

Heifetz, Aviad; Kets, Willemien

**Working Paper**

## All types naive and canny

Discussion Paper, No. 1550

**Provided in Cooperation with:**

Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

*Suggested Citation:* Heifetz, Aviad; Kets, Willemien (2012) : All types naive and canny, Discussion Paper, No. 1550, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/97243>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



CMS-EMS  
Center for Mathematical Studies in Economics  
And Management Science

Discussion Paper #1550

# All Types Naive and Canny

Willemien Kets  
Northwestern University

Aviad Heifetz  
Open University of Israel

August 4, 2012

*JEL Classification:* C700, C720, D800, D830

*Keywords:* Level-k models, cognitive hierarchy models, universal type space, global games



NORTHWESTERN  
UNIVERSITY



# All Types Naive and Canny

Aviad Heifetz\*

Willemien Kets<sup>†</sup>

First version: November 27, 2011

This version: August 4, 2012

## Abstract

This paper constructs a type space that contains all types with a finite depth of reasoning, as well as all types with an infinite depth of reasoning – in particular those types for whom finite-depth types are conceivable, or think that finite-depth types are conceivable in the mind of other players, etcetera. We prove that this type space is universal with respect to the class of type spaces that include types with a finite or infinite depth of reasoning. In particular, we show that it contains the standard universal type space of Mertens and Zamir (1985) as a belief-closed subspace, and that this subspace is characterized by common belief of infinite-depth reasoning. This framework allows us to study the robustness of classical results to small deviations from perfect rationality. As an example, we demonstrate that in the global games of Carlsson and van Damme (1993), a small ‘grain of naiveté’ suffices to overturn the classical uniqueness results in that literature.

---

\*Department of Management and Economics, Open University of Israel. E-mail: aviadhe@openu.ac.il.  
Phone: +972-9-778-1878.

<sup>†</sup>Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu.  
Phone: +1-505-204 8012.

# 1 Introduction

Ever since the conceptual breakthrough of Harsanyi (1967–1968), type spaces are customarily used for analyzing economic applications with incomplete information. Each type has preferences and beliefs about the state of nature and others’ types, and hence, implicitly, also about the others’ beliefs about nature, about the others’ beliefs about *their* peers’ beliefs about nature, and so on, *ad infinitum*. Mertens and Zamir (1985) showed that the space of all such belief hierarchies can be used to construct a type space which is *universal* in the sense that any other type space can be mapped into it using a unique belief-preserving mapping.<sup>1</sup>

In recent years, however, this seemingly all-embracing approach to modeling incomplete information has been challenged by an important literature in experimental economics and behavioral game theory. People do not think *ad infinitum*, so goes the claim. Often they barely think at all. Somewhat less often they think about what others think. Even more rarely do they contemplate the thoughts of others about somebody else, and seldom, if at all, do they form an infinite hierarchy of mutual beliefs in their mind. Indeed, assuming that subjects can be characterized by *cognitive types* with a finite depth of reasoning is consistent with experimental data on beauty contests, certain auctions, and other classes of games.<sup>2</sup> Strzalecki (2009) showed how to construct a universal type space for cognitive types with a finite depth.

The Mertens-Zamir universal space for Harsanyi types and the Strzalecki universal space for finite-depth types are completely disjoint. This might suggest that the two approaches to incomplete information are conceptually divorced. But are they really diametrically opposed? Consider, for instance, comments of some players in beauty-contest experiments (Bosch-Domenech et al., 2002, Appendix A), who are aware of the fact that infinite depth of reasoning would entail the dominance-solvable outcome of the game, but at the same time acknowledge that ‘many different people participate in this game and not everybody will apply [this] reasoning process’. Such players are, on one hand, completely sophisticated in the sense of Harsanyi, but at the same time they admit that others might not be, and act upon this belief.

In this paper we show how to account for such types of players, as well as many others, such as a type who believes that everybody else is fully sophisticated like she is, but who also believes that others may yet entertain suspicions that *their* rivals are not; and so forth. We marry the two seemingly disparate approaches by constructing one universal space that

---

<sup>1</sup>This is true under suitable topological assumptions; also see Brandenburger and Dekel (1993), Heifetz (1993), and Mertens et al. (1994).

<sup>2</sup>See, e.g., Nagel (1995), Stahl and Wilson (1995), Ho et al. (1998), Costa-Gomes et al. (2001), Camerer et al. (2004), and Crawford and Iriberri (2007). See Crawford et al. (2012) for a survey.

contains all such types. In particular, it includes both the universal space of Mertens and Zamir (1985) and Strzalecki (2009) as belief-closed subspaces. This demonstrates how the route recently taken in the behavioral economics literature may be viewed as part and parcel of the generalization proposed here to the classical Harsanyi framework, not a conceptual antithesis to it.<sup>3</sup>

This general framework allows to analyze the robustness of classical results in the literature on games with incomplete information to the presence of various kinds of cognitive types, or even just to the presence of mutual suspicions by fully sophisticated types about the presence of cognitively limited types. We focus on global games, introduced by Carlsson and van Damme (1993). This class of games is of particular interest, because the assumption that players can form beliefs at arbitrarily high orders has stark implications in this class of games.<sup>4</sup> We show that the slightest ‘grain of naiveté’ suffices to overturn the classical result that there is a unique rationalizable action in these games. This result is stronger than a similar result of Strzalecki (2009), who shows that there can be multiple equilibria in the electronic mail game of Rubinstein (1989) when players have a finite depth of reasoning: it suffices that types only think that others think...that other players may have a finite depth of reasoning for both actions of the coordination game to survive as rationalizable. Our framework, which allows for such infinite-depth types with a grain of naiveté is thus essential for formulating and proving this result.

The framework we propose here complements that developed by Kets (2010, 2012), who introduces type spaces in which players can have a finite depth of reasoning. Unlike in our model, players can have nontrivial higher-order beliefs in her framework, even if they have a finite depth of reasoning. Intuitively, types with a finite depth of reasoning can have beliefs about events of higher order if these higher-order events are generated by events of sufficiently low order. In particular, players with a finite depth of reasoning can have common knowledge of nontrivial events in the framework of Kets, something which is not possible in the framework presented here. The framework we present here is thus less general, but has the advantage that it is closer to the models developed in the experimental literature and that it is easy

---

<sup>3</sup>A number of authors have constructed universal type spaces for situations where players do not have unlimited reasoning abilities. In addition to Strzalecki (2009), who constructs a universal type space for type spaces in which all types have a finite depth of reasoning, Heinsalu (2011) and Pintér and Udvari (2011) construct universal type spaces for types with a finite and infinite depth of reasoning that also allow for various forms of unawareness. The present framework encompasses precisely the cognitive types models developed in the experimental literature, and has the advantage that it separates the issue of bounded reasoning from the conceptually distinct issue of unawareness. Our results do not follow from nor entail those in Heinsalu and Pintér and Udvari.

<sup>4</sup>See Heinemann et al. (2004), Heinemann et al. (2009), Cabrales et al. (2007), among others, for experimental studies of games with strategic complementarities, with a focus on the behavior of cognitive types.

to work with. In addition, we define and handle explicitly strategies and solution concepts, which allows us to investigate the implications of introducing a grain of naiveté in games with incomplete information.

The paper is organized as follows. After discussing some preliminaries in Section 2, Section 3 constructs the space of all belief hierarchies and introduces our generalized notion of a type space. Section 4 shows that there is a universal type space that generates all finite and infinite belief hierarchies, and that this space contains the universal space of Mertens and Zamir (1985) for the Harsanyi case. Section 5 introduce games with incomplete information in our setting, and Section 6 presents the analysis of global games.

## 2 Preliminaries

The Cartesian product of a collection of topological spaces  $(X_i)_{i \in I}$  (where  $I$  is an arbitrary index set) is denoted  $X$ . Given  $i \in I$ , we write  $X_{-i}$  for  $\prod_{j \in I \setminus \{i\}} X_j$ , with typical elements  $x$  and  $x_{-i}$ , respectively. The product of a collection of topological spaces is endowed with the product topology. Given a collection of functions  $f_i : X_i \rightarrow Y_i$ , we define the induced functions  $f : X \rightarrow Y$  and  $f_{-i} : X_{-i} \rightarrow Y_{-i}$  by  $f(x) := (f_i(x_i))_{i \in I}$  and  $f_{-i}(x_{-i}) := (f_j(x_j))_{j \in I \setminus \{i\}}$ . If  $(X_i)_{i \in I}$  is a family of topological spaces (possibly made disjoint by replacing some  $X_j$  with a homeomorphic copy), then  $\bigcup_{i \in I} X_i$  is endowed with the sum topology. Note that the sum of a countable collection of Polish spaces is Polish (Kechris, 1995, Prop. 3.3).

With some abuse of notation, we view every topological space  $X$  as a measurable space  $(X, \mathcal{B}(X))$ , where  $\mathcal{B}(X)$  is the Borel  $\sigma$ -algebra on  $X$ . Hence, a function said to be measurable is a Borel-measurable function. Given a Polish space  $X$ , denote by  $\Delta(X)$  the set of probability measures on the Borel  $\sigma$ -algebra  $\mathcal{B}(X)$  in  $X$ . We endow  $\Delta(X)$  with the Borel  $\sigma$ -algebra of the topology of weak convergence. The support of a probability measure  $\mu \in \Delta(X)$  is a closed set in  $X$ , denoted  $\text{supp } \mu$ , such that  $\mu(X \setminus \text{supp } \mu) = 0$ ; and for any  $U \subseteq X$ , we have that  $\mu(U \cap \text{supp } \mu) > 0$  whenever  $U$  is open in  $X$  and  $U \cap \text{supp } \mu \neq \emptyset$ . Any Borel probability measure on a Polish space has a support, and the support is unique (e.g., Aliprantis and Border, 2005, Thm. 12.14).

We extend the definition of a marginal probability measure to a union of measurable spaces. Let  $V = X \cup Y$ ,  $Q \subseteq X \times Z$  and  $W = Q \cup Y$ , where all spaces are assumed to be topological spaces; see Figure 2.1. Then for  $\mu \in \Delta(W)$  denote by  $\text{marg}|_V \mu \in \Delta(V)$  the probability measure defined by

$$\text{marg}|_V \mu(E) = \mu(\{(x, z) \in Q : x \in E\}) + \mu(E \cap Y)$$

for every measurable set  $E \subseteq V$ .

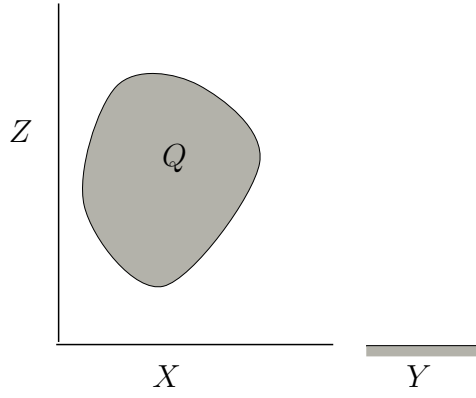


Figure 2.1: The space  $W$  (shaded gray) is the union of  $Q \subseteq X \times Z$  and of  $Y$ . The space  $V$  is the union of  $X$  and  $Y$ .

### 3 Belief hierarchies and types

This section defines the class of type spaces that allow for a finite depth of reasoning. Because each type generates a hierarchy of beliefs, we first construct the space of all belief hierarchies for our setting. Section 3.2 then introduces the type spaces that we consider.

#### 3.1 Coherent belief hierarchies

This section constructs the space of all coherent belief hierarchies. There is a fixed, finite set of players  $N$ , and a common domain of uncertainty  $S$ . A belief hierarchy specifies a player's belief about  $S$ , his beliefs about his opponents' beliefs about  $S$ , and so on, up to a certain order. That is, each hierarchy is associated with a *depth* of reasoning; a hierarchy's depth can either be finite or infinite.

Formally, assume that  $S$  is a Polish space. For a player  $i \in N$ , let  $T_i^0 := \{t_i^0\}$  and  $\tilde{T}_i^0 := \{\tilde{t}_i^0\}$ . The interpretation is that  $\tilde{t}_i^0$  is an extremely 'naive' type who has no beliefs;  $t_i^0$  is a notational "seed" on which the hierarchies of more sophisticated types will be built. Let

$$\begin{aligned}\Omega_i^0 &:= S \times \prod_{j \neq i} (T_j^0 \cup \tilde{T}_j^0) \\ \tilde{\Omega}_i^0 &:= S \times \prod_{j \neq i} \tilde{T}_j^0\end{aligned}$$

and

$$\begin{aligned}T_i^1 &:= T_i^0 \times \Delta(\Omega_i^0), \\ \tilde{T}_i^1 &:= T_i^0 \times \Delta(\tilde{\Omega}_i^0).\end{aligned}$$

These are the first-order beliefs for the types that reason beyond the first order or stop reasoning at the first order, respectively. Suppose, inductively, that we have already defined

$T_j^\ell$  and  $\tilde{T}_j^\ell$  for each player  $j \in N$  and all  $\ell \leq k$ . Define

$$\begin{aligned} T_i^{\leq k} &:= T_i^k \cup \bigcup_{\ell=0}^k \tilde{T}_i^\ell, & \Omega_i^k &:= S \times T_{-i}^{\leq k}, \\ \tilde{T}_i^{\leq k} &:= \bigcup_{\ell=0}^k \tilde{T}_i^\ell, & \tilde{\Omega}_i^k &:= S \times \tilde{T}_{-i}^{\leq k}, \end{aligned}$$

and let

$$\begin{aligned} T_i^{k+1} &:= \left\{ ((\mu_i^0, \dots, \mu_i^k), \mu_i^{k+1}) \in T_i^k \times \Delta(\Omega_i^k) : \text{marg}_{|\Omega_i^{k-1}} \mu_i^{k+1} = \mu_i^k \right\}, \\ \tilde{T}_i^{k+1} &:= \left\{ ((\mu_i^0, \dots, \mu_i^k), \tilde{\mu}_i^{k+1}) \in T_i^k \times \Delta(\tilde{\Omega}_i^k) : \text{marg}_{|\Omega_i^{k-1}} \tilde{\mu}_i^{k+1} = \mu_i^k \right\}. \end{aligned}$$

The interpretation is that  $\tilde{T}_i^k$  is the set of hierarchies that stop reasoning at order  $k$ , while the set  $T_i^k$  contains the hierarchies that also reason at higher orders. We say that a hierarchy  $\tilde{t}_i^k \in \tilde{T}_i^k$  has *depth*  $k$ . The condition on the marginals is a standard coherency condition. It is easy to see that the sets  $\tilde{T}_i^k$  are nonempty and Polish for each  $k$ .

