N-rational probability measures* by

$$(2.2) \quad \Delta^N(X) := \left\{ \mu \in \Delta(X) : \mu(B) \in \mathbb{Q}_N, \forall B \in \mathcal{B} \right\}.$$

Obviously, if  $N \subseteq M \subseteq \mathbb{N}$ , then  $\Delta^N(X) \subseteq \Delta^M(X)$ .

We use *N-rational probability measures* to describe the beliefs of an agent whose computational capacity is bounded in that, not only can she not attach an irrational probability to any event, but moreover she can only divide with finitely many numbers, i.e., those in  $N$ .

PROPOSITION 2.3  $\Delta^N(X)$  is closed in  $\Delta(X)$ .

Obviously, it follows by Proposition 2.2 that  $\Delta^N(X)$  is not closed if  $N$  is infinite. The following result is a direct consequence of Proposition 2.3.

PROPOSITION 2.4  $\Delta^{\mathbb{Q}}(X)$  is a Borel subset of  $\Delta(X)$ .

### 3. RATIONAL BELIEF HIERARCHIES

Let  $\Theta$  be a Polish space together with the Borel  $\sigma$ -algebra,  $\mathcal{B}_0$ . In a game, each  $\theta \in \Theta$  corresponds to a payoff vector (Harsanyi, 1967-68), or a strategy profile (Aumann and Brandenburger, 1995; Tan and Werlang, 1988), or a combination of the two. Throughout the paper, we refer to  $\Theta$  as the underlying — else called, fundamental — space of uncertainty. Let  $I = \{a, b\}$  be the set of agents<sup>7</sup>,

<sup>7</sup>Our analysis can be directly generalized to any finite set of agents, in which case we obviously allow for correlated beliefs, as usual.

with typical elements  $i$  and  $j$ . Each agent forms beliefs about  $\Theta$  (first order beliefs), beliefs about the opponent's beliefs about  $\Theta$  (second order beliefs), and so on. Such a sequence is called a belief hierarchy.

Formally, consider the following sequence of Polish spaces:

$$\begin{aligned}\Psi_0 &:= \Theta \\ \Psi_1 &:= \Psi_0 \times \Delta(\Psi_0) \\ &\vdots \\ \Psi_{k+1} &:= \Psi_k \times \Delta(\Psi_k) \\ &\vdots\end{aligned}$$

A belief hierarchy is an element of

$$(3.1) \quad T_0 := \prod_{k=0}^{\infty} \Delta(\Psi_k).$$

For some  $(\pi_1, \pi_2, \dots) \in T_0$ , the Borel probability measure  $\pi_k \in \Delta(\Psi_{k-1})$  denotes the  $k$ -th order beliefs.

In this paper, we consider agents whose language can express only rational beliefs. That is, for some  $\theta \in \Theta$ , agent  $i$  does not understand the sentence:

$$\boxtimes \theta \text{ occurs with probability } \sqrt{2}/2$$

The latter implies that we restrict  $i$ 's first order beliefs to the space of rational probability measures. Furthermore,  $i$  does not understand any of the following sentences:

$$\begin{aligned}\boxtimes \text{ the event that "j assigns probability } \sqrt{2}/2 \text{ to } \theta \text{ " occurs with probability } p \\ \boxtimes \text{ the event that "j assigns probability } q \text{ to } \theta \text{ " occurs with probability } \sqrt{3}/3\end{aligned}$$

That is, we restrict  $i$ 's second order beliefs to rational probability measures over the space of rational probability measures. Likewise, we restrict higher order beliefs. These restrictions are consistent with the interpretation of rational beliefs as the states of mind of an agent who is unaware of the concept of irrational numbers. We further elaborate on this later in the paper.

Formally, consider the sequence

$$\begin{aligned}\Theta_0 &:= \Theta \\ \Theta_1 &:= \Theta_0 \times \Delta^{\mathbb{Q}}(\Theta_0) \\ &\vdots \\ \Theta_{k+1} &:= \Theta_k \times \Delta^{\mathbb{Q}}(\Theta_k) \\ &\vdots\end{aligned}$$

A *rational belief hierarchy* is a sequence  $(\pi_1, \pi_2, \dots)$ , with  $\pi_k \in \Delta^{\mathbb{Q}}(\Theta_{k-1})$  denoting the  $k$ -th order beliefs. Let

$$(3.2) \quad T_0^{\mathbb{Q}} := \prod_{k=0}^{\infty} \Delta^{\mathbb{Q}}(\Theta_k)$$

denote the space of all rational belief hierarchies, endowed with the product topology.

Intuitively, rational belief hierarchies form a strict subset of all belief hierarchies. However, observe that formally  $T_0^{\mathbb{Q}}$  is not a subset of  $T_0$ , because strictly speaking  $\Delta^{\mathbb{Q}}(\Theta_k)$  is not a subset of  $\Delta(\Psi_k)$ . Therefore, before moving forward, we would like to make sure that the intuitive idea of one being a subset of the other is compatible with our formal model. The following result serves this purpose, by showing that  $T_0^{\mathbb{Q}}$  is embedded as a Borel subset of  $T_0$ .

**PROPOSITION 3.1**  *$T_0^{\mathbb{Q}}$  is homeomorphic to a Borel subset of  $T_0$ .*

Throughout the paper, we denote this embedding by

$$(3.3) \quad h : T_0^{\mathbb{Q}} \hookrightarrow T_0.$$

As usual, with slight abuse of terminology, whenever we talk about a rational belief hierarchy  $t \in T_0^{\mathbb{Q}}$  we actually refer to its image  $h(t) \in T_0$ , and therefore we will informally consider  $T_0^{\mathbb{Q}}$  to be a Borel subset of  $T_0$ .

#### 4. RATIONAL TYPES

In general, belief hierarchies are very large and complex objects, and as such it is really hard directly working with them. [Harsanyi \(1967-68\)](#) was the first one to circumvent this problem by proposing a compact way of expressing belief hierarchies, known in the literature as the *type space model*. Formally, this model consists of a tuple  $(\Theta, T_a, T_b, g_a, g_b)$ , where  $T_i$  is a Polish space of types with typical element  $t_i$ , and  $g_i : T_i \rightarrow \Delta(\Theta \times T_j)$  is a continuous function. In a type space, each  $t_i \in T_i$  is associated with a unique *belief hierarchy* derived as follows: For each type  $t_i \in T_i$ , the *first order beliefs*,  $\pi_1(t_i) \in \Delta(\Psi_0)$ , attach probability

$$(4.1) \quad \pi_1(t_i)(B_0) = \int_{(\theta, t_j): \theta \in B_0} dg_i(t_i)$$

to every Borel event  $B_0 \subseteq \Psi_0$ . The *second order beliefs*,  $\pi_2(t_i) \in \Delta(\Psi_1)$ , attach probability

$$(4.2) \quad \pi_2(t_i)(B_1) = \int_{(\theta, t_j): (\theta, \pi_1(t_j)) \in B_1} dg_i(t_i)$$

to every Borel event  $B_1 \subseteq \Psi_1$ . Likewise, the  $k$ -th order beliefs,  $\pi_k(t_i) \in \Delta(\Psi_{k-1})$ , attach probability

$$(4.3) \quad \pi_k(t_i)(B_{k-1}) = \int_{(\theta, t_j): (\theta, \pi_1(t_j), \dots, \pi_{k-1}(t_j)) \in B_{k-1}} dg_i(t_i)$$

for every Borel subset  $B_{k-1} \subseteq \Psi_{k-1}$ . For a detailed presentation on how the entire belief hierarchy is derived from a type space model, we refer to [Siniscalchi \(2007\)](#).

Observe that every type space induces a state space  $\Theta \times T_i \times T_j$ , and each  $t_i \in T_i$  is mapped to a probability measure  $g_i(t_i)$ , which can be viewed as a conditional belief over the state space given the information set  $\Theta \times \{t_i\} \times T_j$ . We say that  $g_i$  induces a *Bayesian representation* of  $T_i$  whenever  $g_i : T_i \rightarrow \Delta(\Theta \times T_j)$  is injective, thus ruling out the possibility of two different types being associated with the same measure. We call a Bayesian representation *complete*, if  $g_i$  is also surjective, implying that every measure in  $\Delta(\Theta \times T_j)$  is the image of some type in  $T_i$ .

Later, [Mertens and Zamir \(1985\)](#), and [Brandenburger and Dekel \(1993\)](#) independently showed that Harsanyi's framework is sufficiently rich to model all instances of interactive uncertainty, in that there is a type space model  $(\Theta, T_a^*, T_b^*, g_a^*, g_b^*)$ , with  $T_a^* = T_b^* = T^*$  and  $g_a^* = g_b^* = g^*$ , such that (i)  $T^*$  coincides with the set of all belief hierarchies satisfying certain standard coherency restrictions, and (ii)  $g^*$  induces a complete Bayesian representation of  $T^*$ . This construction is called the universal type space<sup>8</sup>.

They started by imposing a standard coherency condition, which states that the  $k$ -th and  $(k+1)$ -th order beliefs cannot contradict each other. Formally, let  $T_c := \{(\pi_1, \pi_2, \dots) \in T_0 : \text{marg}_{\Psi_{k-2}} \pi_k = \pi_{k-1}, \forall k > 1\}$ . Then, they showed ([Brandenburger and Dekel, 1993](#), Prop. 1) that there is a homeomorphism

$$(4.4) \quad f^* : T_c \rightarrow \Delta(\Theta \times T_0).$$

This homeomorphism is a natural one, in that for all  $(\pi_1, \pi_2, \dots) \in T_c$ ,

$$(4.5) \quad \text{marg}_{\Psi_{k-1}} f^*(\pi_1, \pi_2, \dots) = \pi_k.$$

Then, they further restricted attention to belief hierarchies that satisfy, not only coherency, but also

---

<sup>8</sup>[Heifetz \(1993\)](#) generalized this representation result to cases where the underlying space of uncertainty is Hausdorff, while [Heifetz and Samet \(1998\)](#) further considered a purely measurable underlying space of uncertainty.

common certainty in coherency. Formally, consider the following sequence of subsets of  $T_c$ :

$$\begin{aligned} T_1 &:= T_c \\ T_2 &:= \{ t \in T_c : f^*(t)(\Theta \times T_1) = 1 \} \\ &\vdots \\ T_k &:= \{ t \in T_c : f^*(t)(\Theta \times T_{k-1}) = 1 \} \\ &\vdots \end{aligned}$$

Observe that  $T_1$  contains the belief hierarchies satisfying coherency,  $T_2$  those satisfying certainty in everybody's coherency, and so on. Thus,

$$(4.6) \quad T^* := \bigcap_{k=1}^{\infty} T_k$$

contains the belief hierarchies satisfying coherency and common certainty in coherency. Finally, [Brandenburger and Dekel \(1993, Prop. 2\)](#) showed that there is a homeomorphism

$$(4.7) \quad g^* : T^* \rightarrow \Delta(\Theta \times T^*),$$

implying that there is a complete Bayesian representation of  $T^*$ .

