Games and Economic Behavior 104 (2017) 309–328

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

A new epistemic characterization of ε-proper rationalizability ☆

Andrés Perea [a,*], Souvik Roy [b]

[a] Maastricht University, Netherlands
[b] Indian Statistical Institute, India

ARTICLE INFO

Article history:
Received 12 August 2010
Available online 8 May 2017

JEL classification:
C72

Keywords:
Epistemic game theory
Incomplete information
Proper rationalizability

ABSTRACT

For a given $\varepsilon > 0$, the concept of *ε-proper rationalizability* (Schuhmacher, 1999) is based on two assumptions: (1) every player is *cautious*, i.e., does not exclude any opponent's choice from consideration, and (2) every player satisfies the *ε-proper trembling condition*, i.e., the probability he assigns to an opponent's choice a is at most ε times the probability he assigns to b whenever he believes the opponent to prefer b to a. In this paper we show that a belief hierarchy is ε-properly rationalizable in the *complete* information framework, if and only if, there is an equivalent belief hierarchy within the *incomplete* information framework that expresses common belief in the events that (1) players are cautious, (2) the players' beliefs about the opponent's utilities are "centered around the original utilities" in some specific way parametrized by ε, and (3) players rationalize each opponent's choice by a utility function that is as close as possible to the original utility function.

1. Introduction

Epistemic game theory deals with possible ways a player may reason about his opponents before making a decision. More precisely, in epistemic game theory players base their choices on the beliefs about the opponents' behavior, which in turn depend on their beliefs about the opponents' beliefs about others' behavior, and so on. A major goal of epistemic game theory is to study such infinite belief hierarchies, to impose reasonable conditions on these, and to investigate their behavioral implications. See Perea (2012) for a textbook that discusses these issues.

A central idea in epistemic game theory is *common belief in rationality* (Tan and Werlang, 1988), stating that a player believes that his opponents choose rationally, believes that his opponents believe that their opponents choose rationally, and so on. In our view, one of its most natural refinements is the concept of *proper rationalizability* (Schuhmacher, 1999 and Asheim, 2001), which is based on Myerson's (1978) notion of *proper equilibrium*, but without making any equilibrium assumption. Proper rationalizability is based on the following two conditions: The first states that players are *cautious*, meaning that they do not exclude any opponents' choices from consideration. The second condition is known as the *ε-proper trembling condition*, which states that whenever you believe that a choice a is better than another choice b for your opponent, then the probability you assign to b must be at most ε times the probability you assign to a.

☆ We would like to thank an associate editor and two anonymous referees for their valuable comments. In fact, their suggestions have led to a very drastic revision of the first version of our paper.

* Corresponding author.
E-mail address: a.perea@maastrichtuniversity.nl (A. Perea).

http://dx.doi.org/10.1016/j.geb.2017.04.009
0899-8256/

	d	e	f
a	0, 2	1, 1	1, 0
b	1, 2	0, 1	1, 0
c	1, 2	1, 1	0, 0

Fig. 1. An example for ε-proper rationalizability.

Formally, a type within an epistemic model is called *ε-properly rationalizable* if it is cautious, satisfies the ε-proper trembling condition, and expresses *common belief* in these events. The concept of *proper rationalizability* is basically obtained by considering the notion of ε-proper rationalizability, and letting ε tend to zero in the limit. In this paper, however, we will focus on the concept of ε-proper rationalizability for an arbitrary but fixed ε.

We will now explain this concept by means of an example. Consider the game in Fig. 1, where player 1 chooses between a, b and c and player 2 chooses between d, e and f.

Note that for player 2, choice d is better than choice e, and choice e is better than choice f. Hence, any ε-properly rationalizable type of player 1 assigns a probability to choice f that is at most ε times the probability he assigns to choice e, and assigns a probability to e that is at most ε times the probability he assigns to choice d. Therefore, if $\varepsilon < 1$, only choice c for player 1 and choice d for player 2 can be optimal for ε-properly rationalizable types. The usual interpretation of an ε-properly rationalizable type is that you assume that your opponent makes mistakes, but that you deem more costly mistakes much less likely (by a factor ε) than less costly mistakes.

In this paper we offer a rather different foundation for ε-properly rationalizable types, within a model that assumes *incomplete information* about the players' utility functions. More precisely, in our model players do not believe the opponent to make mistakes, but rather believe, with some (small) positive probability, that the opponent has a utility function *different* from the one depicted in the game. Importantly, they *do* believe the opponent to choose rationally given his utility function. In other words, instead of assuming that players are certain about the opponent's utility function but uncertain about his degree of rationality, we assume that players are uncertain about the opponent's utility function but certain about his degree of rationality.