In the limit, define

$$\tilde{T}_i^\infty := \left\{ (\mu_i^0, \mu_i^1, \dots) : (\mu_i^0, \dots, \mu_i^k) \in T_i^k \text{ for all } k \geq 0 \right\},$$

i.e.,  $\tilde{T}_i^\infty$  is the inverse limit space of the sequence  $(T_i^k)_{k \geq 0}$  endowed with the projection operators  $(\pi_{T_i^{k-1}}^{T_i^k})_{k \geq 1}$ . The belief hierarchies in  $\tilde{T}_i^\infty$  are those that “reason up to infinity.” The next result states that  $\tilde{T}_i^\infty$  is well-defined:

**Proposition 3.1.** *The inverse limit space  $\tilde{T}_i^\infty$  is nonempty and Polish.*

The set  $\tilde{T}_i^\infty$  contains the hierarchies with infinite depth, i.e., a hierarchy in  $\tilde{T}_i^\infty$  has well-defined beliefs at each order.

For future reference, it will be convenient to define

$$T_i := \tilde{T}_i^\infty \cup \bigcup_{\ell=0}^{\infty} \tilde{T}_i^\ell$$

to be the set of all belief hierarchies. Thus,  $T_i$  is nonempty and Polish.

Before we turn to the definition of type spaces that allow for finite-order reasoning, let us comment on the present construction of the space of belief hierarchies. First, one might think that the space  $T_i$  of finite and infinite belief hierarchies can also be obtained by adding finite belief hierarchies (modeled in some appropriate way) to the space of infinite belief hierarchies of Mertens and Zamir (1985). This does not work, however: it would give a space of belief hierarchies that is strictly smaller than the one constructed here. The reason is that  $T_i$  not only contains the infinite belief hierarchies constructed by Mertens and Zamir, but also belief hierarchies of infinite depth that assign a positive probability to finite-depth types of



the other player of every given order (for instance, assigning probability  $1/2^k$  to a particular belief hierarchy of depth  $k$ , for  $k \in \mathbb{N}$ ), or assign positive probability to the other player having such beliefs, and so on. It is not clear how these belief hierarchies, which are not present in the space of Mertens and Zamir, can be added ex post without a transfinite construction.

Second, what is crucial in the above construction is that each order  $k < \infty$ , the set of belief hierarchies for a player  $i$  contains both the belief hierarchies that end at that order (viz.,  $\tilde{T}_i^k$ ) as well as the belief hierarchies that continue to “grow” (viz.,  $T_i^k$ ). If the set of belief hierarchies would consist solely of the belief hierarchies that stop at that order, as in Strzalecki (2009), we would not be able to construct a space that contains all belief hierarchies, including the infinite ones.<sup>5</sup>

### 3.2 Type spaces

The previous section provided an explicit description of players’ hierarchies of beliefs. Belief hierarchies can also be described implicitly, using the concept of a type space (cf. Harsanyi, 1967–1968). Here we define type spaces that allow for finite-order reasoning. Formally, given a Polish space  $S$ , an  $S$ -based *type space* (that allows for finite-order reasoning) is a tuple

$$\mathcal{Q} := \left\langle (Q_i)_{i \in N}, (\beta_i^k)_{i \in N, k \in K_i^Q} \right\rangle,$$

where for each  $i \in N$ ,  $Q_i = \tilde{Q}_i^\infty \cup \bigcup_{\ell=0}^\infty \tilde{Q}_i^\ell$  is the type space of player  $i$ , assumed to be nonempty and Polish, and  $K_i^Q$  is the set of indices  $k \in \{0, 1, \dots\} \cup \{\infty\}$  such that  $\tilde{Q}_i^k$  is nonempty. Moreover,

- (a) the function  $\beta_i^0$  maps  $\tilde{Q}_i^0$  into the singleton  $\{\tilde{t}_i^0\}$  whenever  $\tilde{Q}_i^0$  is nonempty, i.e.,  $\beta_i^0(q_i) = \tilde{t}_i^0$  for all  $q_i \in \tilde{Q}_i^0$ ;
- (b) the function  $\beta_i^k$  is measurable and maps  $\tilde{Q}_i^k$  into  $\Delta\left(S \times \tilde{Q}_{-i}^{\leq k-1}\right)$  where  $\tilde{Q}_{-i}^{\leq k} := \bigcup_{\ell=0}^k \tilde{Q}_{-i}^\ell$  whenever  $\tilde{Q}_i^k$  is nonempty, where  $k = 1, 2, \dots$ ;
- (c) the function  $\beta_i^\infty$  is measurable and maps  $\tilde{Q}_i^\infty$  into  $\Delta(S \times Q_{-i})$  whenever  $\tilde{Q}_i^\infty$  is nonempty;
- (d) if there is  $q_i \in \tilde{Q}_i^k$  for some  $i \in N$  and finite  $k > 0$ , then for all  $j \neq i$ , there is  $k_j < k$  such that  $\tilde{Q}_j^{k_j}$  is nonempty.

For  $q_i \in \tilde{Q}_i^k$  ( $k = 0, 1, \dots, \infty$ ) we denote

$$\beta_i(q_i) = \beta_i^k(q_i).$$

---

<sup>5</sup>Again, one could add infinite belief hierarchies to this space of finite belief hierarchies ex post, but then the expanded space would still not contain the belief hierarchies that assign positive probability to types of both finite and infinite depth, and so on.

Thus, each type in  $\tilde{Q}_i^0$  is associated with the ‘naive’ type. Types in  $\tilde{Q}_i^k$  are mapped into a belief over nature and the types of lower depths, while types in  $\tilde{Q}_i^\infty$  have a belief about nature and types of all depths. Condition (d) requires that if there is some player who reasons up to a finite order, then there is a type for each of the other players that has a lower depth of reasoning. Without such a requirement, a type’s belief may not be well-defined, given that types of a finite depth can reason only about types of lower depth.

Intuitively, it is clear that type spaces generate hierarchies of beliefs. Each type (except the level-zero type) defines a belief about the state of nature and about the other players’ types, so it implicitly specifies a belief about the other players’ beliefs about the state of nature, and so on. However, as with the type spaces introduced by Harsanyi (1967–1968), this intuition needs to be made precise. Also, one could ask whether these type spaces can generate *all* belief hierarchies. We investigate these issues in the next section, by constructing a universal type space.

## 4 The universal space

This section constructs a type space that embeds all type spaces, in the sense that any type from any type space can be mapped into this type space in a way that preserves beliefs. We first define a class of belief-preserving mappings and then use these mappings to demonstrate that the space of belief hierarchies defines a universal type space, i.e., a type space that contains all type spaces. Thus, every type generates a well-defined belief hierarchy, and there is a type space, viz., the universal type space, that generates all possible belief hierarchies. We also show that the universal type space for the class of Harsanyi type spaces can be viewed as a (proper) subspace of the universal type space constructed here.

### 4.1 Belief-preserving mappings

Let  $\mathcal{Q} := \langle (Q_i)_{i \in N}, (\beta_i^k)_{i \in N, k \in K_i^{\mathcal{Q}}} \rangle$  and  $\mathcal{R} := \langle (R_i)_{i \in N}, (\lambda_i^k)_{i \in N, k \in K_i^{\mathcal{R}}} \rangle$  be type spaces on  $S$  that allow for finite-order reasoning, where we recall that

$$K_i^{\mathcal{Q}} := \{k \in \{0, 1, \dots\} \cup \{\infty\} : \tilde{Q}_i^k \neq \emptyset\}$$

is the set of indices  $k$  such that the set of types for  $i$  of depth  $k$  is nonempty, and where  $K_i^{\mathcal{R}}$  is similarly defined. We define maps, called type morphisms, from players’ type sets in the space  $\mathcal{Q}$  to the corresponding type sets in  $\mathcal{R}$ , in such a way that higher-order beliefs are preserved.

To define the concept of a type morphism, some preliminary notation will be useful. Suppose  $K_i^{\mathcal{R}} \supseteq K_i^{\mathcal{Q}}$ . For each player  $i \in N$  and  $k \in K_i^{\mathcal{Q}}$ , let  $\varphi_i^k$  be a measurable function from

$\tilde{Q}_i^k$  to  $\tilde{R}_i^k$ . Define  $\varphi_i := (\varphi_i^k)_{k \in K_i^Q}$ , and let  $\varphi := (\varphi_i)_{i \in N}$ . Also, if  $\tilde{Q}_i^k$  is nonempty for some player  $i \in N$  and finite  $k$ , then define

$$\varphi_{-i}^{<k} : \tilde{Q}_{-i}^{<k-1} \rightarrow \tilde{R}_{-i}^{<k-1}$$

by

$$\varphi_{-i}^{<k}((q_j^{m_j})_{j \neq i}) := (\varphi_j^{m_j}(q_j^{m_j}))_{j \neq i}$$

where  $q_j^{m_j} \in \tilde{Q}_j^{m_j}$ ,  $j \neq i$ . Note that by condition (d) above and the assumption that  $K_j^R \supseteq K_j^Q$  for all  $j \in N$ , the induced function  $\varphi_{-i}^{<k}$  is well-defined.

The induced function  $\varphi$  is a *type morphism* from  $\mathcal{Q}$  to  $\mathcal{R}$  if for each player  $i \in N$ ,

(i) for each  $k = 1, 2, \dots$ , type  $q_i \in \tilde{Q}_i^k$ , and  $E \in \mathcal{B}(S \times \tilde{R}_{-i}^{<k-1})$ ,

$$\lambda_i^k(\varphi_i^k(q_i))(E) = \beta_i^k(q_i)((\text{Id}_S, \varphi_{-i}^{<k})^{-1}(E)); \quad (4.1)$$

(ii) for  $q_i \in \tilde{Q}_i^\infty$ ,  $E \in \mathcal{B}(S \times R_{-i})$ ,

$$\lambda_i^\infty(\varphi_i^\infty(q_i))(E) = \beta_i^\infty(q_i)((\text{Id}_S, \varphi_{-i}^\infty)^{-1}(E)), \quad (4.2)$$

where  $\text{Id}_S$  is the identity function on  $S$ .

If  $K_i^Q \supseteq K_i^R$ , then  $\varphi$  is a *type isomorphism* if the inverse of  $\varphi_i$  is measurable for each  $i \in N$ , and satisfies (i)–(ii).

Conditions (i) and (ii) are the analogues of the standard condition that a type morphism preserves beliefs, but take into account that a type may have finite depth. In particular, if a type space only consists of types of infinite depth, the current definition of a type morphism reduces to the standard one. Lemma A.7 in the appendix shows that type morphisms preserve belief hierarchies, as do standard type morphisms (Heifetz and Samet, 1998, Prop. 5.1).

Using the concept of a type morphism, we next show that modeling belief hierarchies by types is without loss of generality in the sense that every (coherent) belief hierarchy can be modeled in this way.

## 4.2 Universality

This section shows that we do not “miss” any belief hierarchies by modeling them by types. This follows from Proposition 4.4 below, which shows that there is a type space that allows for finite-order reasoning that generates all coherent belief hierarchies.

We construct this type space using the space of belief hierarchies constructed in Section 3.1. Specifically, we take the set of types for each player to be the set of belief hierarchies, and define a belief for each type over the state of nature and the types of the other players.

The first step is to show that a belief hierarchy can be associated with a belief over the set of belief hierarchies for the other players. That is, each belief hierarchy specifies a belief about the full hierarchy of other players, not just about the individual levels of the hierarchy:

**Proposition 4.1.** (a) For each belief hierarchy  $(\mu_i^0, \mu_i^1, \dots) \in \tilde{T}_i^\infty$  there exists a unique Borel probability measure  $\mu_i$  on  $S \times T_{-i}$  such that

$$\text{marg}|_{\Omega_i^{\ell-1}} \mu_i = \mu_i^\ell$$

for all  $\ell \in \mathbb{N}$ .

(b) For each  $k > 0$  and every belief hierarchy  $(\mu_i^0, \mu_i^1, \dots, \mu_i^{k-1}, \tilde{\mu}_i^k) \in \tilde{T}_i^k$ , there exists a unique Borel probability measure  $\mu_i$  on  $S \times \tilde{T}_{-i}^{\leq k-1}$  such that

$$\text{marg}|_{\Omega_i^{\ell-1}} \mu_i = \mu_i^\ell$$

for all  $\ell = 1, \dots, k-1$ , and

$$\text{marg}|_{\Omega_i^{k-1}} \mu_i = \tilde{\mu}_i^k.$$

Thus, each belief hierarchy of player  $i$  can be associated with a belief over the basic space of uncertainty  $S$  and the other player's belief hierarchies, in such a way that  $i$ 's belief over his  $\ell$ th-order space of uncertainty coincides with his  $\ell$ th-order belief as specified by his hierarchy of beliefs. That is, the construction is *canonical* in the sense of Brandenburger and Dekel (1993). The result implies that the beliefs of a player at each order he can reason about determine his beliefs about the other players' belief hierarchies. Hence, specifying a player's beliefs about the relevant higher-order spaces of uncertainty fully specifies his beliefs.

Proposition 4.1 implicitly defines a function from the space of belief hierarchies to the spaces of Borel probability measures on nature and the other players' hierarchies. The inverse of this function assigns to each belief  $\mu_i$  over his uncertainty domain a belief hierarchy (possibly finite) by taking the marginal of  $\mu_i$  at each order. It turns out that the functions in Proposition 4.1 and their inverses are continuous, so that we have a homeomorphism for each  $k$ :

**Proposition 4.2.** There is a homeomorphism  $\psi_i^\infty : \tilde{T}_i^\infty \rightarrow \Delta(S \times T_{-i})$ . Moreover, for each  $k \in \mathbb{N}$ , there is a homeomorphism  $\psi_i^k : \tilde{T}_i^k \rightarrow \Delta((S \times \tilde{T}_{-i}^{\leq k-1}))$ .