The first natural question arising at this point is whether we can extend their result to the case of rational belief hierarchies. In other words, is there a type space representation of rational belief hierarchies in the same line as the standard results by [Mertens and Zamir \(1985\)](#) and [Brandenburger and Dekel \(1993\)](#).

We retain the standard coherency restriction. Formally, let

$$(4.8) \quad \begin{aligned} T_c^Q &:= \{ (\pi_1, \pi_2, \dots) \in T_0^Q : \text{marg}_{\Theta_{k-2}} \pi_k = \pi_{k-1}, \forall k > 1 \} \\ &= T_0^Q \cap T_c \end{aligned}$$

denote the set of coherent belief hierarchies.

Similarly to [Brandenburger and Dekel \(1993, Prop. 1\)](#), the following result associates each coherent rational belief hierarchy to a probability measure over the product of the underlying space of uncertainty and the space of the opponent's rational hierarchies. This induces an injective mapping, implying that there is no pair of coherent types associated with the same distribution over  $\Theta \times T_0^Q$ .

**PROPOSITION 4.1** *There is an injection  $f : T_c^Q \rightarrow \Delta(\Theta \times T_0^Q)$ .*

It is rather easy to see that the function  $f$  is in fact the same as  $f^*$  from [Brandenburger and Dekel \(1993, Prop. 1\)](#), but restricted to rational belief hierarchies, i.e., for every  $t \in T_c^Q$  and each Borel

subset  $B \subseteq \Theta \times T_0^Q$ ,

$$(4.9) \quad f(t)(B) = f^*(h(t))(\{(\theta, t) \in \Theta \times T_0 : (\theta, h^{-1}(t)) \in B\}).$$

Throughout the paper, we treat  $f$  and  $f^*$  as the same function. Note that  $f$  inherits from  $f^*$  the property of being a natural mapping, in that every coherent hierarchy is associated with a probability measure over  $\Theta \times \prod_{k=0}^{\infty} \Delta^{\mathbb{Q}}(\Theta_k)$  that has the property that its marginal distribution over  $\Theta_{k-1}$  coincides with the  $k$ -th order beliefs induced by this hierarchy, i.e., for every  $(\pi_1, \pi_2, \dots) \in T_c^Q$

$$(4.10) \quad \text{marg}_{\Theta_{k-1}} f(\pi_1, \pi_2, \dots) = \pi_k.$$

As usual, we further restrict belief hierarchies so that they satisfy, not only coherency, but also common certainty in coherency. Formally, consider the following sequence of subsets of  $T_c^Q$ :

$$\begin{aligned} T_1^Q &:= T_c^Q \\ T_2^Q &:= \{t \in T_c^Q : f(t)(\Theta \times T_1^Q) = 1\} \\ &\vdots \\ T_k^Q &:= \{t \in T_c^Q : f(t)(\Theta \times T_{k-1}^Q) = 1\} \\ &\vdots \end{aligned}$$

Note that  $T_1^Q$  contains the belief hierarchies that satisfy coherency,  $T_2^Q$  contains the belief hierarchies that satisfy certainty in everybody coherency, and so on. Thus, the types in

$$(4.11) \quad T^Q := \bigcap_{k=1}^{\infty} T_k^Q$$

satisfy coherency and common certainty in coherency. Henceforth, whenever we write “rational belief hierarchies” or “rational types”, we implicitly refer to elements of  $T^Q$ , thus omitting to explicitly say that they satisfy coherency and common certainty in coherency. The following result proves the existence of a *Bayesian representation of the space of rational belief hierarchies*, implying that every type in  $T^Q$  is identified by a probability measure on  $\Theta \times T^Q$ .

**PROPOSITION 4.2** *There is an injection  $g : T^Q \rightarrow \Delta(\Theta \times T^Q)$ .*

Once again,  $g$  coincides with the corresponding mapping  $g^*$  used by [Brandenburger and Dekel \(1993, Prop. 2\)](#) when restricted on the domain  $h(T^Q)$ . That is, for every  $t \in T^Q$  and each Borel subset  $B \subseteq \Theta \times T^Q$ ,

$$(4.12) \quad g(t)(B) = g^*(h(t))(\{(\theta, t) \in \Theta \times T^Q : (\theta, h^{-1}(t)) \in B\}).$$

Obviously, the representation induced by  $g$  is not complete, in that there are measures in  $\Delta(\Theta \times T^Q)$  which are not the image of any rational type. The latter is not surprising, as one can easily see that there exist probability measures  $\pi \in \Delta(\Theta \times T^Q)$  with  $\text{marg}_\Theta \pi \notin \Delta^Q(\Theta)$ , e.g., a measure with  $\pi(\{\theta\} \times T^Q) = \sqrt{2}/2$ .

What is really interesting, as well as far from obvious, is the conclusion of the following theorem. Let us first introduce the notion of a *rational representation*: Consider a type space model  $(\Theta, T_a, T_b, g_a, g_b)$  such that  $g_i$  induces a Bayesian representation of  $T_i$ . We call this representation rational whenever  $g_i(t) \in \Delta^Q(\Theta \times T_j)$  for every  $t \in T_i$ . The reason we are particularly interested in rational representations stems from our initial motivation for this project. Recall that we have in mind agents whose language does not contain sentences that express probabilistic beliefs of the form “ $E$  occurs with probability  $p$ ”, where  $p$  is an irrational number. Thus, if the representation induced by  $g$  is not rational, it means that there is some rational type  $t \in T^Q$  attaching an irrational probability to some Borel event in the canonical state space  $\Theta \times T^Q \times T^Q$ . Obviously, the latter would not be consistent with the type of agents we have in mind.

**THEOREM 4.1** *There is some  $t \in T^Q$  such that  $g(t) \notin \Delta^Q(\Theta \times T^Q)$ .*

The previous result has the following surprising implication for the Bayesian representation of rational belief hierarchies: Although the agent’s language can only describe rational beliefs, there are types  $t \in T^Q$  associated with irrational beliefs over the canonical state space  $\Theta \times T^Q \times T^Q$ .

## 5. UNIVERSALLY RATIONAL BELIEF HIERARCHIES

In this section, we restrict our attention to rational types whose language does not contain sentences that express non-rational probabilistic beliefs, i.e., we are interested in rational types which

- (i) are associated with a rational probability measure over the canonical state space,
  - (ii) are certain that the opponent’s beliefs satisfy (i),
  - (iii) are certain that the opponent’s beliefs satisfy (ii),
- and so on *ad infinitum*.

Observe that these conditions rule out not only types with non-rational beliefs over the state space, but also types that believe that their opponent has non-rational beliefs over the state space, and so on. The belief hierarchies that satisfy these restrictions are called universally rational. First, we prove that universally rational types have a complete rational Bayesian representation. Then, we provide a foundation for these types in terms of a second type of bounded rationality which restricts the computational capacity of the agent, and more specifically the agent’s ability to divide natural

numbers.

Formally, a type is *universally rational* (*U-rational*), whenever it belongs to each of the following subsets of  $T^Q$ :

$$\begin{aligned} T_0^U &:= \{ t \in T^Q : g(t) \in \Delta^Q(\Theta \times T^Q) \} \\ T_1^U &:= \{ t \in T^Q : g(t)(\Theta \times T_0^U) = 1 \} \\ &\vdots \\ T_k^U &:= \{ t \in T^Q : g(t)(\Theta \times T_{k-1}^U) = 1 \} \\ &\vdots \end{aligned}$$

Observe that  $T_0^U$  contains exactly the types satisfying condition (i) above,  $T_1^U$  contains exactly the types satisfying (ii), and so on. Throughout the paper we denote the set of *U-rational* types by

$$(5.1) \quad T^U := \bigcap_{k=0}^{\infty} T_k^U.$$

The following result shows that there is a *complete rational Bayesian representation of universally rational belief hierarchies*.

**PROPOSITION 5.1**  *$g : T^U \rightarrow \Delta^Q(\Theta \times T^U)$  is a homeomorphism.*

Note that the previous result is not the product of restrictions on the belief hierarchies. Instead, it rules out certain rational types based on a sequence of conditions on the corresponding Bayesian model. Below, we characterize the space of *U-rational* types in terms of an additional (bounded rationality) condition imposed directly on the beliefs hierarchies — which is our primitive notion — rather on the associated type space which is only a model describing the belief hierarchies.

Recall the notion of *N-rational* measures introduced in Section 2, and consider an agent whose computational capacity is restricted by some finite  $N \subseteq \mathbb{N}$ , in that she can only form *N-rational* beliefs. Intuitively, we have in mind an agent who exhibits a second type of bounded rationality in that she can only divide by finitely many denominators, on top of the fact that her language contains only sentences describing rational beliefs.

Notice that in principle we do not require the agent to believe that everybody else has the same computational capacity as her, i.e., agent *i*'s computational capacity may be restricted by some finite  $N \subseteq \mathbb{N}$  and still attach positive probability to the opponent's computational capacity being restricted by  $M \supsetneq N$ , or even believe that her opponent's computational capacity is not restricted by any finite  $M \subseteq \mathbb{N}$ . That is, bounded computational capacity is weaker than restricting the agent's language. We further discuss this distinction later in the paper.

Formally, for some finite  $N \subseteq \mathbb{N}$ , consider a sequence  $(\pi_1, \pi_2, \dots) \in \prod_{k=0}^{\infty} \Delta^N(\Theta_k)$ , where each  $\pi_k \in \Delta^N(\Theta_{k-1})$  denotes the ( $N$ -rational)  $k$ -th order beliefs. In other words, the agent forms  $N$ -rational beliefs about  $\Theta$  (first order beliefs),  $N$ -rational beliefs about  $\Theta \times \Delta^{\mathbb{Q}}(\Theta)$  (second order beliefs), and so on.

As we have already mentioned above, observe that the second order beliefs may attach positive probability to the opponent holding some rational, but not necessarily  $N$ -rational first order beliefs. This is the case, for instance, when the agent's computational capacity is restricted by  $\{1, \dots, n-1\}$ , but her language is not, e.g., she does understand the sentence “ $E$  occurs with probability  $1/n$ ” and she also deems possible that her opponent attaches probability  $1/n$  to  $E$ , but she herself never attaches probability  $1/n$  to any event because for instance it is computationally very costly to form beliefs beyond the bounds of  $\{1, \dots, n-1\}$ .

Let  $T_0^N := \prod_{k=0}^{\infty} \Delta^N(\Theta_k)$ , and impose the usual restriction of coherency and common certainty in coherency,

$$(5.2) \quad T^N := T_0^N \cap T^{\mathcal{Q}}.$$

Types in  $T^N$  are called  *$N$ -rational*. The following result shows that the agent's computational capacity is necessarily preserved by the Bayesian representation. That is, if an agent's belief hierarchy is restricted by some finite  $N \subseteq \mathbb{N}$ , then the agent attaches an  $N$ -rational probability to every Borel event in the canonical state space  $\Theta \times T^{\mathcal{Q}} \times T^{\mathcal{Q}}$  conditional on her own type. Moreover, every  $N$ -rational probability measure on  $\Theta \times T^{\mathcal{Q}}$  is the image of some  $N$ -rational type.

**PROPOSITION 5.2**  $g : T^N \rightarrow \Delta^N(\Theta \times T^{\mathcal{Q}})$  is homeomorphic.

Let  $\mathcal{N} := \{ \{1, \dots, n\} \mid n \in \mathbb{N} \}$ , and define the space of rational types whose computational capacity is bounded by some finite  $\{1, \dots, n\} \subseteq \mathbb{N}$ , i.e.,

$$(5.3) \quad T_0^{\mathcal{N}} := \bigcup_{N \in \mathcal{N}} T^N.$$

Then, we recursively define the types that (i) are certain that everybody's computational capacity is bounded (by some finite set of denominators), (ii) are certain that everybody is certain that everybody's computational capacity is bounded, and so on. Formally, let

$$\begin{aligned} T_1^{\mathcal{N}} &:= \{ t \in T_0^{\mathcal{N}} : g(t)(\Theta \times T_0^{\mathcal{N}}) = 1 \} \\ &\vdots \\ T_k^{\mathcal{N}} &:= \{ t \in T_0^{\mathcal{N}} : g(t)(\Theta \times T_{k-1}^{\mathcal{N}}) = 1 \} \\ &\vdots \end{aligned}$$

and define the types that have bounded computational capacity, and satisfy common certainty in everybody's computational capacity is bounded by

$$(5.4) \quad T^{\mathcal{N}} := \bigcap_{k=0}^{\infty} T_k^{\mathcal{N}}.$$

Throughout the paper we call these types  $\mathcal{N}$ -rational. The following result characterizes the  $U$ -rational types in terms of a restriction on the bounds of the agent's computational capacity. More specifically, we show that the  $U$ -rational types are exactly those that

- (i) have bounded computational capacity,
  - (ii) are certain that the opponent satisfies (i),
  - (iii) are certain that the opponent satisfies (ii),
- and so on.

In other words, a type is  $U$ -rational if and only if it is  $\mathcal{N}$ -rational.

**THEOREM 5.1**  $T^U = T^{\mathcal{N}}$ .

The previous result is quite interesting as it provides foundations for the set of  $U$ -rational types in terms of conditions imposed directly on the belief hierarchies rather than on the corresponding Bayesian model.

Note that although  $U$ -rational types are shown to be the ones restricted by bounded computational capacity and common certainty in bounded computational capacity, their language may not be restricted by any finite  $N \subseteq \mathbb{N}$ . We provide an example of such a case below.

**EXAMPLE 5.1** Consider the type space model  $(\Theta, T_a, T_b, g_a, g_b)$  such that  $\Theta = \{\theta_1, \theta_2\}$ , and  $T_a = \{t_a^1, t_a^2, \dots\}$  and  $T_b = \{t_b^1, t_b^2, \dots\}$ . Moreover, for each  $k > 0$ , suppose that  $g(t_a^k)$  assigns probability  $1/k$  to  $(\theta_1, t_b^k)$  and probability  $(k-1)/k$  to  $(\theta_2, t_b^k)$ , while  $g(t_b^k)$  assigns probability  $1/k$  to  $(\theta_1, t_a^{k+1})$  and probability  $(k-1)/k$  to  $(\theta_2, t_a^{k+1})$ .

First, observe that  $T_i \subseteq T^{\mathcal{N}}$  and therefore  $T_i \subseteq T^U$  for each  $i \in \{a, b\}$ . Second, notice that for every  $\ell > k$  agent  $t_i^k$ 's language contains the sentence “ $j$  attaches probability  $1/\ell$  to  $\theta_1$ ”, e.g.,  $t_a^k$

- ⊠ is certain that “ $b$  attaches probability  $1/k$  to  $\theta_1$ ”,
  - ⊠ is certain that  $b$  is certain that  $a$  is certain that “ $b$  attaches probability  $1/(k+1)$  to  $\theta_1$ ”,
  - ⊠ is certain that  $b$  is certain that  $a$  is certain that  $b$  is certain that  $a$  is certain that “ $b$  attaches probability  $1/(k+2)$  to  $\theta_1$ ”,
- and so on.

Hence,  $t_a^k$ 's language is not bounded by any finite  $N \subseteq \mathbb{N}$ , even though her computational capacity is bounded by  $\{k\} \in \mathcal{N}$ . ◁

The previous example illustrates the difference between restricting one’s language and restricting her computational capacity. Restricting the agent’s language implies that the agent cannot compute values expressed by sentences outside this language. However, the converse is not necessarily true, e.g., the agent may understand a sentence of the form “ $E$  occurs with probability  $p$ ” and still never hold these beliefs, simply because  $p$  is a very complex.

## 6. TOPOLOGICAL PROPERTIES OF RATIONAL TYPE SPACES

In this section we study the topological properties of the space of rational and universally rational types. In fact, as it turns out, these spaces are sufficiently rich in that every type in the universal type space can be approximated by a sequence of  $U$ -rational types.

**PROPOSITION 6.1**  *$T^U$  is dense in  $T^*$ .*

The previous result is particularly interesting for the analysis of robustness of solution concepts with respect to slight perturbations in the belief hierarchies in finite incomplete information games, as for instance in [Ely and Pęski \(2006\)](#), or [Dekel et al. \(2007\)](#).

## 7. COMMON RATIONAL BELIEF IN RATIONALITY

Consider a finite normal form game  $(I, (S_i)_{i \in I}, (U_i)_{i \in I})$ , where  $I = \{a, b\}$  denotes the finite set of players, with typical elements  $i$  and  $j$ . For every  $i \in I$ , let  $S_i$ , with typical element  $s_i$ , denote player  $i$ ’s finite set of (pure) strategies, and  $U_i : S_i \times S_j \rightarrow \mathbb{R}$  denote  $i$ ’s payoff function. A probability measure  $\mu_i \in \Delta(S_j)$  is called  $i$ ’s conjecture about the opponent’s action, with  $\mu_i(s_j)$  denoting the probability that  $i$  assigns to  $j$  playing  $s_j$ . Given the conjecture  $\mu_i$ , player  $i$ ’s (subjective) expected payoff from playing  $s_i \in S_i$  is equal to

$$u_i(s_i, \mu_i) := \sum_{s_j \in S_j} \mu_i(s_j) U_i(s_i, s_j).$$

We say that  $s_i$  is a best reply to the conjecture  $\mu_i$ , and we write  $s_i \in BR_i(\mu_i)$ , whenever  $u_i(s_i, \mu_i) \geq u_i(s'_i, \mu_i)$  for all  $s'_i \in S_i$ .

Since this is a normal-form game where players do not observe each other’s strategy before they choose their own, each  $i$  forms a belief hierarchy consisting of a conjecture about  $j$ ’s strategy, a conjecture about  $j$ ’s conjecture about  $i$ ’s strategy, and so on. As usual, we represent these belief hierarchies with a type space model,  $(S_a, S_b, T_a, T_b, g_a, g_b)$ , where  $T_i$  is a Polish type space, and  $g_i : T_i \rightarrow \Delta(S_j \times T_j)$  is a continuous function. This construction differs from the one described in Section 4, in that the two players form beliefs over different fundamental spaces, i.e., each  $i$  is

uncertain only about  $S_j$ . This is because, as usual, we focus on belief hierarchies restricted by the condition that players know their own actual strategy. An element  $(s_i, t_i) \in S_i \times T_i$  is called strategy-type pair of player  $i$ , whereas each  $(s_a, s_b, t_a, t_b) \in S_a \times S_b \times T_a \times T_b$  is called state (of the world). We say that  $i$  is rational at a state  $(s_a, s_b, t_a, t_b)$  whenever  $s_i$  is a best reply to the conjecture about  $j$ 's strategy induced by  $t_j$ .