If we define  $\psi_i^0 : \tilde{T}_i^0 \rightarrow \{\tilde{t}_i^0\}$  by setting  $\psi_i^0(\tilde{t}_i^0) := \tilde{t}_i^0$ , and extend the range of the functions  $\psi_i^k$ ,  $k < \infty$ , to  $\Delta(S \times T_{-i})$  in the usual way, then we have the following corollary:

**Corollary 4.3.** There is an embedding  $\psi_i$  from  $T_i$  to  $\Delta(S \times T_{-i})$ , which coincides with  $\psi_i^k$  on  $\tilde{T}_i^k$  for  $k = 0, 1, \dots$ , and with  $\psi_i^\infty$  on  $\tilde{T}_i^\infty$ .

Corollary 4.3 says that each  $t_i \in T_i$  for a player  $i \in N$  is associated with a belief  $\psi_i(t_i)$  on  $S \times T_{-i}$  and vice versa, and that these mappings are continuous (and therefore Borel measurable). This justifies using the term *type* to refer to an element  $t_i$  of  $T_i$ , and to  $T_i$  as the *type set* for player  $i$ . We can thus write  $\mathcal{T} := \langle (T_i)_{i \in N}, (\psi_i^k)_{i \in N, k \in \{0, 1, \dots\} \cup \{\infty\}} \rangle$  for the resulting type space.

We next show that  $\mathcal{T}$  is universal. A type space  $\mathcal{R}$  is *universal* if for any type space  $\mathcal{Q}$ , there is a unique type morphism from  $\mathcal{Q}$  to  $\mathcal{R}$  (Mertens and Zamir, 1985; Heifetz and Samet, 1998).

**Proposition 4.4.** *The type space  $\mathcal{T}$  is universal, and the universal space is unique (up to type isomorphism).*

Thus, the type space  $\mathcal{T}$  “contains” all the type spaces that allow for finite-order reasoning.

The next result shows that every type space that satisfies a nonredundancy condition can be seen as a belief-closed subset of the universal space, and, conversely, any belief-closed subset corresponds to a type space. To state the result, say that a subset  $T'$  of the set of type profiles  $T$  is a *belief-closed subset* if for all  $i \in N$  and  $t_i \in T'_i \setminus \tilde{T}_i^0$ ,

$$\psi_i(t_i)(S \times T'_{-i}) = 1.$$

Note that  $\prod_{i \in N} \tilde{T}_i^0$  is trivially a belief-closed subset. A type structure  $\mathcal{Q}$  is *nonredundant* if the hierarchy map  $h_j^{Q,k}$  is injective for all players  $j \in N$  and all  $k \in \{0, 1, \dots\} \cup \{\infty\}$ , whenever  $h_j^{Q,k}$  is defined.

**Proposition 4.5.** *Suppose  $\mathcal{Q}$  is a type space, and suppose  $\varphi$  is a type morphism from  $\mathcal{Q}$  to the universal type space  $\mathcal{T}$ . If  $\mathcal{Q}$  is nonredundant, then, for all  $i \in N$  and  $q_i \in Q_i \setminus \tilde{Q}_i^0$ ,*

$$\psi_i(\varphi_i^{\kappa(q_i)}(q_i)) \left( S \times \prod_{j \neq i} \{t_j \in T_j : t_j = \varphi_j^{\kappa(q_j)}(q_j) \text{ for some } q_j \in Q_j\} \right) = 1,$$

where  $\kappa(q_\ell) = k$  for  $\ell \in N$  and  $q_\ell \in \tilde{Q}_\ell^k$ . Conversely, if  $T'_i \subseteq T_i$ ,  $i \in N$ , is such that

$$\psi_i(q_i)(S \times T'_{-i}) = 1$$

for all  $i \in N$ , then there is a type structure  $\mathcal{Q}$  and a type morphism  $\varphi$  from  $\mathcal{Q}$  to  $\mathcal{T}$  such that for all  $i$ ,

$$T'_i = \{t_i \in T_i : t_i = \varphi_i^{\kappa(q_i)}(q_i) \text{ for some } q_i \in Q_i\}.$$

Thus, the type space  $\mathcal{T}$  is universal and contains all nonredundant type spaces as belief-closed subsets. We next turn to the question of how the universal space  $\mathcal{T}$  relates to the universal space constructed for the standard case by Mertens and Zamir (1985) and others.

### 4.3 Common belief in infinite depth of reasoning

Here we show that the universal Harsanyi space, constructed by Mertens and Zamir (1985) and others, is a belief-closed subset of the universal space constructed in the previous section, and is characterized by the event that players have an infinite depth of reasoning, and commonly believe that all players have an infinite depth of reasoning.

Formally, given a Polish space  $S$ , an  $S$ -based *Harsanyi type space* is a tuple  $\langle (\hat{Q}_i)_{i \in N}, (\hat{\beta}_i)_{i \in N} \rangle$ , where  $\hat{Q}_i$  is a Polish space for each player  $i \in N$ , and the measurable function  $\hat{\beta}_i$  maps  $\hat{Q}_i$  into the set of Borel probability measures  $\Delta(S \times \hat{Q}_{-i})$  on the set of states of nature and other players' types.

The universal Harsanyi type space can be constructed in a similar way as the universal type space  $\mathcal{T}$  (that allows for finite-order reasoning). Let  $\hat{Z}_i^0 := \{\hat{z}_i^0\}$  be an arbitrary singleton, and define

$$\hat{\Omega}_i^0 := S \times \hat{Z}_i^0,$$

and

$$\hat{Z}_i^1 := \hat{Z}_i^0 \times \Delta(\hat{\Omega}_i^0).$$

Again, assume, inductively, that we have already defined  $\hat{Z}_j^\ell$  for each player  $j \in N$  and all  $\ell \leq k$ . Define

$$\hat{\Omega}_i^k := S \times \hat{Z}_{-i}^k,$$

and let

$$\hat{Z}_i^{k+1} := \left\{ ((\mu_i^0, \dots, \mu_i^k), \mu_i^{k+1}) \in \hat{Z}_i^k \times \Delta(\hat{\Omega}_i^k) : \text{marg}_{|\hat{\Omega}_i^{k-1}} \mu_i^{k+1} = \mu_i^k \right\}.$$

The inverse limit space  $\hat{Z}_i$  for player  $i$  is the set of all  $(\mu_i^0, \mu_i^1, \dots)$  such that  $(\mu_i^0, \mu_i^1, \dots, \mu_i^k) \in \hat{Z}_i^k$  for all  $k$ . It can be shown that the analogue of Propositions 3.1 holds. Moreover, the analogue of Proposition 4.2 holds for this case: there is a homeomorphism  $\hat{\chi}_i$  from  $\hat{Z}_i$  to the set of Borel probability measures  $\Delta(S \times \hat{Z}_{-i})$  for each  $i \in N$  (cf. Heifetz, 1993). This means that we can view  $\hat{\mathcal{Z}} := \langle (\hat{Z}_i)_{i \in N}, (\hat{\chi}_i)_{i \in N} \rangle$  as a Harsanyi type space. The Harsanyi type space  $\hat{\mathcal{Z}}$  is universal with respect to the class of Harsanyi type spaces, in the sense that every Harsanyi type space can be embedded into  $\hat{\mathcal{Z}}$  via a unique type morphism for Harsanyi type spaces.

We show that the universal Harsanyi type space  $\hat{\mathcal{Z}}$  can be viewed as a belief-closed subset of the universal type space  $\mathcal{T}$ , characterized by the event that players have an infinite depth of reasoning, and there is common belief that players have an infinite depth of reasoning. To see this, note that  $\hat{\mathcal{Z}}$  corresponds to a type space  $\mathcal{Z}$  with the type set for player  $i \in N$  given by  $Z_i = \tilde{Z}_i^\infty \cup \bigcup_{k=0}^\infty \tilde{Z}_i^k$ , where  $\tilde{Z}_i^\infty := \hat{Z}_i$ , and  $\tilde{Z}_i^k = \emptyset$  for  $k < \infty$ , and the belief map  $\chi_i^\infty := \hat{\chi}_i$ .

It follows from Proposition 4.4 that  $\mathcal{Z}$  can be embedded in the universal type space  $\mathcal{T}$  via a unique type morphism. The converse clearly does not hold, as  $\mathcal{T}$  contains types that

have a finite depth of reasoning, types that assign a positive probability to types with a finite depth of reasoning, types that assign a positive probability to types that assign a positive probability to types with a finite depth of reasoning, and so on.

Moreover, because the space  $\mathcal{Z}$  is nonredundant by construction, the type space  $\mathcal{Z}$  derived from the universal Harsanyi type space  $\widehat{\mathcal{Z}}$  corresponds to a belief-closed subspace of the universal type space  $\mathcal{T}$  (by Proposition 4.5).

Can we characterize this subspace of  $\mathcal{T}$  in terms of players' higher-order beliefs? Proposition 4.7 below establishes that the space  $\mathcal{Z}$  is characterized by the event that players have an infinite depth of reasoning, and that there is common belief in the event that players have an infinite depth of reasoning.

To state the result, we define the event that a player  $i \in N$  believes an event  $E \in \mathcal{B}(S \times Q)$ . To that aim, let  $E \in \mathcal{B}(S \times Q)$ ,  $i \in N$ , and  $q_i \in Q_i$ . Then,

$$E_{q_i} := \{(s, q_{-i}) \in S \times Q_{-i} : (s, q_i, q_{-i}) \in E\}$$

is the set of tuples  $(s, q_{-i})$  that are consistent with  $E$  and the event that player  $i$  has type  $q_i$ . The set of states where  $i$  believes  $E$  (with probability 1) is then

$$B_i(E) := \{(s, q_i, q_{-i}) \in S \times (Q_i \setminus \widetilde{Q}_i^0) \times Q_{-i} : \beta_i^{\kappa(q_i)}(q_i)(E_{q_i}) = 1\},$$

where  $\kappa(q_i) = k$  if  $q_i \in \widetilde{Q}_i^k$ .

**Lemma 4.6.** *For each  $i \in N$  and  $E \in \mathcal{B}(S \times Q)$ , we have that  $B_i(E) \in \mathcal{B}(S \times Q)$ .*

Then, for  $E \in \mathcal{B}(S \times Q)$ ,

$$B(E) := \bigcap_{i \in N} B_i(E)$$

is the event that all players believe  $E$ , and we say that  $E$  is *common belief* at a state  $(s, q)$  if

$$(s, q) \in CB(E) := \bigcap_{\ell \in \mathbb{N}} [B]^\ell(E),$$

where  $[B]^1(E) := B(E)$ , and  $[B]^\ell(E) := B \circ [B]^{\ell-1}(E)$  for  $\ell > 1$ . It is immediate from Lemma 4.6 that  $B(E)$  and  $CB(E)$  are events whenever  $E$  is an event. Finally, let  $E^\infty := S \times \prod_{i \in N} \widetilde{T}_i^\infty$  be the event that players have an infinite depth of reasoning.<sup>6</sup> We then have:

**Proposition 4.7.** *Let  $\varphi$  be the unique type morphism from  $\mathcal{Z}$  to the universal type space  $\mathcal{T}$ . Then,*

$$S \times \prod_{i \in N} \varphi_i^\infty(Z_i) = E^\infty \cap CB[E^\infty].$$

---

<sup>6</sup>The set  $E^\infty$  is indeed an event. To see this, note that  $T_i$  is endowed with the sum topology and the Borel  $\sigma$ -algebra. Moreover,  $\mathcal{B}(S \times \prod_i T_i) = \mathcal{B}(S) \otimes \bigotimes_i \mathcal{B}(T_i)$  (Aliprantis and Border, 2005, Thm. 4.44).

## 5 Games with incomplete information

This section defines a class of games with incomplete information for our setting, and introduces a suitable solution concept. Formally, given a set of players  $N$  and set  $S$  of states of nature, a (*generalized*) *Bayesian game* is a tuple  $((A_i)_{i \in N}, (u_i)_{i \in N}, \mathcal{Q})$ , where for each player  $i \in N$ ,  $A_i$  is  $i$ 's action set, assumed to be measurable,  $u_i : S \times A \rightarrow \mathbb{R}$  is her utility function, and

$$\mathcal{Q} := \left\langle (Q_i)_{i \in N}, (\beta_i^k)_{i \in N, k \in K_i^Q} \right\rangle,$$

is a type space that allows for finite-order reasoning. The action set  $A_i$  may be the set of mixtures over some space of pure actions. It will be convenient to define  $\beta_i$  to be the function from  $Q_i$  to  $\Delta(S \times Q_{-i})$  which coincides with  $\beta_i^k$  on  $\tilde{Q}_i^k$  for each  $k$  such that  $\tilde{Q}_i^k$  is nonempty.

It is straightforward to extend the notion of interim correlated rationalizability (Dekel et al., 2007). In standard Bayesian games this solution concept embodies common belief of rationality; and it allows a type to believe that her opponents' actions are correlated even conditional on them having a particular profile of types and given that a particular state of nature obtains (see also Chen et al. (2010) and Battigalli et al. (2011)). For each player  $i \in N$ , let

$$R_i^0(q_i) := A_i$$

and, for  $k > 1$ , define inductively

$$R_i^k(q_i) := \left\{ \begin{array}{l} a_i \in A_i : \text{there exists a measurable } \mu : S \times Q_{-i} \rightarrow \Delta(A_{-i}) \text{ s.t.} \\ \mu((s, q_{-i})) \in \Delta\left(\prod_{j \neq i} R_j^{k-1}(q_j)\right) \quad \forall q_{-i} \in Q_{-i}, s \in S; \text{ and} \\ a_i \in \arg \max_{a'_i \in A_i} \int_{S \times Q_{-i}} d\beta_i(q_i) \int_{A_{-i}} u_i(s, a'_i, \cdot) d\mu(s, q_{-i}) \end{array} \right\}.$$

The *interim correlated rationalizable actions* of type  $q_i$  for player  $i \in N$  are

$$R_i^\infty(q_i) = \bigcap_{k=0}^{\infty} R_i^k(q_i).$$

**Remark.** At first sight, it may seem that our solution concept is not entirely consistent with the idea that players can have a finite depth of reasoning. Specifically, the conjecture  $\mu$  in the definition of  $R_i^k(q_i)$  is defined for every type profile  $q_{-i} \in Q_{-i}$  of  $i$ 's opponents for every  $q_i$ , *y compris* type profiles  $\check{q}_{-i}$  that correspond to a depth of reasoning that is greater than that of  $q_i$ . But since such type profiles  $\check{q}_{-i}$  are outside the support of the belief  $\beta_i(q_i)$  for type  $q_i$ , the beliefs  $\mu(\check{q}_{-i})$  do not affect the definition of  $R_i^k(q_i)$ . In other words, at the cost of additional notation we could have restricted the domain of  $\mu$  in the definition of  $R_i^k(q_i)$  to the type profiles which  $q_i$  “conceives” without altering the surviving set of actions for type  $q_i$ .  $\triangleleft$



In the next section, we study the set of rationalizable actions in a game with strategic complements, and show that introducing a ‘grain of naiveté’ can change the set of rationalizable outcomes.