Consider the set of states in  $S_a \times S_b \times T_a^* \times T_b^*$  where both players are rational:

$$R_0^* := \left\{ (s_a, s_b, t_a, t_b) \in S_a \times S_b \times T_a^* \times T_b^* : s_i \in BR_i(\text{marg}_{S_j} g_i^*(t_j)), \text{ for all } i \in \{a, b\} \right\}.$$

For every  $k > 0$ , inductively define

$$R_k^* := \left\{ (s_a, s_b, t_a, t_b) \in S_a \times S_b \times T_a^* \times T_b^* : g_i^*(t_i)(\text{marg}_{S_j \times T_j^*} R_{k-1}^*) = 1, \text{ for all } i \in \{a, b\} \right\},$$

and obtain the states that satisfy rationality and common belief in rationality, by

$$(7.1) \quad R^* := \bigcap_{k=0}^{\infty} R_k^*.$$

At these states, not only is everybody rational, but also everybody believes that everybody is rational, everybody believes that everybody believes that everybody is rational, and so on. The strategy profiles rationally played under common belief in rationality are the ones in

$$(7.2) \quad \begin{aligned} S^* &:= S_a^* \times S_b^* \\ &:= \text{proj}_{S_a} R^* \times \text{proj}_{S_b} R^*. \end{aligned}$$

These strategy profiles are exactly those played under correlated rationalizability, and are the ones surviving iterated elimination of strictly dominated strategies ([Brandenburger and Dekel, 1987](#)). In two-player games, they also coincide with the set of rationalizable<sup>9</sup> outcomes ([Bernheim, 1984](#); [Pearce, 1984](#)).

Now, suppose that we restrict attention to players with rational belief hierarchies. Belief hierarchies are represented by types in the type space  $(S_a, S_b, T_a^Q, T_b^Q, g_a, g_b)$ , where  $T_i^Q$  is the space of all rational belief hierarchies, and  $g_i : T_i^Q \rightarrow \Delta(S_j \times T_j^Q)$  is an injective mapping derived analogously to the one in Proposition 4.2. Let  $S_a \times S_b \times T_a^Q \times T_b^Q$  denote the set of states where both players have rational belief hierarchies.

---

<sup>9</sup>The equivalence between rationalizability and correlated rationalizability does not hold in games with more than two players ([Pearce, 1984](#), p. 1035).

Similarly to the unrestricted case, we recursively impose the following restrictions

$$\begin{aligned}
R_0^Q &:= \{ (s_a, s_b, t_a, t_b) \in S_a \times S_b \times T_a^Q \times T_b^Q : s_i \in BR_i(\text{marg}_{S_j} g_i(t_i)), \text{ for all } i \in \{a, b\} \} \\
&\vdots \\
R_k^Q &:= \{ (s_a, s_b, t_a, t_b) \in S_a \times S_b \times T_a^Q \times T_b^Q : g_i(t_i)(\text{marg}_{S_j \times T_j^Q} R_{k-1}^Q) = 1, \text{ for all } i \in \{a, b\} \} \\
&\vdots
\end{aligned}$$

thus obtaining the states satisfying *rationality and common rational belief in rationality* by

$$(7.3) \quad R^Q := \bigcap_{k=0}^{\infty} R_k^Q.$$

The strategy profiles rationally played under common rational belief in rationality are

$$\begin{aligned}
(7.4) \quad S^Q &:= S_a^Q \times S_b^Q \\
&:= \text{proj}_{S_a} R^Q \times \text{proj}_{S_b} R^Q.
\end{aligned}$$

Likewise, we can also define the states that satisfy *rationality and common universally rational belief in rationality*, denoted by  $R^U$ . Finally, let  $S^U$  denote the strategy profiles that can be played at some state in  $R^U$ . Obviously, it follows by construction that

$$(7.5) \quad S^U \subseteq S^Q \subseteq S^*.$$

In general, there are games where the second inclusion is strict<sup>10</sup>, e.g., there may exist pure strategies which are optimal only to irrational conjectures, as illustrated by the following example.

**EXAMPLE 7.1** Consider, for instance, the following game played between Alexandra and Barney, who choose among the row and column strategies respectively, i.e.,  $S_a = \{a_1, a_2, a_3\}$  and  $S_b = \{b_1, b_2\}$ . Observe that the only first order beliefs of Alexandra that would make  $a_3$  rational would be to

	$b_1$	$b_2$
$a_1$	$\sqrt{2}$ 0	0 2
$a_2$	0 2	$2 + \sqrt{2}$ 0
$a_3$	1 1	1 1

<sup>10</sup>Since we only consider finite games, the first inclusion is always an equality, i.e.,  $S^U = S^Q$ . This is because for all  $k > 0$ , and for every  $t \in T^Q$ , there is some  $t' \in T^U$  such that  $\pi_\ell(t) = \pi_\ell(t')$  for all  $\ell \leq k$ .

attach probability  $\sqrt{2}/2$  to the event that Barney plays  $b_1$ . Thus,  $a_3$  belongs to  $\text{proj}_{S_a} R_0^*$  but not to  $\text{proj}_{S_a} R_0^Q$ . Moreover, observe that  $\text{proj}_{S_a \times S_b} R_k^* = \text{proj}_{S_a \times S_b} R_0^*$  and  $\text{proj}_{S_a \times S_b} R_k^Q = \text{proj}_{S_a \times S_b} R_0^Q$  for all  $k > 0$ , implying that  $a_3$  belongs to  $S_a^*$  but not to  $S_a^Q$ .  $\triangleleft$

However, as the following theorem shows, the previous case is non-generic. Recall that a result holds *generically*, whenever for an arbitrary strategy space  $S_a \times S_b$ , the payoff vectors  $(U_a, U_b) \in \mathbb{R}^{|S_a \times S_b|} \times \mathbb{R}^{|S_a \times S_b|}$  that make this result hold are of Lebesgue measure 1 in the space  $\mathbb{R}^{|S_a \times S_b|} \times \mathbb{R}^{|S_a \times S_b|}$ . In fact, we show that the strategy profiles rationally played under common  $U$ -rational belief in rationality generically coincide to those rationally played under common belief in rationality.

**THEOREM 7.1**  $S^U = S^*$  *generically*.

Obviously, the latter implies that  $S^U$  generically contains the strategy profiles surviving iterated elimination of strictly dominated strategies.

## 8. DISCUSSION

In this section we relate the notion of rational belief hierarchies to other well-known concepts, and we discuss possible interpretations of different rational type spaces.

*Rational beliefs and unawareness.* As we have already mentioned, rational probability measures can be thought as beliefs of an agent whose language does not contain sentences of the form “ $E$  occurs with probability  $p$ ”, for any irrational number  $p \in [0, 1]$ . This idea is consistent with considering an agent who is unaware of the notion of irrational numbers and therefore does not understand sentences that express probabilistic beliefs which involve irrational numbers. Identifying the agent’s language with the collection of events that the agent is aware is not new in the literature (e.g., [Modica and Rustichini, 1999](#); [Halpern, 2001](#); [Heifetz et al., 2006](#); [Li, 2009](#)). The common characteristic of these models is that each agent is endowed with a language — subjective view of the world — and she can only reason within the bounds of this language. A straightforward consequence is that she attaches positive probability only to events generated by this language.

[Heifetz et al. \(2006\)](#) introduced a generalized framework accommodating interactive unawareness. In their model, each agent’s language restricts not only what the agent is aware of, but also what the agent believes that others are aware of. More specifically, if an agent is unaware of a sentence, she is also unaware of the possibility that someone else is aware of this sentence<sup>11</sup>. The latter is consistent

---

<sup>11</sup>[Halpern and Rêgo \(2009\)](#) go one step further by allowing agents to be aware of the possibility of being unaware of an event, thus accommodating situations where the agent deems possible that somebody else attaches positive

with the way we construct rational belief hierarchies in Section 3 in that the  $k$ -th order rational beliefs attach positive probability only to  $(k - 1)$ -th order *rational* beliefs. From the previous discussion it becomes clear why Theorem 4.1 is an intuitive contradiction. More specifically, recall that we have in mind an agent who is unaware of the concept of rational numbers. However, Theorem 4.1 says that there are types associated with a rational belief hierarchy, which at the same time “assign probability  $p \in [0, 1] \setminus \mathbb{Q}$  to some Borel event  $B$ ” in the canonical state space. Since we only consider agents who hold conscious beliefs, it follows that these specific types will know the event “ $B$  occurs with probability  $p$ ”, and therefore this last sentence will be part of their language, which in turn contradicts our initial assumption that our agent is unaware of the concept of irrational numbers.

*Bounded computational capacity.* In Section 5, we provided a characterization of universally rational belief hierarchies in terms of a bounded rationality condition which restricts the computational capacity of the players. Firstly, let us elaborate a bit on the difference between bounded computational capacity on the one hand and bounded language on the other. As we have already mentioned above, when an agent’s language is not sufficiently rich to describe certain probabilistic beliefs, the agent is not even aware of the existence of these degrees of uncertainty. Therefore, discussing about the agent’s capacity to compute these beliefs is meaningless in the first place. Conversely, an agent may understand what  $\sqrt{2}/2$  means but still find it unnecessarily complex to compute beliefs with such precision. Consider for instance an agent who simply adopts as her own probabilistic assessment the estimation yielded by a computer whose output is restricted to at most five decimals. Obviously, the agent may still understand the notion of an irrational number. From the previous discussion it becomes apparent that in our model bounded language is strictly stronger than bounded computational capacity.

There are quite a few papers in the literature that account for bounded computational capacity in the players’ belief formation. Megiddo (1989) models belief hierarchies when we restrict attention only to computable beliefs. More recently, Eliaz (2003) and Spiegler (2004) have studied repeated games with players who prefer to hold simpler beliefs about their opponents’ strategies.

*Department of Economics, Maastricht University, The Netherlands*

[e.tsakas@maastrichtuniversity.nl](mailto:e.tsakas@maastrichtuniversity.nl)

---

probability to an event of which she herself is unaware.