## 6 A game with a grain of naiveté

To exemplify the insights the framework can yield, we consider the following game, taken from Carlsson and van Damme (1993) and Morris and Shin (2003). Two players, indexed by  $i = 1, 2$ , decide simultaneously whether to invest ( $I$ ) or not to invest ( $N$ ). The payoff matrix is

	$I$	$N$
$I$	$1 - s, 1 - s$	$-s, 0$
$N$	$0, -s$	$0, 0$

The state of the world  $s$  is drawn uniformly from the interval  $[-1, 2]$ .<sup>7</sup> Thus when  $s < 0$  investing ( $I$ ) is strongly dominant for each player. At the other end, when  $s > 1$  not investing ( $N$ ) is strongly dominant for each player. In the middle range, when  $s \in [0, 1]$ , investment is strictly preferable if and only if the player believes that the other player invests with probability greater than  $s$ . Thus, if the game were played with complete information, both actions would be rationalizable.

Prior to playing, however, for each realized state of nature  $s$ , the two players get uniform i.i.d. signals from the interval  $[s - \varepsilon, s + \varepsilon]$ , where  $\varepsilon > 0$  is small ( $\varepsilon \leq \frac{1}{2}$ , say). With the signal  $x_i$ , player  $i$ ’s posterior on the state of nature is uniform in  $[x_i - \varepsilon, x_i + \varepsilon] \cap [-1, 2]$  (by Bayes’ rule). Moreover, conditional on each such conceivable state of nature  $s \in [x_i - \varepsilon, x_i + \varepsilon] \cap [-1, 2]$ , the player believes that the other player’s signal  $x_{-i}$  is itself uniformly distributed in the interval  $[s - \varepsilon, s + \varepsilon]$ . In particular, upon receiving a signal  $x_i \in [-1 + \varepsilon, 2 - \varepsilon]$ , the posterior probability that  $i$  assigns to the other player having received a signal  $x_{-i}$  less than  $z \in [x_i - \varepsilon, x_i]$  is

$$f(z; x_i, \varepsilon) := \int_{x_i - \varepsilon}^{z + \varepsilon} \left( \int_{s - \varepsilon}^z \frac{dy}{2\varepsilon} \right) \frac{ds}{2\varepsilon} = \frac{1}{8\varepsilon^2} (z - x_i + 2\varepsilon)^2.$$

If all types have an infinite depth of reasoning, then investing is the only rationalizable action for a player who received a signal less than  $\frac{1}{2}$ , and not investing is the unique rationalizable action for a player who received a signal greater than  $\frac{1}{2}$  (Carlsson and van Damme, 1993). Thus, even as signals become arbitrarily accurate ( $\varepsilon \rightarrow 0$ ), so that the game, in a sense, converges to one with complete information, each player has a unique rationalizable action for every possible signal (except for the knife-edge signal  $\frac{1}{2}$ ). This is in contrast with

---

<sup>7</sup>With respect to the notation in Morris and Shin (2003), we use the change of variable  $s = 1 - \theta$ .

the complete-information case, where both actions are rationalizable whenever  $s \in (0, 1)$ . In this sense the concept of interim correlated rationalizability suffers from a particular kind of discontinuity.

We show that this discontinuity need *not* occur as long as there is even a slight ‘grain of naiveté’, i.e., when a finite depth of reasoning is conceivable, though perhaps only after an arbitrarily long, or even *transfinitely* long, chain of mutual beliefs, and with a vanishing probability.

Consider the universal space  $\langle (T_i)_{i=1,2}, (\psi_i^k)_{i=1,2,k \in \{0,1,\dots\} \cup \{\infty\}} \rangle$  from Section 4 with the set of states of nature  $S = [-1, 2]$ . Restrict attention to its maximal belief-closed subspace  $(G_i)_{i=1,2}$  pertinent to the above information structure, i.e., the union of all belief-closed subspaces in which to each type  $t_i, i = 1, 2$ , there corresponds some signal  $x_i \in [-1 - \varepsilon, 2 + \varepsilon]$  such that  $t_i$ ’s first order belief on  $S$  (if it has any) is uniform in the interval  $[x_i - \varepsilon, x_i + \varepsilon] \cap [-1, 2]$ , its second order belief (if it has any) is such that conditional on each such  $s \in [x_i - \varepsilon, x_i + \varepsilon] \cap [-1, 2]$  the other player’s signal  $x_{-i}$  is uniformly distributed in the interval  $[s - \varepsilon, s + \varepsilon]$ .

Within this subspace  $(G_i)_{i=1,2}$  of the universal space, consider, first, for  $i = 1, 2$ , the sequence of subsets of finite-depth types

$$\bar{G}_i^n \subseteq \tilde{T}_i^n \cap G_i, \quad n = 0, 1, 2, \dots,$$

where  $\bar{G}_i^0 = \tilde{T}_i^0 \cap G_i$  and, inductively,  $\bar{G}_i^n$  are the types in  $\tilde{T}_i^n \cap G_i$  who assign probability 1 to the types  $\bar{G}_{-i}^{n-1}$  of the other player. That is,  $\bar{G}_i^n$  are  $i$ ’s types of depth  $n$  who are certain that the other player has a type of depth  $n - 1$  who is certain that *his* opponent is of depth  $n - 2$ , etc. Thus, types in  $\bar{G}_i^n$  are as optimistic as possible, given their depth  $n$  of reasoning, about the (mutual beliefs about) the other’s depth of reasoning.

The unique 0-depth type in  $\bar{G}_i^0$  ‘behaves erratically’ (‘it doesn’t think’)—by definition both actions are rationalizable for it. Types in  $\bar{G}_i^1$  only reason about the state of nature (and not about the other player’s belief and corresponding choice), so investing is the unique rationalizable action for them if and only if they assign probability greater than  $\frac{1}{2}$  to the negative states  $s < 0$ . This is the case if and only if  $x_i < 0$ .

Consider now a type in  $\bar{G}_i^2$ . For such a type who received the signal  $x_i$ , investing is the only rationalizable action if and only if it assigns probability greater than  $\frac{1}{2}$  to the negative states (which is the case if  $x_i < 0$ ), or otherwise it assigns probability greater than  $x_i$ —the expected value of the state of nature  $s$  according to the player’s posterior—to the event that the other player is in  $\bar{G}_{-i}^1$  and has received a signal  $x_{-i} < 0$  (so that the other player’s only rationalizable action is to invest). A type with signal  $x_i$  assigns probability greater than  $x_i$  to the event that the other player has received a signal  $x_{-i} < 0$  if and only if  $x_i > 0$  and

$$f(0; x_i, \varepsilon) > x_i,$$

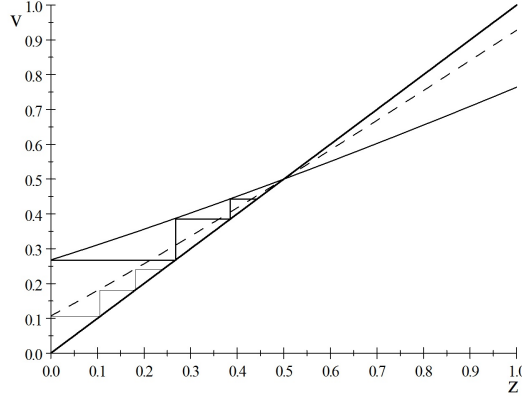


Figure 6.1: The function  $v(z; \varepsilon)$  for  $\varepsilon = 0.5$  (solid line) and  $\varepsilon = 0.1$  (dashed line) around the 45-degree line (thick solid line). The thick zigzag staircase traces on the vertical axis the values of  $v(0; 0.5)$ ,  $v^2(0; 0.5)$  and  $v^3(0; 0.5)$ ; the thin zigzag staircase traces on the vertical axis the values of  $v(0; 0.1)$ ,  $v^2(0; 0.1)$  and  $v^3(0; 0.1)$ .

where we recall that  $f(z; x_i, \varepsilon)$  is the posterior probability that  $i$  assigns to the other player having received a signal  $x_{-i}$  less than  $z$ . In general, we have that  $x_i > z$  and

$$f(z; x_i, \varepsilon) > x_i$$

if and only if  $x_i$  is less than

$$v(z; \varepsilon) := (z + 2\varepsilon + 4\varepsilon^2) - 2\varepsilon\sqrt{2(z + 2\varepsilon + 2\varepsilon^2)},$$

where the right-hand side is the smaller of the two roots of the quadratic equation  $f(z; x_i, \varepsilon) = x_i$ . Hence, for a type in  $\bar{G}_i^2$  who has received the signal  $x_i$ , the unique rationalizable action is  $I$  if and only if

$$x_i < v(0; \varepsilon).$$

By the same logic, for a type in  $\bar{G}_i^3$  who received the signal  $x_i$  investing is the only rationalizable action if and only if

$$x_i < v^2(0; \varepsilon) := v(v(0; \varepsilon); \varepsilon)$$

Inductively, for a type in  $\bar{G}_i^{k+1}$  who received the signal  $x_i$  the only rationalizable action is to invest if and only if

$$x_i < v^k(0; \varepsilon) := v(v^{k-1}(0; \varepsilon); \varepsilon)$$

The function  $v(z; \varepsilon)$  is depicted in Figure 6.1 for various values of  $\varepsilon$ . The graph demonstrates how  $v^k(0; \varepsilon) \xrightarrow[k \rightarrow \infty]{} \frac{1}{2}$  even when  $\varepsilon$  is small ( $\varepsilon = 0.1$  in the graph). At the same time, the graph also demonstrates how for any fixed  $k$  it is the case that  $v^k(0; \varepsilon) \xrightarrow[\varepsilon \rightarrow 0]{} 0$  (the graph demonstrates this for  $k = 1, 2, 3$ ). These properties hold in general:

**Lemma 6.1.** *Investing is the unique rationalizable action for a type in  $\bar{G}_i^1$  if and only if its signal  $x_i$  is less than 0. Moreover, there is a sequence of thresholds  $v^k(0; \varepsilon) > 0$  such that*

investing is the unique rationalizable action for a type in  $\bar{G}_i^{k+1}$  if and only if its signal  $x_i$  is less than  $v^k(0; \varepsilon)$ . For a fixed  $\varepsilon > 0$ , we have

$$v^k(0; \varepsilon) \xrightarrow{k \rightarrow \infty} \frac{1}{2},$$

while for every fixed  $k$  we have

$$v^k(0; \varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 0.$$

A symmetric argument implies:

**Lemma 6.2.** *Not investing is the unique rationalizable action for a type in  $\bar{G}_i^1$  if and only if its signal  $x_i$  is greater than 1. Moreover, there is a sequence of thresholds  $w^k(1; \varepsilon) < 1$  such that not investing is the unique rationalizable action for a type in  $\bar{G}_i^{k+1}$  if and only if its signal  $x_i$  is greater than  $w^k(1; \varepsilon)$ . For any fixed  $\varepsilon > 0$ , we have*

$$w^k(1; \varepsilon) \xrightarrow{k \rightarrow \infty} \frac{1}{2},$$

while for every fixed  $k$ , we have

$$w^k(1; \varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 1.$$

Together, Lemmas 6.1 and 6.2 immediately imply:

**Proposition 6.3.** *For every  $k$  and for every state of nature  $s \in (0, 1)$ , there exists  $\varepsilon_k(s) > 0$  such that if the accuracy of the signals is high enough, i.e.,  $\varepsilon < \varepsilon_k(s)$ , both investing and not investing are rationalizable for types in  $\bar{G}_i^k$  that have a signal  $x_i \in [s - \varepsilon, s + \varepsilon]$ .*

Thus, when  $s \in (0, 1)$ , for all types in  $\bar{G}_i^k$ , the set of rationalizable actions eventually contains both actions as  $\varepsilon \rightarrow 0$ , as in the limit game with complete information. This means that the rationalizability correspondence is continuous in this respect.