## APPENDIX A: PROOFS OF SECTION 2

PROOF OF PROPOSITION 2.1: Consider an arbitrary  $\mu \in \Delta^{\mathbb{Q}}(X)$ , and consider the set of singletons with positive measure,

$$(A.1) \quad \Gamma := \{x \in X : \mu(\{x\}) > 0\}.$$

First, we show that  $\Gamma$  is non-empty. Suppose that  $\mu$  is a non-atomic measure. Then, it follows from [Fremlin \(2003, p. 46\)](#) that for every  $\xi \in (0, 1)$  there is some  $B \in \mathcal{B}$  such that  $\mu(B) = \xi$ , which contradicts  $\mu \in \Delta^{\mathbb{Q}}(X)$  if we consider some  $\xi \in \mathbb{R} \setminus \mathbb{Q}$ . Hence, there is at least one atom  $A \in \mathcal{B}$ . Now, it follows from [Aliprantis and Border \(1994, Lem. 12.18\)](#) that  $A$  contains a singleton of positive measure, implying that  $\Gamma$  is non-empty.

Second, we show that  $\Gamma$  is countable. Let  $\{\Gamma_n ; n \geq 1\}$  be the countable partition of  $\Gamma$ , defined by

$$\Gamma_n := \left\{ x \in \Gamma : \frac{1}{n+1} < \mu(\{x\}) \leq \frac{1}{n} \right\}.$$

If  $\Gamma$  is uncountable, there is some  $n \geq 1$  such that  $\Gamma_n$  is uncountable, implying that there is a countably infinite  $\{x_1, x_2, \dots\} \subseteq \Gamma_n$ . Finally, observe that

$$\begin{aligned} \mu(X) &\geq \mu(\Gamma_n) \\ &\geq \sum_{k=1}^{\infty} \mu(\{x_k\}) \\ &> \sum_{k=1}^{\infty} \frac{1}{n+1} \\ &= \infty, \end{aligned}$$

which is a contradiction.

Third, we show that  $\mu(\Gamma) = 1$ . Assume otherwise, implying that  $\mu(X \setminus \Gamma) > 0$ . Since  $\Gamma$  is countable, it is Borel, implying that  $X \setminus \Gamma$  is also Borel. Hence, it follows from [Aliprantis and Border \(1994, Lem. 12.18\)](#) that there is  $x \in X \setminus \Gamma$  with  $\mu(\{x\}) > 0$ , implying, by Eq. (A.1), that  $x \in \Gamma$ , which is a contradiction.

Now, suppose that  $\Gamma = \{x_1, x_2, \dots\}$  is infinite. Observe that the sequence of rational numbers  $\{\mu(\{x_k\})\}_{k>0}$  satisfies  $\sum_{k=1}^{\infty} \mu(\{x_k\}) = 1$ . Then, it follows from [Badea \(1987, Prop., p. 225\)](#) that there is a subsequence  $\{y_k\}_{k>0}$  of  $\{x_k\}_{k>0}$  such that  $\sum_{k=1}^{\infty} \mu(\{y_k\})$  is an irrational number, thus contradicting the hypothesis that  $\mu \in \Delta^{\mathbb{Q}}(X)$ .

Therefore,  $\Gamma$  is necessarily finite. Moreover, it is closed, as it is the finite union of singletons, implying that  $\text{supp}(\mu) = \Gamma$ , which completes the proof. *Q.E.D.*

PROOF OF PROPOSITION 2.2: It follows from [Billingsley \(1968, p. 237\)](#) that the space of Borel probability measures with finite support — denoted by  $\Delta^S(X)$  and called space of simple probability measures — is dense in  $\Delta(X)$ . It follows from Proposition 2.1 that  $\Delta^{\mathbb{Q}}(X)$  is a subspace of  $\Delta^S(X)$ . Recall that denseness is transitive ([Gupta, 2000, Thm. 4.12](#)), implying that it suffices to show that  $\Delta^{\mathbb{Q}}(X)$  is dense in  $\Delta^S(X)$ .

Consider an arbitrary  $\mu \in \Delta^S(X)$ , with  $\Gamma(\mu) = \{x_1, \dots, x_n\}$ , and suppose without loss of generality that  $0 \leq \mu(\{x_1\}) \leq \dots \leq \mu(\{x_n\}) \leq 1$ . Let  $\{U_i^k\}_{k=1}^{\infty}$  be a decreasing sequence of open neighborhoods of  $\mu(\{x_i\})$  with its radius converging to 0, such that  $U_i^1 \cap U_{i+1}^1 = \emptyset$  whenever  $\mu(\{x_i\}) < \mu(\{x_{i+1}\})$ . Observe that for every  $k > 0$  and each  $i \in \{1, \dots, n\}$  there is  $\alpha_i^k \in [0, 1] \cap \mathbb{Q} \cap U_i^k$  such that  $\alpha_1^k + \dots + \alpha_n^k = 1$ , thus inducing a sequence of rational probability measures  $\{\mu_k\}$  such that

$$(\mu_k(\{x_1\}), \dots, \mu_k(\{x_n\})) \rightarrow (\mu(\{x_1\}), \dots, \mu(\{x_n\}))$$

implying that  $\mu_k \xrightarrow{w^*} \mu$  which completes the proof.

*Q.E.D.*

PROOF OF PROPOSITION 2.3: It suffices to show that an arbitrary convergent sequence  $\{\mu_k\}$  of elements of  $\Delta^N(X)$  has its limit in  $\Delta^N(X)$ , i.e., if  $\mu_k \xrightarrow{w^*} \mu$ , then  $\mu \in \Delta^N(X)$ .

Let  $\bar{N} := \max_{n \in \mathbb{N}} n$ . Let also  $d : X \times X \rightarrow \mathbb{R}$  be a metric compatible with the topology on  $X$ , and for every  $x \in X$  and  $\delta > 0$ , define an open neighborhood of  $x$  as  $B(x, \delta) := \{x' \in X : d(x, x') < \delta\}$ .

Consider an arbitrary  $x \in X$ , and suppose there is some  $\delta > 0$  such that there are finitely many  $k > 0$  with  $\mu_k(B(x, \delta)) > 0$ . Then, obviously, there are infinitely many  $k > 0$  such that  $\mu_k(B(x, \delta)) = 0$ , implying that  $\liminf \mu_k(B(x, \delta)) = 0$ . Hence, it follows from  $\mu_k \xrightarrow{w^*} \mu$  that  $\mu(B(x, \delta)) \leq \liminf \mu_k(B(x, \delta)) = 0$  (Aliprantis and Border, 1994, Thm. 15.3), implying that  $x \notin \Gamma \text{supp}(\mu)$ . If, on the other hand, for every  $\delta > 0$  there are infinitely many  $k > 0$  such that  $\mu_k(B(x, \delta)) > 0$ , it follows from  $\mu_k \in \Delta^N(X)$  that there are infinitely many  $k > 0$  such that  $\mu_k(\overline{B(x, \delta)}) \geq 1/\bar{N}$ , where  $\overline{B(x, \delta)} := \{x' \in X : d(x, x') \leq \delta\}$  is the closure of  $B(x, \delta)$ . Therefore,  $\mu(\overline{B(x, \delta)}) \geq \limsup \mu_k(\overline{B(x, \delta)}) \geq 1/\bar{N}$  (Aliprantis and Border, 1994, Thm. 15.3). Now, consider a sequence of positive reals  $\{\delta_n\}$  with  $\delta_n \downarrow 0$ , which induces a sequence of Borel events  $\{\overline{B(x, \delta_n)}\}$  such that  $\limsup_{n>0} \overline{B(x, \delta_n)} = \{x\}$ . Then, it follows from  $\mu(\limsup_{n>0} \overline{B(x, \delta_n)}) \geq \limsup_{n>0} \mu(\overline{B(x, \delta_n)}) \geq 1/\bar{N}$  (Billingsley, 1995, Thm. 4.1) that  $\mu(\{x\}) \geq 1/\bar{N}$ . Hence,  $x \in \text{supp}(\mu)$  if and only if  $\mu(\{x\}) \geq 1/\bar{N}$ , implying that  $\text{supp}(\mu)$  is finite.

Let  $x \in \text{supp}(\mu)$ . It follows from Aliprantis and Border (1994, Thm. 15.3) that for every  $\delta > 0$ ,

$$\begin{aligned}
\mu(\overline{B(x, \delta)}) &\geq \limsup \mu_k(\overline{B(x, \delta)}) \\
&\geq \limsup \mu_k(B(x, \delta)) \\
&\geq \liminf \mu_k(B(x, \delta)) \\
\text{(A.2)} \quad &\geq \mu(B(x, \delta)).
\end{aligned}$$

Since  $\text{supp}(\mu)$  is finite, there is some  $\rho > 0$  such that  $x' \notin B(x, \rho)$  for any  $x' \in \text{supp}(\mu) \setminus \{x\}$ , implying that  $\mu(\overline{B(x, \delta)}) = \mu(B(x, \delta)) = \mu(\{x\})$  for every  $\delta < \rho$ . Hence, it follows from (A.2) that  $\mu(\{x\}) = \lim \mu_k(B(x, \delta))$ . Finally, since the sequence  $\{\mu_k(B(x, \delta))\}$  contains only elements of the finite set  $\mathbb{Q}_N$ , it follows that  $\lim \mu_k(B(x, \delta)) \in \mathbb{Q}_N$ , which completes the proof.

*Q.E.D.*

PROOF OF PROPOSITION 2.4: First, observe that  $\Delta^{\mathbb{Q}}(X) = \bigcup_{n \in \mathbb{N}} \Delta^{\{1, \dots, n\}}(X)$ . It follows from Proposition 2.3 that  $\Delta^{\{1, \dots, n\}}(X)$  is closed, and therefore Borel in  $\Delta(X)$  for every  $n \in \mathbb{N}$ , which completes the proof.

*Q.E.D.*

## APPENDIX B: PROOFS OF SECTION 3

PROOF OF PROPOSITION 3.1: We proceed inductively to show that for every  $k \geq 0$

- ⊗  $\Theta_k$  is embedded as a Borel subset of  $\Psi_k$ , and
- ⊗  $\Delta^{\mathbb{Q}}(\Theta_k)$  is embedded as a Borel subset of  $\Delta(\Psi_k)$ .