**Remark.** In fact, lemmas 6.1 and 6.2 imply a somewhat stronger result, namely that the same threshold  $\varepsilon_k(s)$  applies to an interval of states of nature, not only to a unique state  $s$  (and the same holds true for propositions 6.4, 6.5 and 6.6 below). For example, lemma 6.1 implies that for every  $k$  and for every state of nature  $s \in (0, \frac{1}{2})$ , there exists  $\varepsilon_k(s) > 0$  such that if  $\varepsilon < \varepsilon_k(s)$ , both investing and not investing are rationalizable for types in  $\bar{G}_i^k$  that have a signal  $x_i \in [s' - \varepsilon, s' + \varepsilon]$  for every  $s' \in [s, \frac{1}{2} - \varepsilon_k(s)]$ .  $\triangleleft$

Consider now a type  $t_i \in G_i$  of *infinite* depth of player  $i$  that for a given  $p \in (0, 1)$  assigns probability  $p^k(1 - p)$  to the set of types  $\bar{G}_{-i}^k$  of the other player, for  $k = 0, 1, 2, \dots$ . As  $p \rightarrow 1$  this infinite-depth type  $t_i$  assigns larger weights to types of large finite depth. Proposition 6.3 nevertheless implies:

**Proposition 6.4.** *For every  $p < 1$  and state of nature  $s \in (0, 1)$ , there exists a small enough  $\varepsilon_p(s) > 0$  such that if the accuracy of the signals is high enough, i.e.,  $\varepsilon < \varepsilon_p(s)$ , both investing and not investing are rationalizable for the types  $t_i \in G_i$  with signals  $x_i \in [s - \varepsilon, s + \varepsilon]$  that have infinite depth and that assign probability  $p^k(1 - p)$  to the set of types  $\bar{G}_{-i}^k$  of the other player, for  $k = 0, 1, 2, \dots$ .*

Next, consider a sequence of types  $t_{m,-i} \in G_{-i}$ ,  $m = 0, 1, 2, \dots$  of player  $-i$  that assign probability  $p_m^k(1 - p_m)$  to the event that the other player  $i$  is in  $\bar{G}_i^k$ ,  $k = 1, 2, \dots$ , where  $p_m \nearrow 1$ . Then, consider also a type  $t_i \in G_i$  for the other player  $i$  that has an infinite depth and assigns probability  $r^m(1 - r)$  to  $t_{m,-i}$ . By a similar argument to that in the proof of Proposition 6.4, one can show:

**Proposition 6.5.** *For every  $r < 1$  and state of nature  $s \in (0, 1)$ , there exists a small enough  $\varepsilon_r(s) > 0$  such that if the accuracy of the signals is good enough, i.e.,  $\varepsilon < \varepsilon_r(s)$ , both investing and not investing are rationalizable for the types  $t_i \in G_i$  with signals  $x_i \in [s - \varepsilon, s + \varepsilon]$  that have infinite depth and that assign probability  $r^m(1 - r)$  to the type  $t_{m,-i}$ ,  $m = 0, 1, 2, \dots$ .*

Notice that the types  $t_i \in G_i$  to which Proposition 6.5 applies not only have infinite depth, but on top of that they assign probability 1 to the event that the other player  $-i$  is one of  $t_{m,-i}$ ,  $m = 0, 1, 2, \dots$  and thus also has infinite depth. That these types  $t_i$  have multiple rationalizable actions is the result of the ‘grain of naiveté’ in their belief: these types believe that even though the other player  $-i$  has an infinite depth of reasoning, that other player  $-i$  believes that player  $i$  herself has a finite-depth of reasoning (though most probably a very large finite depth).

Repeating the same logic again and again, we can make this ‘grain of naiveté’ as small as we like, and still have that both actions are rationalizable:

**Proposition 6.6.** *For every state of nature  $s \in (0, 1)$  and for every  $\ell \geq 1$  there exists a type  $t_{i,\ell} \in G_i$  with a signal  $x_i \in [s - \varepsilon, s + \varepsilon]$  that has infinite depth, is certain that the other player has infinite depth, and is certain that  $\dots$  ( $\ell$  times)  $\dots$  that the other player has infinite depth of reasoning, and nevertheless there exists a small enough  $\varepsilon(s; t_{i,\ell}) > 0$  such that when the accuracy of the signals is good enough, i.e.,  $\varepsilon < \varepsilon(s; t_{i,\ell})$ , both investing and not investing are rationalizable for  $t_{i,\ell}$ .*

To sum up, the classical result in the literature on global games, by which as  $\varepsilon$  vanishes a unique rationalizable action survives, is known to hold for the Mertens-Zamir types in  $G_i$ , for whom there is common belief in infinite reasoning among the players. But these Mertens-Zamir types form only one particular subspace of  $(G_i)_{i=1,2}$ , and in other parts of  $(G_i)_{i=1,2}$  both actions may remain rationalizable even as the noise level  $\varepsilon$  tends to zero.

# A Proofs

## A.1 Proof of Proposition 3.1

The proof follows from a number of lemmas:

**Lemma A.1.** *For  $i \in N$  and  $k \in \mathbb{N}$ ,  $\Omega_i^k$ ,  $\tilde{\Omega}_i^k$ ,  $T_i^k$  and  $\tilde{T}_i^k$  are Polish.*

**Proof.** The proof is by induction. Clearly,  $T_i^0$  and  $\tilde{T}_i^0$  are Polish for each  $i \in \mathbb{N}$ , so that  $\Omega_i^0$ ,  $\tilde{\Omega}_i^0$  and  $T_i^1$  and  $\tilde{T}_i^1$  are also Polish. Suppose  $\Omega_i^\ell$ ,  $\tilde{\Omega}_i^\ell$ ,  $T_i^{\ell+1}$  and  $\tilde{T}_i^{\ell+1}$  are Polish spaces for each  $i \in N$  and  $\ell \leq k-1$ . It follows immediately that  $\Omega_i^k$  and  $\tilde{\Omega}_i^k$  are Polish, so that it remains to show that  $T_i^{k+1}$  and  $\tilde{T}_i^{k+1}$  are Polish spaces. First note that  $\Delta(\Omega_i^k)$  and  $\Delta(\tilde{\Omega}_i^k)$  are Polish. We thus need to establish that  $T_i^{k+1}$  and  $\tilde{T}_i^{k+1}$  are a closed subset of  $T_i^k \times \Delta(\Omega_i^k)$  and  $T_i^k \times \Delta(\tilde{\Omega}_i^k)$ , respectively. We prove the claim for  $T_i^{k+1}$ ; the proof for  $\tilde{T}_i^{k+1}$  is similar. Let  $t_i^{k+1} = (\mu_i^0, \dots, \mu_i^{k+1}) \in T_i^k \times \Delta(\Omega_i^k)$  and suppose there is a sequence  $(t_i^n)_{n \in \mathbb{N}}$  in  $T_i^{k+1}$ , where  $t_i^n = (\mu_i^{0,n}, \mu_i^{2,n}, \dots, \mu_i^{k+1,n})$ , such that  $t_i^n \rightarrow t_i$ . It is sufficient to show that  $t_i \in T_i^k$ . If we show that

$$\text{marg}|_{\Omega_i^{k-1}} \mu_i^{k+1,n} \rightarrow \text{marg}|_{\Omega_i^{k-1}} \mu_i^{k+1}, \quad (\text{A.1})$$

and

$$\mu_i^{k,n} \rightarrow \mu_i^k, \quad (\text{A.2})$$

the proof is complete: Because  $t_i^n \in T_i^{k+1}$  for all  $n$ , it follows that

$$\text{marg}|_{\Omega_i^{k-1}} \mu_i^{k+1} = \mu_i^k$$

so that  $t_i \in T_i^{k+1}$ . But using that  $T_i^k \times \Delta(\Omega_i^k)$  is endowed with the product topology, (A.1) and (A.2) follow immediately from the assumption that  $t_i^n \rightarrow t_i^{k+1}$ .  $\square$

**Lemma A.2. (Heifetz, 1993, Thm. 6)** *For any  $(\mu_i^0, \dots, \mu_i^k) \in T_i^k$ , there exists  $\mu_i^{k+1} \in \Delta(\Omega_i^k)$  such that  $(\mu_i^0, \dots, \mu_i^k, \mu_i^{k+1}) \in T_i^{k+1}$ .*

**Proof.** Let  $i \in N$ , and fix  $(\mu_i^0, \dots, \mu_i^k) \in T_i^k$ . It suffices to show that there is a continuous mapping  $f_i^k : \Delta(S \times T_{-i}^{k-1}) \rightarrow \Delta(S \times T_{-i}^k)$  such that  $(\mu_i^0, \dots, \mu_i^k, f_i^k(\mu_i^k)) \in T_i^{k+1}$ .

To show this, we construct a continuous mapping  $F_i^k : S \times T_{-i}^{k-1} \rightarrow S \times T_{-i}^k$  for  $k = 1, 2, \dots$ , such that  $\pi_{S \times T_{-i}^{k-1}}^{S \times T_{-i}^k} \circ F_i^k$  is the identity function on  $\Omega_i^{k-1}$ , where, with some abuse of notation,  $S \times T_{-i}^{-1} := S$ . To construct such a function  $F_i^k$  for all  $k$ , fix  $s^* \in S$  and define  $F_i^0 : S \rightarrow S \times T_{-i}^0$  by

$$F_i^0(s) = (s, (\delta_{s^*})_{j \in N \setminus \{i\}})$$

for  $s \in S$ . Clearly,  $F_i^0$  is continuous, and  $\pi_{\Omega_i^{-1}}^{\Omega_i^0} \circ F_i^0$  is the identity function on  $\Omega_i^{-1}$ .

Suppose, by induction, that we have defined  $F_i^k$  for  $i \in N$  and  $\ell \leq k$ . Then let  $f_i^k : \Delta(S \times T_{-i}^{k-1}) \rightarrow S \times T_{-i}^k$  be defined by:

$$f_i^k(\mu_i^k) = \mu_i^k \circ (F_i^k)^{-1}$$

for  $\mu_i^k \in \Delta(S \times T_{-i}^{k-1})$ . Then  $f_i^k$  is continuous, and, by the induction hypothesis,  $\pi_{S \times T_{-i}^{k-1}}^{S \times T_{-i}^k} \circ F_i^k$  is the identity function on  $S \times T_{-i}^{k-1}$ . It follows that

$$\text{marg}|_{\Omega_i^{k-1}} f_i^k(\mu_i^k) = f_i^k(\mu_i^k) \circ \left( \pi_{S \times T_{-i}^{k-1}}^{S \times T_{-i}^k} \right) = \mu_i^k.$$

It remains to define  $F_i^{k+1}$ . For  $(s, (\mu_j^0, \dots, \mu_j^k)_{j \in N \setminus \{i\}}) \in S \times T_i^k$ , let

$$F_i^{k+1}(s, (\mu_j^0, \dots, \mu_j^k)_{j \in N \setminus \{i\}}) = (s, (\mu_j^0, \dots, \mu_j^k, f_j^k(\mu_j^k))_{j \in N \setminus \{i\}}).$$

Again,  $F_i^{k+1}$  is continuous, and  $\pi_{S \times T_{-i}^k}^{S \times T_{-i}^{k+1}} \circ F_i^{k+1}$  is the identity function on  $S \times T_i^k$ .  $\square$

By Lemma A.2,  $T_i^k$  is nonempty. Also, the projection  $\pi_{T_i^{k-1}}^{T_i^k}$  is surjective. It follows that the inverse limit space  $\tilde{T}_i^\infty$  is nonempty, where  $T_i \subseteq \prod_{k \in \mathbb{N}} T_i^k$  is endowed with the relative product topology (e.g., Hocking and Young, 1988, Lemma 2.84). Since  $\tilde{T}_i^\infty$  is a closed subset of the Polish space  $\prod_{k \in \mathbb{N}} T_i^k$ , it is Polish (Hocking and Young, 1988).  $\square$

## A.2 Proof of Lemma A.5

We start with some preliminary results.

**Lemma A.3.** *Let  $X = \bigcup_{m \in \Lambda} X^m$  be a countable union of topological spaces, endowed with the sum topology. Let  $B \in \mathcal{B}(X)$  and  $\ell \in \Lambda$ . Then  $B \cap X^\ell \in \mathcal{B}(X^\ell)$ .*

**Proof.** Suppose not. Then there is a  $\sigma$ -algebra  $\mathcal{A}^\ell$  on  $X^\ell$  that contains the open sets in  $X^\ell$  such that  $B \cap X^\ell \notin \mathcal{A}^\ell$ . It is sufficient to show that there is a  $\sigma$ -algebra  $\mathcal{A}$  on  $X$  that contains the open sets in  $X$  such that  $B \notin \mathcal{A}$ .

To show this, let  $\mathcal{A}$  be the  $\sigma$ -algebra on  $X$  generated by the open sets in  $X^m$ , where  $m \in \Lambda$ ,  $m \neq \ell$ , and by the sets in  $\mathcal{A}^\ell$ . We claim that  $\mathcal{A}$  contains the open sets in  $X$ . To see this, suppose that  $U$  is open in  $X$ . As  $X$  is endowed with the sum topology, it follows that  $U \cap X^m$  is open in  $X^m$  for all  $m \in \Lambda$ . Since  $\mathcal{A}^\ell$  contains the open sets, it follows that  $U \cap X^m \in \mathcal{A}$  for all  $m \in \Lambda$ . As  $U$  is a countable union of the sets  $U \cap X^m$ ,  $m \in \Lambda$ , in  $\mathcal{A}$ , and since  $\mathcal{A}$  is a  $\sigma$ -algebra, it follows that  $U \in \mathcal{A}$ .

We claim that  $B \notin \mathcal{A}$ . To see this, note that if  $B \in \mathcal{A}$ , it follows that  $B \cap X^\ell \in \mathcal{A}$ . But then  $B \cap X^\ell \in \mathcal{A}^\ell$ , a contradiction.  $\square$

**Lemma A.4.** *Let  $X = \bigcup_{m \in \Lambda} X^m$  be a union of topological spaces, endowed with the sum topology. Let  $B \in \mathcal{B}(X^\ell)$  for some  $\ell \in \Lambda$ . Then  $B \in \mathcal{B}(X)$ .*

**Proof.** Let  $\mathcal{A}$  be the  $\sigma$ -algebra on  $X$  generated by the open sets in  $X^\ell$  and the sets  $X^m$ ,  $m \in \Lambda$ . Then  $\mathcal{B}(X) \supseteq \mathcal{A}$ . It is therefore sufficient to show that  $B \in \mathcal{A}$ . But this follows directly from the definitions.  $\square$

We are now ready to prove Lemma A.5. The proof is by induction. As noted above, the functions  $h_i^{Q,0,0}$ ,  $h_i^{Q,1,0}$ , and  $h_i^{Q,1,1}$  are well-defined and measurable (as is  $h_i^{Q,<1,0}$ ) for every player  $i$  (whenever the respective domains are nonempty). Let  $k = 1, 2, \dots$ . Suppose that the functions  $h_i^{Q,k,\ell}$  and  $h_i^{Q,k,k}$  are well-defined and measurable whenever  $\tilde{Q}_i^k$  is nonempty. It suffices to show that:

- (i) The function  $h_i^{Q,<k+1,\ell}$  is well-defined and measurable for  $\ell = 0, 1, \dots, k$ .
- (ii) The function  $h_i^{Q,k+1,\ell}$  is well-defined and measurable for  $\ell = 0, 1, \dots, k+1$ .

To prove (i), first note that  $\tilde{Q}_i^{\leq k}$  is nonempty whenever  $\tilde{Q}_i^k$  is nonempty. It follows directly from the induction hypothesis that  $h_i^{Q,<k+1,\ell}$  and  $h_i^{Q,<k+1,k}$  are well-defined for  $\ell = 0, 1, \dots, k-1$ , i.e.,

$$h_i^{Q,<k+1,\ell}(\tilde{Q}_i^{\leq k}) \subseteq T_i^{\leq \ell}, \quad \text{and} \quad h_i^{Q,<k+1,k}(\tilde{Q}_i^{\leq k}) \subseteq \tilde{T}_i^{\leq k}.$$

To show that  $h_i^{Q,<k+1,k}$  is measurable, let  $B \in \mathcal{B}(\tilde{T}_i^{\leq k})$ . Then,

$$\begin{aligned} (h_i^{Q,<k+1,k})^{-1}(B) &= \{q_i \in \tilde{Q}_i^{\leq k} : h_i^{Q,<k+1,k}(q_i) \in B\} \\ &= \bigcup_{m=0}^k \{q_i \in \tilde{Q}_i^m : h_i^{Q,m,m}(q_i) \in B \cap \tilde{T}_i^m\}. \end{aligned}$$

Hence, it suffices to show that for all  $\ell = 0, \dots, k$ ,

$$\{q_i \in \tilde{Q}_i^\ell : h_i^{Q,\ell,\ell}(q_i) \in B \cap \tilde{T}_i^\ell\} \in \mathcal{B}(\tilde{Q}_i^{\leq k}). \quad (\text{A.3})$$

By Lemma A.3, we have that  $B \cap \tilde{T}_i^\ell \in \mathcal{B}(\tilde{T}_i^\ell)$ . It then follows from the measurability of  $h_i^{Q,\ell,\ell}$  that

$$\{q_i \in \tilde{Q}_i^\ell : h_i^{Q,\ell,\ell}(q_i) \in B \cap \tilde{T}_i^\ell\} \in \mathcal{B}(\tilde{Q}_i^\ell),$$

so that (A.3) follows from Lemma A.4. The proof that  $h_i^{Q,<k+1,\ell}$  is measurable for  $\ell = 0, \dots, k-1$  is similar and thus omitted.

The proof of (ii) consists of two parts. We first show that  $h_i^{Q,k+1,\ell}$  and  $h_i^{Q,k+1,k+1}$  are well-defined for  $\ell = 0, 1, \dots, k$  whenever  $\tilde{Q}_i^{k+1}$  is nonempty. That is, suppose  $\tilde{Q}_i^{k+1}$  is nonempty. Then,

$$h_i^{Q,k+1,\ell}(\tilde{Q}_i^{k+1}) \subseteq T_i^\ell \quad \text{and} \quad h_i^{Q,k+1,k+1}(\tilde{Q}_i^{k+1}) \subseteq \tilde{T}_i^{k+1}$$



Clearly,  $h_i^{Q,k+1,0}(\tilde{Q}_i^{k+1}) \subseteq T_i^0$ . Let  $\ell = 1, \dots, k-1$ , and suppose  $h_i^{Q,k+1,\ell-1}(\tilde{Q}_i^{k+1}) \subseteq T_i^{\ell-1}$ . We will show that  $h_i^{Q,k+1,\ell}(\tilde{Q}_i^{k+1}) \subseteq T_i^\ell$ . From the induction hypothesis and (i) it follows that  $h_{-i}^{Q,<k+1,\ell-1}$  is well-defined and measurable (recall condition (d) in the definition of a type space). Hence, for all  $q_i \in \tilde{Q}_i^{k+1}$ ,

$$h_i^{Q,k+1,\ell}(q_i) = (h_i^{Q,k+1,\ell-1}(q_i), \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1}) \in T_i^\ell \times \Delta(S \times T_{-i}^{\leq \ell-1})$$

where we have used the induction hypothesis. If  $\ell = 1$ , then we are done. So suppose  $\ell = 2, 3, \dots, k$ . We need to show that a player's higher-order beliefs are coherent, i.e., for each  $q_i \in \tilde{Q}_i^{k+1}$ ,

$$\text{marg}|_{\Omega_i^{\ell-2}} \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1} = \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-2})^{-1}.$$

Fix  $E \in \mathcal{B}(\Omega_i^{\ell-2})$ . Then, using the extended definition of the marginal,

$$\begin{aligned} & \text{marg}|_{\Omega_i^{\ell-2}} \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1}(E) \\ &= \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1}(\{(s, (\mu_{-i}^0, \dots, \mu_{-i}^{\ell-2}, \mu_{-i}^{\ell-1})) \in (S \times T_{-i}^{\leq \ell-1}) : \\ & \quad (s, (\mu_{-i}^0, \dots, \mu_{-i}^{\ell-2})) \in E\}) + \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1}(E \cap (S \times \tilde{T}_{-i}^{\leq \ell-2})) \\ &= \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-2})^{-1}(\{(s, (\mu_{-i}^0, \dots, \mu_{-i}^{\ell-2})) \in S \times T_{-i}^{\ell-2} : \\ & \quad (s, (\mu_{-i}^0, \dots, \mu_{-i}^{\ell-2})) \in E\}) + \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-2})^{-1}(E \cap (S \times \tilde{T}_{-i}^{\leq \ell-2})) \\ &= \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1}(E), \end{aligned}$$

so that  $h_i^{Q,k+1,\ell}(q_i) \in T_i^\ell$  for  $\ell = 2, 3, \dots, k$ . A similar argument shows that  $h_i^{Q,k+1,k+1}(q_i) \in \tilde{T}_i^{k+1}$ .

Next, we show that  $h_i^{Q,k+1,\ell}$  is measurable, where  $\ell = 0, 1, \dots, k+1$ . For  $\ell = 0$ , this is immediate. So let  $\ell = 1, 2, \dots, k+1$ , and suppose the claim is true for  $\ell-1$ . It then follows directly from the induction hypothesis and (i) that the claim is true for  $\ell$  (recall that the image measure induced by a measurable function from a metrizable space into a metrizable space is measurable).  $\square$

### A.3 Proof of Proposition 4.1

We first prove the first claim. By Proposition 3.1, the space  $S \times \tilde{T}_{-i}^\infty$  is a nonempty Polish space. By a version of the Kolmogorov consistency theorem, for any  $t_i^\infty = (\mu_i^0, \mu_i^1, \dots) \in \tilde{T}_i^\infty$ , there exists a unique Borel probability measure  $\mu_i^\infty$  on  $S \times T_{-i}$  such that

$$\text{marg}|_{\Omega_i^k} \mu_i^\infty = \mu_i^{k+1},$$

i.e., the mapping is canonical (Parthasarathy, 1978, Prop. 27.4). The last claim follows immediately by associating the belief  $\tilde{\mu}_i^k$  to the finite hierarchy  $t_i^k = (\mu_i^0, \dots, \mu_i^{k-1}, \tilde{\mu}_i^k) \in T_i^k$ .

$\square$

## A.4 Proof of Proposition 4.2

First consider the infinite hierarchies. Proposition 4.1 shows that each infinite belief hierarchy  $t_i^\infty = (\mu_i^0, \mu_i^1, \dots) \in \tilde{T}_i^\infty$  corresponds to a unique Borel probability measure on  $S \times T_{-i}$ , and the mapping is canonical. Denote the function that maps  $\tilde{T}_i^\infty$  to  $\Delta(S \times T_{-i})$  in this way by  $\psi_i^\infty$ . Conversely, let  $r_i^\infty : \Delta(S \times T_{-i}) \rightarrow \tilde{T}_i^\infty$  be the mapping that assigns to each  $\mu_i \in \Delta(S \times T_{-i})$  the hierarchy  $(\text{marg}|_S \mu_i, \text{marg}|_{\Omega_i^0} \mu_i, \text{marg}|_{\Omega_i^1} \mu_i, \dots) \in \Delta(S) \times \prod_{k \geq 0} \Delta(\Omega_i^k)$ . The function  $r_i^\infty$  is the inverse of  $\psi_i^\infty$ ; it remains to show that  $\psi_i^\infty$  and  $r_i^\infty$  are continuous. The function  $\psi_i^\infty$  is continuous if and only if  $t_i^n \rightarrow t_i$  in  $\tilde{T}_i^\infty$  implies  $\psi_i^\infty(t_i^n) \rightarrow \psi_i^\infty(t_i)$  in  $\Delta(S \times T_{-i})$ . But the cylinders form a convergence-determining class in  $S \times T_{-i}$  (Billingsley, 1999, Thm. 2.4), and the value of  $\psi_i^\infty(t_i)$  for  $t_i = (\mu_i^0, \mu_i^1, \dots)$  on the cylinders is given by the  $\mu_i^k$ 's. Finally, it follows from the continuity of the marginal operator that  $r_i^\infty$  is continuous.

For the case of finite hierarchies, simply set  $\psi_i^k(t_i^k) := \tilde{\mu}_i^k$  for each  $t_i^k = (\mu_i^0, \dots, \mu_i^{k-1}, \tilde{\mu}_i^k) \in T_i^k$ . Continuity of the mapping  $\psi_i^k$  is immediate.  $\square$

## A.5 Proof of Proposition 4.4

Let  $\mathcal{Q} = \langle (Q_i)_{i \in N}, (\beta_i^k)_{i \in N, k \in K^Q} \rangle$  be a type space that allows for finite-order reasoning. To construct a type morphism from the types in  $\mathcal{Q}$  to the types in the space  $\mathcal{T}$ , we first construct a collection of functions that maps each type into the associated hierarchy of beliefs (Step 1). Step 2 establishes that these mappings define a type morphism from  $\mathcal{Q}$  to  $\mathcal{T}$ . Step 3 then shows that this type morphism is unique.

### *Step 1: From types to belief hierarchies*

Each type induces a belief hierarchy of the kind discussed in Section 3.1, as we show now. The mapping from types to belief hierarchies is standard,<sup>8</sup> except that we need to take into account that hierarchies may be finite.

We define a collection of mappings. Lemma A.5 below shows that these functions are well-defined. For  $i \in N$ , if  $\tilde{Q}_i^0 \neq \emptyset$ , let  $h_i^{Q,0,0} : \tilde{Q}_i^0 \rightarrow \tilde{T}_i^0$  be the trivial mapping that assigns to each  $q_i \in \tilde{Q}_i^0$  the ‘naive’ type  $\tilde{t}_i^0$ , i.e.,  $h_i^{Q,0,0}(q_i) = \tilde{t}_i^0$ . Clearly,  $h_i^{Q,0,0}(\tilde{Q}_i^0) \subseteq \tilde{T}_i^0$ . Also,  $h_i^{Q,0,0}$  is measurable.

Similarly, define  $h_i^{Q,1,0} : \tilde{Q}_i^1 \rightarrow T_i^0$  to be the trivial mapping whenever  $\tilde{Q}_i^1$  is nonempty. Again, it is easy to see that  $h_i^{Q,1,0}(\tilde{Q}_i^1) \subseteq T_i^0$ , and that  $h_i^{Q,1,0}$  is measurable. If  $\tilde{Q}_i^0$  is nonempty, define the function  $h_i^{Q,<1,0} : \tilde{Q}_i^0 \rightarrow \tilde{T}_i^0$  by

$$h_i^{Q,<1,0}(q_i) := h_i^{Q,0,0}(q_i).$$

---

<sup>8</sup>See, for example, Mertens and Zamir (1985), and Heifetz and Samet (1998).

Again,  $h_i^{Q,<1,0}(\tilde{Q}_i^0) \subseteq \tilde{T}_i^0$ , and  $h_i^{Q,<1,0}$  is measurable. Finally, define the function  $h_i^{Q,1,1} : \tilde{Q}_i^1 \rightarrow \tilde{T}_i^1$  by

$$h_i^{Q,1,1}(q_i) := (h_i^{Q,1,0}(q_i), \beta_i^1(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<1,0})^{-1}),$$

where  $\text{Id}_S$  is the identity function on  $S$ . It is easy to verify that  $h_i^{Q,1,1}(\tilde{Q}_i^1) \subseteq \tilde{T}_i^1$ . Using that an image measure induced by a measurable function from a metrizable space into a metrizable space is measurable,<sup>9</sup> it follows directly that  $h_i^{Q,1,1}$  is measurable.

Fix  $k = 1, 2, \dots$ , and let  $\ell = 0, \dots, k-1$ . Suppose, inductively, that the mappings  $h_i^{Q,m,\ell}$  have been defined for  $m = 0, 1, \dots, k$  whenever the relevant domain is nonempty. If  $\tilde{Q}_i^{\leq k} = \bigcup_{m=0}^k \tilde{Q}_i^m \neq \emptyset$ , then define

$$h_i^{Q,<k+1,\ell} : \tilde{Q}_i^{\leq k} \rightarrow T_i^{\leq \ell}$$

by

$$\forall m = 0, 1, \dots, k, \quad q_i \in \tilde{Q}_i^m : \quad h_i^{Q,<k+1,\ell}(q_i) := \begin{cases} h_i^{Q,m,\ell}(q_i) & \text{if } m > \ell; \\ h_i^{Q,m,m}(q_i) & \text{if } m \leq \ell; \end{cases}$$

Also, for  $k > 0$ , let

$$h_i^{Q,<k+1,k} : \tilde{Q}_i^{\leq k} \rightarrow \tilde{T}_i^{\leq k}$$

be defined by

$$\forall m = 0, 1, \dots, k, \quad q_i \in \tilde{Q}_i^m : \quad h_i^{Q,<k+1,k}(q_i) := h_i^{Q,m,m}(q_i)$$

Then, if  $\tilde{Q}_i^{k+1} \neq \emptyset$ , let  $h_i^{Q,k+1,0} : \tilde{Q}_i^{k+1} \rightarrow T_i^0$  be the trivial mapping, as before, and for  $\ell = 1, \dots, k$ , define  $h_i^{Q,k+1,\ell} : \tilde{Q}_i^{k+1} \rightarrow T_i^\ell$  by

$$h_i^{Q,k+1,\ell}(q_i) := \left( h_i^{Q,k+1,\ell-1}(q_i), \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,\ell-1})^{-1} \right),$$

where  $\text{Id}_S$  is the identity function on  $S$ . Finally, define  $h_i^{Q,k+1,k+1} : \tilde{Q}_i^{k+1} \rightarrow \tilde{T}_i^{k+1}$  by

$$h_i^{Q,k+1,k+1}(q_i) := \left( h_i^{Q,k+1,k}(q_i), \beta_i^{k+1}(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k+1,k})^{-1} \right).$$

The next lemma states that these functions are well-defined:

**Lemma A.5.** *Let  $i \in N$  and  $k = 0, 1, \dots$*

(a) *If  $\tilde{Q}_i^k \neq \emptyset$ , then  $h_i^{Q,k,\ell}$  is well-defined and measurable for  $\ell = 0, 1, \dots, k$ .*

(b) *If  $\tilde{Q}_i^{\leq k} \neq \emptyset$ , then  $h_i^{Q,<k+1,\ell}$  is well-defined and measurable for  $\ell = 0, 1, \dots, k$ .*

---

<sup>9</sup>Given a measurable function  $f : X \rightarrow Y$ , the image measure  $\underline{f} : \Delta(X) \rightarrow \Delta(Y)$  induced by  $f$  is defined by  $\underline{f}(\mu) := \mu \circ f^{-1}$  for all  $\mu \in \Delta(X)$ . If  $Z$  is Polish, then the Borel  $\sigma$ -algebra on  $\Delta(Z)$  is generated by sets of the form  $\{\mu \in \Delta(Z) : \mu(E) \geq p\}$ , where  $E \in \mathcal{B}(Z)$  and  $p \in [0, 1]$ .