First, observe that  $\Theta_0 = \Psi_0 = \Theta$ , implying that  $\Theta_0$  is embedded as a Borel subset of  $\Psi_0$  via the identity function. It follows from Proposition 2.4 that  $\Delta^{\mathbb{Q}}(\Theta_0)$  is a Borel subset of  $\Delta(\Psi_0)$ . Hence,  $\Delta^{\mathbb{Q}}(\Theta_0)$  is embedded as a Borel subset of  $\Delta(\Psi_0)$  via the identity function.

Now, suppose that  $\Theta_k$  is embedded as a Borel subset of  $\Psi_k$  via  $\vartheta_k : \Theta_k \rightarrow \Psi_k$ , and  $\Delta^{\mathbb{Q}}(\Theta_k)$  is embedded as a Borel subset of  $\Delta(\Psi_k)$  via  $\delta_k : \Delta^{\mathbb{Q}}(\Theta_k) \rightarrow \Delta(\Psi_k)$ . Define the function  $\vartheta_{k+1} : \Theta_{k+1} \rightarrow \Psi_{k+1}$  such that for each  $(\theta_k, \mu_k) \in \Theta_k \times \Delta^{\mathbb{Q}}(\Theta_k)$ ,  $\vartheta_{k+1}(\theta_k, \mu_k) := (\vartheta_k(\theta_k), \delta_k(\mu_k))$ . Obviously, it follows from above that  $\Theta_{k+1}$  is embedded as

a Borel subset of  $\Psi_{k+1}$  via  $\vartheta_{k+1}$ , and therefore  $\Theta_{k+1}$  is homeomorphic to  $\vartheta_{k+1}(\Theta_{k+1})$ . Hence,  $\Delta(\Theta_{k+1})$  is homeomorphic to  $\Delta(\vartheta_{k+1}(\Theta_{k+1}))$ . Since  $\vartheta_{k+1}(\Theta_{k+1})$  is Borel in  $\Psi_{k+1}$ , there is a homeomorphism  $\Delta(\vartheta_{k+1}(\Theta_{k+1})) \mapsto \{\mu \in \Delta(\Psi_{k+1}) : \mu(\vartheta_{k+1}(\Theta_{k+1})) = 1\}$ , where  $\{\mu \in \Delta(\Psi_{k+1}) : \mu(\vartheta_{k+1}(\Theta_{k+1})) = 1\}$  is a Borel subset of  $\Delta(\Psi_{k+1})$ . The latter implies that there is a homeomorphism  $\nu_{k+1} : \Delta(\Theta_{k+1}) \rightarrow \{\mu \in \Delta(\Psi_{k+1}) : \mu(\vartheta_{k+1}(\Theta_{k+1})) = 1\}$ . By Proposition 2.4,  $\Delta^{\mathbb{Q}}(\Theta_{k+1})$  is Borel in  $\Delta(\Theta_{k+1})$ . Let  $\delta_{k+1} : \Delta^{\mathbb{Q}}(\Theta_{k+1}) \rightarrow \{\mu \in \Delta(\Psi_{k+1}) : \mu(\vartheta_{k+1}(\Theta_{k+1})) = 1\}$  be the same mapping as  $\nu_{k+1}$  but restricted in the domain  $\Delta^{\mathbb{Q}}(\Theta_{k+1})$ . Then, it follows directly that  $\delta_{k+1}$  embeds  $\Delta^{\mathbb{Q}}(\Theta_{k+1})$  as a Borel subset on  $\{\mu \in \Delta(\Psi_{k+1}) : \mu(\vartheta_{k+1}(\Theta_{k+1})) = 1\}$  and therefore on  $\Delta(\Psi_{k+1})$ , which completes the proof by induction. *Q.E.D.*

## APPENDIX C: PROOFS OF SECTION 4

PROOF OF PROPOSITION 4.1: Since  $\Theta_0 = \Theta$  is Polish, it is by definition separable and metrizable. Supposing that  $\Theta_k$  is separable and metrizable,  $\Delta^{\mathbb{Q}}(\Theta_k)$  is separable and metrizable too, as it is a subspace of  $\Delta(\Theta_k)$ , which is also separable and metrizable (Aliprantis and Border, 1994, Thm 15.12). Thus,  $\Theta_{k+1} = \Theta_k \times \Delta^{\mathbb{Q}}(\Theta_k)$  is also separable and metrizable, and therefore it follows by induction that every  $\Theta_k$  is separable and metrizable. Since  $\pi_{k+1} \in \Delta^{\mathbb{Q}}(\Theta_k)$  has a finite support (by Proposition 2.1), it follows that it is tight, and therefore, by applying a version of Kolmogorov extension theorem (Aliprantis and Border, 1994, Cor. 15.28), we prove the existence of a unique measure  $\pi \in \Delta(\Theta \times T_0^{\mathbb{Q}})$  that extends every  $\pi_k$ . *Q.E.D.*

PROOF OF PROPOSITION 4.2: The proof follows directly from  $T^{\mathbb{Q}} = \{t \in T_c^{\mathbb{Q}} : f(t)(\Theta \times T^{\mathbb{Q}}) = 1\}$ . *Q.E.D.*

PROOF OF THEOREM 4.1: Take two arbitrary  $\theta_1, \theta_2 \in \Theta$ , and consider the following sequence:

$$\begin{aligned} \mathcal{P}_1 &:= \{ p_1 \in \Delta^{\mathbb{Q}}(\Theta_0) : p_1(\theta) = 1, \text{ for some } \theta \in \{\theta_1, \theta_2\} \} \\ \mathcal{P}_2 &:= \{ p_2 \in \Delta^{\mathbb{Q}}(\Theta_1) : p_2(\theta, p_1) = 1, \text{ for some } (\theta, p_1) \in \{\theta_1, \theta_2\} \times \mathcal{P}_1 \} \\ &\vdots \\ \mathcal{P}_k &:= \{ p_k \in \Delta^{\mathbb{Q}}(\Theta_{k-1}) : p_k(\theta, p_1, \dots, p_{k-1}) = 1, \text{ for some } (\theta, p_1, \dots, p_{k-1}) \in \{\theta_1, \theta_2\} \times \mathcal{P}_1 \times \dots \times \mathcal{P}_{k-1} \} \\ &\vdots \end{aligned}$$

Let  $T_p$  be the set of types  $(p_1, p_2, \dots) \in \prod_{k>0} \mathcal{P}_k$  that satisfy coherency and common certainty in coherency. Observe that for every  $p_1 \in \mathcal{P}_1$  there are exactly two measures in  $\mathcal{P}_2$  such that  $(p_1, p_2)$  does not contradict coherency. Likewise, for every  $(p_1, p_2) \in \mathcal{P}_1 \times \mathcal{P}_2$  that does not contradict coherency, there are exactly two measures  $p_3 \in \mathcal{P}_3$  such that  $(p_1, p_2, p_3)$  does not contradict coherency and 1-fold certainty in coherency. Inductively, for each  $k > 1$ , for every  $(p_1, \dots, p_{k-1}) \in \mathcal{P}_1 \times \dots \times \mathcal{P}_{k-1}$  that does not contradict coherency, 1-fold,  $\dots$ , and  $(k-3)$ -fold certainty in coherency, there are exactly two measures  $p_k \in \mathcal{P}_k$  such that  $(p_1, \dots, p_k)$  does not contradict coherency, 1-fold,  $\dots$ , and  $(k-2)$ -fold certainty in coherency. Therefore,  $T_p$  has the same cardinality as  $\{0, 1\}^{\mathbb{N}}$ , implying that it is uncountable. Now, consider a belief hierarchy  $(\pi_1, \pi_2, \dots)$  such that

$$\begin{aligned} \pi_1 &\text{ is uniformly distributed over } \Theta, \\ \pi_2 &\text{ is uniformly distributed over } \Theta \times \text{proj}_{\mathcal{P}_1} T_p, \\ &\vdots \\ \pi_k &\text{ is uniformly distributed over } \Theta \times \text{proj}_{\mathcal{P}_1 \times \dots \times \mathcal{P}_{k-1}} T_p \end{aligned}$$

⋮

First observe that  $(\pi_1, \pi_2, \dots)$  satisfies coherency and common certainty in coherency. Moreover, by construction  $(\pi_1, \pi_2, \dots) \in T^Q$ , and therefore  $g(\pi_1, \pi_2, \dots) \in \Delta(\Theta \times T^Q)$ . However, observe that  $g(\pi_1, \pi_2, \dots)$  has an infinite support, and therefore by Proposition 2.1,  $g(\pi_1, \pi_2, \dots) \notin \Delta^Q(\Theta \times T^Q)$ , which completes the proof. *Q.E.D.*

#### APPENDIX D: PROOFS OF SECTION 5

**PROOF OF PROPOSITION 5.1:** The proof is rather straightforward. Consider an arbitrary  $t \in T^U$ . It follows from  $t \in T_0^U$  that  $g(t) \in \Delta^Q(\Theta \times T^Q)$ . Now, suppose that there is some  $t' \in T^Q \setminus T^U$  such that  $g(t)(\Theta \times \{t'\}) > 0$ . Since  $t' \notin T^U$ , it follows that there is some  $k \geq 0$  such that  $t' \notin T_k^U$ , implying that  $t \notin T_{k+1}^U$ , which contradicts  $t \in T^U$ . Therefore,  $g(T^U) \subseteq \Delta^Q(\Theta \times T^U)$ , implying, together with Proposition 4.2, that  $g : T^U \rightarrow \Delta^Q(\Theta \times T^U)$  is an injection. Now, take an arbitrary  $\mu \in \Delta^Q(\Theta \times T^U)$ . It follows by construction that  $\mu$  is associated with a  $U$ -rational belief hierarchy, implying that  $g : T^U \rightarrow \Delta^Q(\Theta \times T^U)$  is surjective. Finally, continuity follows directly from [Brandenburger and Dekel \(1993, Prop. 2\)](#). *Q.E.D.*