For  $i \in N$  and  $k < \infty$  such that  $\tilde{Q}_i^k$  is nonempty, define  $h_i^{Q,k} : \tilde{Q}_i^k \rightarrow \tilde{T}_i^k$  by:

$$h_i^{Q,k}(q_i) := \left( h_i^{Q,k,0}(q_i), \beta_i^k(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k,0})^{-1}, \beta_i^k(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k,1})^{-1}, \dots, \right. \\ \left. \beta_i^k(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k,k-1})^{-1} \right),$$

i.e.,  $h_i^{Q,k}(q_i)$  is the belief hierarchy (of depth  $k$ ) induced by  $q_i$ . It follows directly from the above that  $h_i^{Q,k}$  is well-defined and measurable.

We next define a collection of functions that will be used to obtain the belief hierarchies of infinite depth. For  $i \in N$ , if  $\tilde{Q}_i^\infty \neq \emptyset$ , let  $h_i^{Q,\infty,0} : Q_i^\infty \rightarrow T_i^0$  again be the trivial mapping. For  $\ell = 1, 2, \dots$ , assume that  $h_i^{Q,\infty,\ell-1} : Q_i^\infty \rightarrow T_i^{\ell-1}$  has been defined and is measurable. Define  $h_i^{Q,\leq\infty,\ell-1} : \tilde{Q}_i^\infty \cup \bigcup_{m=0}^\infty \tilde{Q}_i^m \rightarrow T_i^{\leq\ell-1}$  by

$$\forall m = \infty, 0, 1, \dots, q_i \in \tilde{Q}_i^m : \quad h_i^{Q,\leq\infty,\ell-1}(q_i) = \begin{cases} h_i^{Q,m,\ell-1}(q_i) & \text{if } m > \ell - 1; \\ h_i^{Q,m,m}(q_i) & \text{if } m \leq \ell - 1; \end{cases}$$

Also, define  $h_i^{Q,\infty,\ell} : \tilde{Q}_i^\infty \rightarrow T_i^\ell$  by

$$h_i^{Q,\infty,\ell}(q_i) := \left( h_i^{Q,\infty,\ell-1}(q_i), \beta_i^\infty(q_i) \circ (\text{Id}_S, h_{-i}^{Q,\leq\infty,\ell-1})^{-1} \right).$$

Again, these functions are well-defined:

**Lemma A.6.** *Let  $i \in N$ .*

(a) *If  $\tilde{Q}_i^\infty \neq \emptyset$ , then  $h_i^{Q,\infty,\ell}$  is well-defined and measurable for  $\ell = 0, 1, \dots$*

(b) *The function  $h_i^{Q,\leq\infty,\ell}$  is well-defined and measurable for  $\ell = 0, 1, \dots$*

The proof is similar to that of Lemma A.5, and thus omitted. Define  $h_i^{Q,\infty} : \tilde{Q}_i^\infty \rightarrow \tilde{T}_i^\infty$  by:

$$h_i^{Q,\infty}(q_i) := \left( h_i^{Q,\infty,0}(q_i), \beta_i^\infty(q_i) \circ (\text{Id}_S, h_{-i}^{Q,\leq\infty,0})^{-1}, \beta_i^\infty(q_i) \circ (\text{Id}_S, h_{-i}^{Q,\leq\infty,1})^{-1}, \dots \right).$$

That is,  $h_i^{Q,\infty}(q_i)$  is the belief hierarchy (of infinite depth) induced by  $q_i$ . By the above,  $h_i^{Q,\infty}$  is well-defined and measurable. Thus, each type generates a well-defined belief hierarchy.

We next define a type morphism from an arbitrary type space  $\mathcal{Q}$  to  $\mathcal{T}$ , using the mappings defined in Step 1.

*Step 2: Constructing a type morphism*

Recall that  $K_i^Q$  is the set of indices  $k \in \{0, 1, \dots\} \cup \{\infty\}$  such that  $\tilde{Q}_i^k$  is nonempty. For  $i \in N$ , define  $\varphi_i := (\varphi_i^k)_{k \in K_i^Q}$  as follows. If  $k \in K_i^Q$  is finite, then define  $\varphi_i^k : \tilde{Q}_i^k \rightarrow \tilde{T}_i^k$  by:

$$\varphi_i^k(q_i) := h_i^{Q,k}(q_i).$$

If  $Q_i^\infty$  is nonempty, then define  $\varphi_i^\infty : Q_i^\infty \rightarrow \tilde{T}_i^\infty$  by:

$$\varphi_i^\infty(q_i) := h_i^{Q,\infty}(q_i).$$

We show that  $\varphi = (\varphi_i)_{i \in N}$  is a type morphism. By Lemmas A.5 and A.6, the functions  $\varphi_i^k$ ,  $i \in N$ ,  $k \in K_i^Q$ , are well-defined and measurable. It remains to show that the mappings preserve higher-order beliefs.

To show this, let  $i \in N$  and suppose there is  $k < \infty$  such that  $\tilde{Q}_i^k \neq \emptyset$ . We need to show that for each  $q_i \in \tilde{Q}_i^k$  and  $E \in \mathcal{B}(S \times \tilde{T}_{-i}^{\leq k-1})$ ,

$$\psi_i^k(\varphi_i^k(q_i))(E) = \beta_i^k(q_i) \left( (\text{Id}_S, \varphi_i^{\leq k})^{-1}(E) \right).$$

Let  $q_i \in \tilde{Q}_i^k$ . Using that the belief mappings in  $\mathcal{T}$  are canonical, it follows that

$$\begin{aligned} \psi_i^k(\varphi_i^k(q_i))(E) &= \psi_i^k \left( h_i^{Q,k,0}(q_i), \beta_i^k \circ (\text{Id}_S, h_{-i}^{Q,<k,0})^{-1}, \dots, \beta_i^k \circ (\text{Id}_S, h_{-i}^{Q,<k,k-1})^{-1} \right)(E) \\ &= \beta_i^k(q_i) \left( (\text{Id}_S, h_{-i}^{Q,<k,k-1})^{-1}(E) \right). \end{aligned}$$

Next suppose that  $\tilde{Q}_i^\infty \neq \emptyset$ , and let  $q_i \in Q_i^\infty$ . We need to show that for each  $E \in \mathcal{B}(S \times T_{-i})$ ,

$$\psi_i^\infty(\varphi_i^\infty(q_i))(E) = \beta_i^\infty(q_i) \left( (\text{Id}_S, \varphi_i^\infty)^{-1}(E) \right).$$

Let  $q_i \in Q_i^\infty$ . Again using that the belief maps in  $\mathcal{T}$  are canonical, we have

$$\begin{aligned} \psi_i^\infty(\varphi_i^\infty(q_i))(E) &= \psi_i^\infty \left( h_i^{Q,\infty,0}(q_i), \beta_i^\infty \circ (\text{Id}_S, h_{-i}^{Q,\leq\infty,0})^{-1}, \dots \right)(E) \\ &= \beta_i^\infty(q_i) \left( (\text{Id}_S, h_{-i}^{Q,\infty})^{-1}(E) \right). \end{aligned}$$

It follows that  $\varphi$  is a type morphism.

*Step 3: There is a unique type morphism from any type space to  $\mathcal{T}$*

We show that for any type space  $\mathcal{Q}$ , there is a unique type morphism from  $\mathcal{Q}$  to  $\mathcal{T}$ . The proof uses the following lemmas. Lemma A.7 shows that type morphisms preserve belief hierarchies (cf. Heifetz and Samet, 1998, Prop. 5.1):

**Lemma A.7.** *Fix arbitrary type spaces  $\mathcal{Q}$  and  $\mathcal{R}$ , and let  $\varphi$  be a type morphism from  $\mathcal{Q}$  to  $\mathcal{R}$ . Then, for each  $i \in N$ ,*

- (a) *if  $\tilde{Q}_i^k \neq \emptyset$  for  $k < \infty$ , then  $h_i^{R,k}(\varphi_i^k(q_i)) = h_i^{Q,k}(q_i)$ ;*
- (b) *if  $\tilde{Q}_i^\infty \neq \emptyset$ , then  $h_i^{R,\infty}(\varphi_i^\infty(q_i)) = h_i^{Q,\infty}(q_i)$ .*

**Proof.** Here we show (a); the proof that (b) holds is similar. The claim clearly holds for  $k = 0$  for every player  $i$  such that  $\tilde{Q}_i^0 \neq \emptyset$ , as  $\tilde{T}_i^0$  is a singleton. Let  $k = 1, 2, \dots$ , and suppose the claim is true for all  $i \in N$  and  $m = 0, 1, \dots, k-1$  such that  $\tilde{Q}_i^m$  is nonempty. Again, for each  $i \in N$  such that  $\tilde{Q}_i^k \neq \emptyset$ , it is easy to see that  $h_i^{R,k,0}(\varphi_i^k(q_i)) = h_i^{Q,k,0}(q_i)$  for every  $q_i \in \tilde{Q}_i^k$ , where  $h_i^{R,k,0}$  is defined analogously to  $h_i^{Q,k,0}$  (recall that  $K_i^R \supseteq K_i^Q$ , so that  $h_i^{R,k,0}$  is well-defined). Let  $\ell = 1, \dots, k$  and suppose that

$$h_i^{R,k,m}(\varphi_i^k(q_i)) = h_i^{Q,k,m}(q_i)$$

for every  $q_i \in \tilde{Q}_i^k$  and  $m \leq \ell - 1$ . Denoting the belief maps for player  $i$  in  $\mathcal{R}$  by  $\lambda_i^k$ , where  $k \in K_i^R$ , we can use condition (4.1) to obtain

$$\begin{aligned} \lambda_i^k(\varphi_i^k(q_i)) \circ (\text{Id}_S, h_{-i}^{R,<k,\ell-1})^{-1} &= \beta_i^k(q_i) \circ (\text{Id}_S, \varphi_{-i}^{\leq k})^{-1} \circ (\text{Id}_S, h_{-i}^{R,<k,\ell-1})^{-1} \\ &= \beta_i^k(q_i) \circ (\text{Id}_S, h_{-i}^{R,<k,\ell-1} \circ \varphi_{-i}^{\leq k})^{-1} \\ &= \beta_i^k(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k,\ell-1})^{-1}, \end{aligned}$$

where the last line uses the induction hypothesis. Again using the induction hypothesis, this gives

$$\begin{aligned} h_i^{R,k,\ell}(\varphi_i^k(q_i)) &= (h_i^{R,k-1,\ell}(\varphi_i^k(q_i)), \lambda_i^k(\varphi_i^k(q_i)) \circ (\text{Id}_S, h_{-i}^{R,<k,\ell-1})^{-1}) \\ &= (h_i^{Q,k,\ell-1}(q_i), \beta_i^k(q_i) \circ (\text{Id}_S, h_{-i}^{Q,<k,\ell-1})^{-1}) \\ &= h_i^{Q,k,\ell}(q_i), \end{aligned}$$

for every  $q_i \in \tilde{Q}_i^k$ . □

**Lemma A.8.** *Let  $i \in N$  and  $k = 0, 1, \dots$  or  $k = \infty$ . Then  $h_i^{T,k} : \tilde{T}_i^k \rightarrow \tilde{T}_i^k$  is the identity function.*

The proof of Lemma A.8 follows directly from Propositions 4.1 and 4.2.

To show that  $\varphi$  is the unique type morphism from  $\mathcal{Q}$  to  $\mathcal{T}$ , suppose that  $\tilde{\varphi}$  is a type morphism from  $\mathcal{Q}$  to  $\mathcal{T}$ . Then, it follows from Lemma A.7 that for every  $i \in N$  and  $k \in K_i^Q$ ,

$$h_i^{T,k}(\tilde{\varphi}_i^k(q_i)) = h_i^{Q,k}(q_i).$$

But by Lemma A.8,

$$h_i^{T,k}(\tilde{\varphi}_i^k(q_i)) = \tilde{\varphi}_i^k(q_i),$$

so that  $\tilde{\varphi}_i^k(q_i) = h_i^{Q,k}(q_i)$ . The result then follows by noting that  $\varphi_i^k = h_i^k$ .

To summarize: Step 2 shows that for any type space  $\mathcal{Q}$ , there is a type morphism from  $\mathcal{Q}$  to  $\mathcal{T}$ , using the functions defined in Step 1. Step 3 shows that this type morphism is unique.

Hence,  $\mathcal{T}$  is universal. It remains to show that there is at most one universal space, up to type isomorphism (cf. Heifetz and Samet, 1998, Prop. 3.5).