**PROOF OF PROPOSITION 5.2:** First, we show that  $g(T^N) \subseteq \Delta^N(\Theta \times T^Q)$ . Consider an arbitrary  $(\pi_1, \pi_2, \dots) \in T^N$ . It follows from Proposition 4.2 that there is some  $\pi \in \Delta(\Theta \times T^Q)$  such that  $g(\pi_1, \pi_2, \dots) = \pi$ . It suffices to show that  $\pi \in \Delta^N(\Theta \times T^Q)$ . For each  $k \geq 0$ , let  $\mathcal{B}_k$  denote the Borel  $\sigma$ -algebra in  $\Theta_k$ . Since  $\pi$  extends every  $\pi_{k+1}$ , it follows that for every  $B_k \in \mathcal{B}_k$ ,

$$(D.1) \quad \pi_{k+1}(B_k) = \pi \left( B_k \times \prod_{\ell=k}^{\infty} \Delta^Q(\Theta_\ell) \right).$$

Observe that every Borel event  $B \subseteq \Theta \times T^Q$  is also Borel in  $\Theta \times \prod_{k=0}^{\infty} \Delta^Q(\Theta_k)$ , and

$$(D.2) \quad B = \bigcap_{k=0}^{\infty} \left( \text{proj}_{\Theta_k} B \times \prod_{\ell=k}^{\infty} \Delta^Q(\Theta_\ell) \right).$$

Then, it follows from [Billingsley \(1995, Thm. 4.1\)](#), together with Eq. (D.2), that

$$\begin{aligned} \pi(B) &= \lim_{k \rightarrow \infty} \pi \left( \text{proj}_{\Theta_k} B \times \prod_{\ell=k}^{\infty} \Delta^Q(\Theta_\ell) \right) \\ &\stackrel{(D.1)}{=} \lim_{k \rightarrow \infty} \pi_{k+1}(\text{proj}_{\Theta_k} B). \end{aligned}$$

Since  $(\text{proj}_{\Theta_k} B) \in \mathcal{B}_k$  and  $\pi_{k+1} \in \Delta^N(\Theta_k)$ , it follows that  $\pi_{k+1}(\text{proj}_{\Theta_k} B) \in \mathbb{Q}_N$ . Since  $\{\pi_{k+1}(\text{proj}_{\Theta_k} B)\}_{k>0}$  is a convergent sequence taking finitely many values, it follows that the limit converges to one of these values. Therefore,  $\pi(B) \in \mathbb{Q}_N$ , which proves that  $\pi \in \Delta^N(\Theta \times T^Q)$ .

Showing that  $\Delta^N(\Theta \times T^Q) \subseteq g(T^N)$  is straightforward, and it follows directly from  $\Delta^N(\Theta \times T^Q) \subseteq \Delta^Q(\Theta \times T^Q)$ . Thus, we conclude that  $\Delta^N(\Theta \times T^Q) = g(T^N)$ , which together with Proposition 4.2 completes the proof. *Q.E.D.*

**PROOF OF THEOREM 5.1:** First, we show that  $T_0^U = T_0^N$ . Consider some  $t \in T_0^U$ , implying that  $g(t) \in \Delta^Q(\Theta \times T^Q)$ . By Proposition 2.1,  $g(t)$  has a finite support, implying that there is some  $n > 0$  such that  $g(t) \in \Delta^{\{1, \dots, n\}}(\Theta \times T^Q)$ . Hence,  $t \in T_0^N$ , which proves that  $T_0^U \subseteq T_0^N$ . Now, consider some  $t \in T_0^N$ , implying that  $t \in T^N$  for some  $N \in \mathbb{N}$ . Therefore, by Proposition 5.2,  $g(t) \in \Delta^N(\Theta \times T^Q)$ , and therefore  $g(t) \in \Delta^Q(\Theta \times T^Q)$ , implying that  $t \in T_0^U$ , thus proving  $T_0^N \subseteq T_0^U$ . Hence,  $T_0^N = T_0^U$ .

Now, for an arbitrary  $k \geq 0$  suppose that for all  $\ell \leq k$ ,

$$(D.3) \quad T_0^U \cap \cdots \cap T_\ell^U = T_0^{\mathcal{N}} \cap \cdots \cap T_\ell^{\mathcal{N}}.$$

Then, we show that  $T_0^U \cap \cdots \cap T_{k+1}^U = T_0^{\mathcal{N}} \cap \cdots \cap T_{k+1}^{\mathcal{N}}$ :

$$\begin{aligned} T_0^U \cap \cdots \cap T_{k+1}^U &= T_0^U \cap \left( \bigcap_{\ell=0}^k \{ t \in T^Q : g(t)(\Theta \times T_\ell^U) = 1 \} \right) \\ &= T_0^U \cap \{ t \in T^Q : g(t)(\Theta \times (T_0^U \cap \cdots \cap T_k^U)) = 1 \} \\ &= \{ t \in T_0^U : g(t)(\Theta \times (T_0^U \cap \cdots \cap T_k^U)) = 1 \} \\ &\stackrel{(D.3)}{=} \{ t \in T_0^{\mathcal{N}} : g(t)(\Theta \times (T_0^{\mathcal{N}} \cap \cdots \cap T_k^{\mathcal{N}})) = 1 \} \\ &= \bigcap_{\ell=0}^k \{ t \in T_0^{\mathcal{N}} : g(t)(\Theta \times T_\ell^{\mathcal{N}}) = 1 \} \\ &= T_0^{\mathcal{N}} \cap \left( \bigcap_{\ell=0}^k \{ t \in T_0^{\mathcal{N}} : g(t)(\Theta \times T_\ell^{\mathcal{N}}) = 1 \} \right) \\ &= T_0^{\mathcal{N}} \cap \cdots \cap T_{k+1}^{\mathcal{N}}. \end{aligned}$$

Thus, it follows by induction that  $T_0^U \cap \cdots \cap T_k^U = T_0^{\mathcal{N}} \cap \cdots \cap T_k^{\mathcal{N}}$  for all  $k \geq 0$ , implying that  $T^U = T^{\mathcal{N}}$ . *Q.E.D.*

## APPENDIX E: PROOFS OF SECTION 6

LEMMA E.1  $T^Q$  is dense in  $T^*$ .

PROOF: Consider an arbitrary  $t^* \in T^*$ , and we show that there is a sequence of rational types converging to  $t^*$  in the topology of weak convergence.

It follows from Proposition 2.2 that for every  $t \in T^*$  there is a sequence of rational probability measures on the space  $\Theta \times T^*$  that converges to  $g(t)$ , i.e., there is a sequence  $\{\mu_k^t\}_{k>0}$  with  $\mu_k^t \in \Delta^Q(\Theta \times T^*)$ , such that  $\mu_k^t \xrightarrow{w^*} g(t)$ . Now, let  $T_i$  be a copy of  $T^*$ , inducing a homeomorphism  $r_i : T_i \rightarrow T^*$ . The elements of  $T_i$  are not yet associated with any belief hierarchy. For each  $k > 0$ , define the type space  $(\Theta, T_a, T_b, g_a^k, g_b^k)$ , such that for each  $t_i \in T_i$ , and every  $(\theta, t_j) \in \Theta \times T_j$ ,

$$(E.1) \quad g_i^k(t_i)(\theta, t_j) = \mu_k^{r_i(t_i)}(\theta, r_j(t_j)).$$

Observe that, by construction, for each  $k > 0$ , every  $t_i$  is associated with a rational belief hierarchy. The latter induces a mapping

$$\beta_i^k : T_i \rightarrow T^Q$$

that associates every  $t_i$  to a sequence of rational types,  $\{\beta_i^k(t_i)\}_{k>0}$ . In other words, in the type space  $(\Theta, T_a, T_b, g_a^k, g_b^k)$ , the type  $t_i$  yields the same rational hierarchy as  $\beta_i^k(t_i)$  does in  $(\Theta, T_a^Q, T_b^Q, g_a, g_b)$ . Note that for different  $k$ 's we may obtain different hierarchies associated with  $t_i$ . Finally, it suffices to show that

$$(E.2) \quad \beta_i^k(r_i^{-1}(t^*)) \xrightarrow{w^*} t^*.$$

Observe that, for every  $k > 0$  and each  $t_i \in T_i$ , the first order beliefs associated with  $\beta_i^k(t_i)$  are given by

$$\begin{aligned} \pi_1(\beta_i^k(t_i)) &= \text{marg}_{\Theta} g(\beta_i^k(t_i)) \\ &= \text{marg}_{\Theta} g_i^k(t_i) \\ \stackrel{\text{(E.1)}}{=} & \text{marg}_{\Theta} \mu_k^{r_i(t_i)}. \end{aligned}$$

Since,  $\mu_k^{r_i(t_i)} \xrightarrow{w^*} g(r_i(t_i))$ , it follows that  $\{\pi_1(\beta_i^k(t_i))\}_{k>0}$  weakly converges to the first order beliefs associated with  $r_i(t_i)$ . Since the latter is true for each  $i \in \{a, b\}$  and every  $t_i \in T_i$ , it follows that for every  $n > 0$  the sequence of  $n$ -th order rational beliefs  $\{\pi_n(\beta_i^k(t_i))\}_{k>0}$  weakly converges to the  $n$ -th order beliefs associated with  $r_i(t_i)$ . Therefore, the sequence of rational types  $\{\beta_i^k(t_i)\}_{k>0}$  weakly converges to  $r_i(t_i)$ , i.e.,  $\beta_i^k(t_i) \xrightarrow{w^*} r_i(t_i)$ . Finally, set  $t_i = r_i^{-1}(t^*)$ , and obtain Eq. (E.2), which completes the proof. *Q.E.D.*

LEMMA E.2  $T^U$  is dense in  $T^Q$ .

PROOF: Let  $(\pi_1, \pi_2, \dots) \in T^Q$ . Then, it suffices to show that there is a sequence of  $\mathcal{N}$ -rational — and therefore by Theorem 5.1,  $U$ -rational — types  $\{t^\ell\}_{\ell=1}^\infty$ , such that  $\pi_k(t^\ell) \xrightarrow{w^*} \pi_k$  for all  $k > 0$ . Now, for each  $\ell > 0$ , define  $t^\ell := (\pi_1(t^\ell), \pi_2(t^\ell), \dots)$  by

- ⊗  $\pi_k(t^\ell) := \pi_k$  for all  $k < \ell$ , and
- ⊗  $\pi_k(t^\ell)$  is such that  $\text{marg}_{\Theta} \pi_k(t^\ell) = \pi_1(t^\ell)$  and also  $\pi_k(t^\ell)(\Theta \times \{\pi_1(t^\ell)\} \times \dots \times \{\pi_{k-1}(t^\ell)\}) = 1$  for all  $k \geq \ell$ .

First, observe that by construction, for every  $\ell > 0$ , the type  $t^\ell$  is  $\mathcal{N}$ -rational. Now, consider an arbitrary  $k > 0$ . Then, it suffices, for every  $\varepsilon > 0$ , there is some  $\ell_0$  such that for every bounded and continuous real-valued function  $f$ ,

$$\left| \int f d\pi_k(t^\ell) - \int f d\pi_k \right| < \varepsilon$$

for all  $\ell > \ell_0$ . Set  $\ell_0 = k$  and the proof is completed. *Q.E.D.*

PROOF OF PROPOSITION 6.1: Recall that denseness is transitive. Then, the proof follows directly from Lemmas E.1 and E.2. *Q.E.D.*

## APPENDIX F: PROOFS OF SECTION 7

PROOF OF THEOREM 7.1: We show that generically each pure strategy is rationally played under some conjecture if and only if it is rationally played under some rational conjecture. Then, since this is a finite game, for each rational conjecture there is some  $U$ -rational belief hierarchy inducing this conjecture. First, we simplify the notation we use throughout the proof: Let  $S_i = (s_i^1, \dots, s_i^K)$  and  $S_j = (s_j^1, \dots, s_j^L)$ . For each  $s_i \in S_i$ , let  $U_i(s_i) := (U_i(s_i, s_j^1), \dots, U_i(s_i, s_j^L))$ . Fix some arbitrary  $U_j \in \mathbb{R}^{|S_i \times S_j|}$  and  $(U_i(s_i^1), \dots, U_i(s_i^{k-1}), U_i(s_i^{k+1}), \dots, U_i(s_i^K)) \in \mathbb{R}^{|(S_i^{-1}) \times S_j|}$ . Then, we show that the set

$$\text{(F.1)} \quad Z := \{ U_i(s_i^k) \in \mathbb{R}^{|S_j|} : s_i^k \in (\text{proj}_{S_i} R_0^*) \setminus (\text{proj}_{S_i} R_0^U) \}$$

is of Lebesgue measure 0 in  $\mathbb{R}^{|S_j|}$ . Define  $\lambda : \Delta(S_j) \rightarrow \mathbb{R}$  by

$$\text{(F.2)} \quad \lambda(\mu_i) := \max_{s'_i \in S_i \setminus \{s_i^k\}} u_i(s'_i, \mu_i),$$

which is obviously continuous and convex in  $\Delta(S_j)$ . Now, suppose that  $(U_i(s_i^k, s_j^1), \dots, U_i(s_i^k, s_j^L)) \in Z$ . This implies that

$$(F.3) \quad u_i(s_i^k, \mu_i) \geq \lambda(\mu_i)$$

only holds with equality. Finally, observe that for every  $\varepsilon > 0$ ,

- ⊠  $(U_i(s_i^k, s_j^1) + \varepsilon, \dots, U_i(s_i^k, s_j^L) + \varepsilon) \notin Z$ . This is because in this case, (F.3) holds with strict inequality for the irrational conjecture  $\mu_i \in \Delta(S_j)$  that solved it before. Therefore, by Proposition 2.2 there exists some  $\mu'_i \in \Delta^{\mathbb{Q}}(S_j)$  which would be sufficiently close to  $\mu_i$  and therefore by continuity it also satisfies (F.3).
- ⊠  $(U_i(s_i^k, s_j^1) - \varepsilon, \dots, U_i(s_i^k, s_j^L) - \varepsilon) \notin Z$ . The reason is that in this case, (F.3) does not hold for any conjecture, rational or irrational.

Repeat the same argument for all  $U_j \in \mathbb{R}^{|S_i \times S_j|}$  and  $(U_i(s_i^1), \dots, U_i(s_i^{k-1}), U_i(s_i^{k+1}), \dots, U_i(s_i^K)) \in \mathbb{R}^{|(S_i-1) \times S_j|}$  and the proof is completed. Q.E.D.

## REFERENCES

- AHN, D.S. (2007). Hierarchies of ambiguous beliefs. *Journal of Economic Theory* 136, 286–301.
- ALIPRANTIS, C. & BORDER, K. (1994). *Infinite dimensional analysis*. Springer Verlag, Berlin.
- AUMANN, R.J. (1976). Agreeing to disagree. *Annals of Statistics* 4, 1236–1239.
- (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55, 1–18.
- AUMANN, R.J. & BRANDENBURGER, A. (1995). Epistemic conditions for Nash equilibrium. *Econometrica* 63, 1161–1180.
- BADEA, C. (1987). Irrationality of certain infinity series. *Glasgow Mathematical Journal* 29, 221–228.
- BATTIGALLI, P. & FRIEDENBERG, A. (2012). Forward induction reasoning revisited. *Theoretical Economics* 7, 57–98.
- BATTIGALLI, P. & SINISCALCHI, M. (2003). Rationalization and incomplete information. *Advances in Theoretical Economics* 3, 1–44.
- BERNHEIM, D. (1984). Rationalizable strategic behavior. *Econometrica* 52, 1007–1028.
- BILLINGSLEY, P. (1968). *Convergence of probability measures*. John Wiley & Sons, New York.
- (1995). *Probability and measure*. John Wiley & Sons, New York.
- BRANDENBURGER, A. (2008). Epistemic game theory: complete information. *The New Palgrave Dictionary of Economics*, edited by L. Blume & S. Durlauf, McMillan, forthcoming.
- BRANDENBURGER, A. & DEKEL, E. (1987). Rationalizability and correlated equilibria. *Econometrica* 55, 1391–1402.
- (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory* 59, 189–198.
- DEKEL, E., FUDENBERG, D. & MORRIS, S. (2006). Topologies on types. *Theoretical Economics* 1, 275–309.
- (2007). Interim correlated rationalizability. *Theoretical Economics* 2, 15–40.
- DEKEL, E., LIPMAN, B. & RUSTICHINI, A. (1998). Standard state-space models preclude unawareness. *Econometrica* 66, 159–173.
- ELIAZ, K. (2003). Nash equilibrium when players account for the complexity of their forecasts. *Games and Economic Behavior* 44, 286–310.
- ELY, J.C. & PEŒSKI, M. (2006). Hierarchies of belief and interim correlated rationalizability. *Theoretical Economics* 1, 19–65.
- FREMLIN, D.H. (2003). *Measure theory: Broad foundations*. Vol. 2, Torres Fremlin, Essex.

- FRIEDENBERG, A. & KEISLER, H.J. (2011). Iterated dominance revisited. *Working Paper, Arizona State University*.
- GUPTA, A. (2000). *Real and abstract analysis*. Academic Publishers, Calcutta.
- HALPERN, J.Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior* 37, 321–339.
- HALPERN, J.Y. & RÊGO, L.C. (2009). Reasoning about knowledge of unawareness. *Games and Economic Behavior* 67, 503–525.
- HARSANYI, J. (1967-68). Games with incomplete information played by Bayesian players, I-III. *Management Science* 14, 159–182, 320–334, 486–502.
- HEIFETZ, A. (1993). The Bayesian formulation of incomplete information — The non-compact case. *International Journal of Game Theory* 21, 329–338
- HEIFETZ, A. & KETS, W. (2011). All types naive and canny. *Working Paper, Northwestern University*.
- HEIFETZ, A., MEIER, M. & SCHIPPER, B. (2006). Interactive unawareness. *Journal of Economic Theory* 130, 78–94.
- HEIFETZ, A. & SAMET (1998). Topology-free typology of beliefs. *Journal of Economic Theory* 82, 324–341.
- KECHRIS, A. (1995). *Classical descriptive set theory*. Springer Verlag, Berlin.
- KETS, W. (2010). Bounded reasoning and higher order uncertainty. *Working Paper, Santa Fe Institute*.
- LEHTINEN, E., MERENLUOTO, K. & KASANEN, E. (1997). Conceptual change in mathematics: From rational to (un)real numbers. *European Journal of Psychology of Education* 12, 131–145.
- LI, J. (2009). Information structures with unawareness. *Journal of Economic Theory* 144, 977–993.
- LIPMAN, B.L. (2003). Finite order implications of common priors. *Econometrica* 71, 1255–1267.
- MEGIDDO, N. (1989). On computable beliefs of rational machines. *Games and Economic Behavior* 1, 144–169.
- MEIER, M. (2006). Finitely additive beliefs and universal type spaces. *Annals of Probability* 34, 386–422.
- MERTENS, J.F. & ZAMIR, S. (1985). Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory* 14, 1–29.
- MODICA, S. & RUSTICHINI, A. (1999). Unawareness and partitioned information structures. *Games and Economic Behavior* 27, 265–298.
- PARTHASARATHY, K.R. (1967). *Probability measures on metric spaces*. AMS Chelsea Publishing, Providence, Rhode Island.
- PEARCE, D.G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52, 1029–1050.
- PEREA, A. (2012). *Reasoning and choice: An epistemic course in game theory*. Cambridge University Press. UK.
- SINISCALCHI, M. (2007). Epistemic game theory: Beliefs and types. *The New Palgrave Dictionary of Economics*, edited by L. Blume & S. Durlauf, McMillan, forthcoming.
- SPIEGLER, R. (2004). Simplicity of beliefs and delay tactics in a concession game. *Games and Economic Behavior* 47, 200–220.
- SRIVASTAVA, S.M. (1991). *A course on Borel sets*. Springer Verlag, Berlin.
- STRZALECKI, T. (2011). Depth of reasoning and higher order beliefs. *Working Paper, Harvard University*
- TAN, T. & WERLANG, S. (1988). The Bayesian foundations of solution concepts of games. *Journal of Economic Theory* 45, 370–391.