To see this, suppose that  $\mathcal{T}$  and  $\mathcal{T}'$  are universal. Then, there is a type morphism  $\varphi$  from  $\mathcal{T}$  to  $\mathcal{T}'$  and a type morphism  $\varphi'$  from  $\mathcal{T}'$  to  $\mathcal{T}$ . Hence,  $\varphi' \circ \varphi$  is a type morphism from  $\mathcal{T}$  to  $\mathcal{T}$ . But the tuple  $\text{Id}_T := (\text{Id}_{\tilde{T}_i^k})_{k \in K_i^T}$ , where  $\text{Id}_{\tilde{T}_i^k}$  is the identity mapping on  $\tilde{T}_i^k$ , is also a type morphism from  $\mathcal{T}$  to  $\mathcal{T}$ . By uniqueness of the type morphisms, it follows that  $\varphi' \circ \varphi = \text{Id}_T$ . Similarly,  $\varphi \circ \varphi' = \text{Id}_{T'}$ . Hence,  $\varphi$  is an isomorphism.  $\square$

## A.6 Proof of Proposition 4.5

Let  $\mathcal{Q}$  be a type structure that allows for finite-order reasoning. We first prove the first claim. If  $h_j^Q$  is injective, then it is trivially countably uncountable, i.e., the set  $\{t_j \in T_j : (h_j^{Q,k})^{-1}(t_j) \text{ is uncountable}\}$  is countable (as it is empty) for every  $k$ . It then follows from Purves' theorem that  $h_j^{Q,k}$  is bimeasurable (Purves, 1966), i.e., for all  $B \in \mathcal{B}(\tilde{Q}_j^k)$ ,  $h_j^{Q,k}(B) \in \mathcal{B}(\tilde{T}_j^k)$ . In particular,

$$\{t_j \in T_j : t_j = \varphi_j^{\kappa(q_j)}(q_j) \text{ for some } q_j \in Q_j\} = \bigcup_{k \in K_j^Q} \varphi_j^k(\tilde{Q}_j^k) \in \mathcal{B}(T_j).$$

Hence,  $S \times \prod_{j \neq i} \{t_j \in T_j : t_j = \varphi_j^{\kappa(q_j)}(q_j) \text{ for some } q_j \in Q_j\}$  is indeed an event in  $\mathcal{B}(S \times T_{-i})$ . The result now follows directly from the definition of a type morphism.

The second claim follows directly by setting  $Q_i := T'_i$  and  $\beta_i^k := \psi_i^k$  for all  $i \in N$  and  $k \in K_i^Q$ .  $\square$

## A.7 Proof of Lemma 4.6

The result follows immediately by noting that the Borel  $\sigma$ -algebra on  $\Delta(S \times Q_{-i})$  is generated by sets of the form

$$E \in \mathcal{B}(S \times Q_{-i}), p \in [0, 1] : \quad \{\mu \in \Delta(S \times Q_{-i}) : \mu(E) \geq p\}.$$

## A.8 Proof of Proposition 4.7

Let  $\varphi$  be the type morphism from  $\mathcal{Z}$  to  $\mathcal{T}$ . Clearly,  $\varphi_i^\infty(z_i) \in \tilde{T}_i^\infty$  for all  $i \in N$  and  $z_i \in Z_i$ . Hence,

$$S \times \prod_{j \in N} \{t_j \in T_j : t_j = \varphi_j^{\kappa(z_j)}(z_j) \text{ for some } z_j \in Z_j\} \subseteq E^\infty.$$

The type structure  $\mathcal{Z}$  is nonredundant by construction. Hence, by Proposition 4.5,

$$\psi_i(\varphi_i^\infty(z_i)) \left( S \times \prod_{j \in N} \{t_j \in T_j : t_j = \varphi_j^{\kappa(z_j)}(z_j) \text{ for some } z_j \in Z_j\} \right) = 1$$

for each  $i \in N$  and  $z_i \in Z_i$ , so that

$$S \times \prod_{j \in N} \{t_j \in T_j : t_j = \varphi_j^{\kappa(z_j)}(z_j) \text{ for some } z_j \in Z_j\} \subseteq CB(E^\infty).$$

To prove the reverse inclusion, it is sufficient to show that for each  $i \in N$ , there is  $Y_i^\infty \subseteq Z_i^\infty$  such that

$$\varphi_i^\infty(Y_i^\infty) = \pi_{T_i}^{S \times T}(E^\infty \cap CB(E^\infty)).$$

To show this, we construct a map  $\hat{f}$  from  $E^\infty \cap CB(E^\infty)$  to  $S \times \hat{Z}$ . First note that  $E^\infty \cap CB(E^\infty) \subseteq S \times \tilde{T}^\infty$ . For  $i \in N$  and  $(\mu_i^0, \mu_i^1, \dots) \in E^\infty \cap CB(E^\infty)$ , let  $\hat{f}_i^0(\mu_i^0) := z_i^0$ . For  $k = 1, 2, \dots$ , suppose  $f_j^{k-1} : \pi_{T_j^{k-1}}^{S \times T}(E^\infty \cap CB(E^\infty)) \rightarrow \hat{Z}_j^{k-1}$  has been defined for all  $j \in N$ . For  $i \in N$  and  $(\mu_i^0, \mu_i^1, \dots) \in E^\infty \cap CB(E^\infty)$ , define

$$\hat{f}_i^k(\mu_i^0, \dots, \mu_i^k) := (\hat{f}_i^{k-1}(\mu_i^0, \dots, \mu_i^{k-1}), \mu_i^k \circ (\text{Id}_S, \hat{f}_{-i}^{k-1})^{-1}).$$

It is easy to check that  $\hat{f}_i^k$  is well-defined, given that the beliefs specified by the belief hierarchies in  $E^\infty \cap CB(E^\infty)$  are coherent. Then, for each  $(s, t) \in E^\infty \cap CB(E^\infty)$ , with  $t_i = (\mu_i^0, \mu_i^1, \dots)$  for  $i \in N$ , define

$$\hat{f}(s, t) := (s, (\hat{f}_i^0(\mu_i^0), \mu_i^1 \circ (\text{Id}_S, \hat{f}_{-i}^0)^{-1}, \dots)_{i \in N}).$$

Again, it is easy to verify that  $\hat{f}(E^\infty \cap CB(E^\infty)) \subseteq S \times \hat{Z}$ , so that the set  $\pi_{\hat{Z}_i}^{S \times \hat{Z}}(E^\infty \cap CB(E^\infty))$  corresponds to a subset  $Y_i^\infty$  of  $Z_i^\infty$ . Given that there is a unique type morphism  $\varphi$  from  $\mathcal{Z}$  to  $\mathcal{T}$ , it must be the case that  $\varphi_i^\infty(Y_i^\infty) = \pi_{T_i}^{S \times T}(E^\infty \cap CB(E^\infty))$ , and the result follows.  $\square$

## A.9 Proof of Lemma 6.1

Since

$$\frac{\partial v}{\partial z} = 1 - \frac{\sqrt{2}\varepsilon}{\sqrt{2\varepsilon^2 + 2\varepsilon + z}} \in (0, 1)$$

and  $v(\frac{1}{2}; \varepsilon) = \frac{1}{2}$ , it follows that  $v(0; \varepsilon) < \frac{1}{2}$  and that

$$v^k(0; \varepsilon) \nearrow_{k \rightarrow \infty} \frac{1}{2}.$$

On the other hand,

$$v(0; \varepsilon) = (2\varepsilon + 4\varepsilon^2) - 2\varepsilon\sqrt{2(2\varepsilon + 2\varepsilon^2)} \xrightarrow{\varepsilon \rightarrow 0} 0.$$

Inductively, if we have already shown that

$$v^{k-1}(0; \varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 0$$

then it follows that

$$v^k(0; \varepsilon) = v(v^{k-1}(0; \varepsilon); \varepsilon) = (v^{k-1}(0; \varepsilon) + 2\varepsilon + 4\varepsilon^2) - 2\varepsilon\sqrt{2(v^{k-1}(0; \varepsilon) + 2\varepsilon + 2\varepsilon^2)} \xrightarrow{\varepsilon \rightarrow 0} 0$$

as well.  $\square$



## A.10 Proof of Proposition 6.4

Fix  $p < 1$ , and let  $t_i$  be a type for player  $i$  that has an infinite depth of reasoning and assigns probability  $p^k(1-p)$  to the event that the other player  $-i$ 's type is in  $\bar{G}_{-i}^k$ , where  $k = 0, 1, 2, \dots$ . There is  $k^*$  such that  $t_i$  assigns probability less than  $\min\{\frac{s}{4}, \frac{1-s}{4}\}$  to types in  $\bigcup_{k > k^*} \bar{G}_{-i}^k$ .

Given any  $\varepsilon_{k^*}(s)$  as defined in Proposition 6.3, let  $\varepsilon_p(s) := \min\{\varepsilon_{k^*}(s), \frac{s}{4}, \frac{1-s}{4}\}$ . Then, if  $\varepsilon < \varepsilon_p(s)$ , both investing and not investing are rationalizable for types in  $\bigcup_{k \leq k^*} \bar{G}_{-i}^k$ . Since  $\varepsilon_p(s) \leq \min\{\frac{s}{4}, \frac{1-s}{4}\}$ , both players' signals are confined to the interval  $[s - \frac{s}{4}, s + \frac{1-s}{4}]$ . Therefore, both players assign probability 1 to the event that the state of nature is in the interval  $[s - \frac{s}{2}, s + \frac{1-s}{2}]$ . But given that the probability that  $t_i$  assigns to the types in  $\bigcup_{k > k^*} \bar{G}_{-i}^k$  (who may have a unique rationalizable action) is less than  $\min\{\frac{s}{4}, \frac{1-s}{4}\}$ , both actions are rationalizable for  $t_i$ .  $\square$

## A.11 Proof of Proposition 6.5

The proof is analogous to the proof of Proposition 6.4. Fix  $r < 1$ , and let  $t_i$  be the type of player  $i$  that assigns probability  $r^k(1-r)$  to the event that the other player  $-i$ 's type is  $t_{m,-i}$ , where  $m = 0, 1, 2, \dots$ . There is  $m^*$  such that  $t_i$  assigns probability less than  $\min\{\frac{s}{4}, \frac{1-s}{4}\}$  to types in  $\bigcup_{m > m^*} \{t_{m,-i}\}$ .

Given any  $\varepsilon_{p^*}(s)$  as defined in Proposition 6.4, let  $\varepsilon_r(s) := \min\{\varepsilon_{p^*}(s), \frac{s}{4}, \frac{1-s}{4}\}$ . Then, if  $\varepsilon < \varepsilon_r(s)$ , both investing and not investing are rationalizable for types in  $\bigcup_{m \leq m^*} \{t_{m,-i}\}$ . Since  $\varepsilon_r(s) \leq \min\{\frac{s}{4}, \frac{1-s}{4}\}$ , both players' signals are confined to the interval  $[s - \frac{s}{4}, s + \frac{1-s}{4}]$ . Therefore, both players assign probability 1 to the event that the state of nature is in the interval  $[s - \frac{s}{2}, s + \frac{1-s}{2}]$ . But given that the probability that  $t_i$  assigns to the types in  $\bigcup_{m > m^*} \{t_{m,-i}\}$  (who may have a unique rationalizable action) is less than  $\min\{\frac{s}{4}, \frac{1-s}{4}\}$ , both actions are rationalizable for  $t_i$ .  $\square$

## References

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

- Billingsley, P. (1999). *Convergence of Probability Measures* (2nd ed.). Wiley Series in Probability and Statistics. New York: Wiley.
- Bosch-Domenech, A., R. Nagel, A. Satorra, and J. García-Montalvo (2002). One, two, (three), infinity: A comparison of behavior of newspaper readers, game theorists, and students in beauty-contest experiments. *American Economic Review* 92, 1687–1701.
- Brandenburger, A. and E. Dekel (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory* 59, 189–198.
- Cabrales, A., R. Nagel, and R. Armenter (2007). Equilibrium selection through incomplete information in coordination games: An experimental study. *Experimental Economics* 10, 221–234.
- Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics* 119(3), 861–898.
- Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica* 61, 989–1018.
- Chen, Y.-C., A. Di Tillio, E. Faingold, and S. Xiong (2010). Uniform topologies on types. *Theoretical Economics* 5, 445–478.
- 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 S. Morris (2007). Interim correlated rationalizability. *Theoretical Economics* 2, 15–40.
- Harsanyi, J. (1967–1968). Games on incomplete information played by Bayesian players. Parts I, II, III. *Management Science* 14, 159–182, 320–334, 486–502.
- Heifetz, A. (1993). The Bayesian formulation of incomplete information—The non-compact case. *International Journal of Game Theory* 21, 329–338.

- Heifetz, A. and D. Samet (1998). Topology-free typology of beliefs. *Journal of Economic Theory* 82, 324–341.
- Heinemann, F., R. Nagel, and P. Ockenfels (2004). The theory of global games on test: Experimental analysis of coordination games with public and private information. *Econometrica* 72, 1583–1599.
- Heinemann, F., R. Nagel, and P. Ockenfels (2009). Measuring strategic uncertainty in coordination games. *Review of Economic Studies* 76, 181–221.
- Heinsalu, S. (2011). Universal type spaces with unawareness. Mimeo, Yale University.
- 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.
- Hocking, J. G. and G. S. Young (1988). *Topology*. Mineola, NY: Dover Publications.
- Kechris, A. S. (1995). *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Berlin: Springer-Verlag.
- Kets, W. (2010). Bounded reasoning and higher-order uncertainty. Working paper, Northwestern University.
- Kets, W. (2012). Bounded reasoning: A canonical type space. Mimeo.
- 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. and H. S. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances in economics and econometrics: Eighth World Congress*, Volume 1. Cambridge University Press.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review* 85, 1313–1326.
- Parthasarathy, K. R. (1978). *Introduction to probability and measure*. New York: Springer-Verlag.
- Pintér, M. and Z. Udvari (2011). Generalized type spaces. Working paper 34107, MPRA.

- Purves, R. (1966). Bimeasurable functions. *Fundamenta Mathematicae* 58, 149–157.
- Rubinstein, A. (1989). The electronic mail game: Strategic behavior under “almost common knowledge”. *American Economic Review* 79, 385–391.
- 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. mimeo, Harvard University.