Within such a framework we offer epistemic conditions that *characterize* the concept of ε-proper rationalizability for two-player games. The epistemic conditions we introduce are *caution, ε-centered beliefs around u*, and *belief in the opponent's rationality under the closest utility function.* Here, u is the original profile of utility functions as depicted in the game. By *caution* we mean that whenever a type deems some opponent's belief hierarchy possible, it does not exclude any opponent's choice for that particular belief hierarchy. A type has *ε-centered beliefs around u* if it deems opponent's utility functions closer to u much more likely (by a factor ε) than utility functions that are further away from u. Finally, a type *believes in the opponent's rationality under the closest utility function* if it rationalizes every opponent's choice by an opponent's utility function that is as close as possible to u.

Our main result shows that a type in a model with complete information is ε-properly rationalizable, if and only if, there is a type in the model with incomplete information that (a) expresses common belief in *caution, ε-centered beliefs around u* and *rationality under the closest utility function,* and (b) generates the same belief hierarchy on choices as the ε-properly rationalizable type. By doing so, we thus provide a new, alternative epistemic characterization of ε-proper rationalizability within an incomplete information setting.

The crucial difference with the usual interpretation of ε-proper rationalizability is that in our new characterization, a player believes – with probability 1 – that his opponent chooses rationally, but faces some small uncertainty about the opponent's true utility function, restricted by the *ε-centered beliefs* condition. In the usual interpretation, a player assigns probability 1 to his opponent's actual utility function, but faces some small uncertainty about the opponent's degree of rationality, restricted by the ε-proper trembling condition. Hence, we show that the new conditions of *ε-centered beliefs around u* and *belief in the opponent's rationality under the closest utility function* in the incomplete information setting are, in a sense, equivalent to the ε-proper trembling condition in the complete information setting.

The driving force behind this equivalence is our Choice Ranking Lemma (Lemma 5.5). There we show that for a given belief and a given utility function u_i, choice a is better than choice b if and only if the closest utility function to u_i needed to rationalize a is closer to u_i than the closest utility function to u_i needed to rationalize b. Hence, "inferior choices" in the complete information setting can be translated into "more distant utility functions needed to rationalize the choice" in the incomplete information setting. Building upon this insight, the ε-proper trembling condition, stating that inferior choices must deemed less probable, by a factor ε, than better choices, becomes equivalent to saying that (a) choices must be rationalized by utility functions that are as close as possible to the original one, and (b) more distant utility functions must be deemed less probable, by a factor ε, than closer utility functions. Since the conditions (a) and (b) summarize *ε-centered beliefs around u* and *belief in the opponent's rationality under the closest utility function,* these conditions can thus be seen as an "incomplete information counterpart" to the ε-proper trembling condition. This, in fact, is the main message of our paper.

Our setting with incomplete information is related to the model used in Dekel and Fudenberg (1990). They also consider games with incomplete information where the players face some small uncertainty about the opponent's utilities. One important difference with our approach is that Dekel and Fudenberg apply the concept of *iterated elimination of weakly dominated choices* to such games with incomplete information. They show that if the uncertainty about the opponent's utilities vanishes, then we obtain one round of deletion of weakly dominated strategies, followed by iterated deletion of strongly dominated strategies, in the original game. The latter procedure is also called the Dekel–Fudenberg procedure in the literature. In contrast, we apply common belief in *caution*, *ε-centered beliefs around u* and *rationality under the closest utility function* to games with incomplete information. We then show that these types are in one-to-one correspondence with the ε-properly rationalizable types.

Another fundamental difference between our paper and Dekel–Fudenberg lies in the restrictions imposed on the uncertainty about the opponent's utilities. Their model assumes that players only deem possible finitely many utility functions for the opponent, and that a large probability must be assigned to the opponent's original utility function u. Our condition of *ε-centered beliefs around u* also imposes these conditions, but additionally requires that utility functions closer to u must be deemed much more likely than utility functions further away from u – something that is not required in the Dekel–Fudenberg setting.

The paper is organized as follows. In Section 2 we introduce the notion of ε-properly rationalizable types for games with complete information. In Section 3 we introduce an epistemic model for games with incomplete information, and define common belief in *caution*, *ε-centered beliefs around u* and *rationality under the closest utility function* within that setting. In Section 4 we show how to derive, for a given type, the full belief hierarchy on choices it induces. We do so for the complete information setting and the incomplete information setting. In Section 5 we state some preparatory results that are needed to prove our characterization result. In Section 6 we present and prove our epistemic characterization of ε-proper rationalizability. We give some concluding remarks in Section 7. Finally, Section 8 contains the proofs of the preparatory results.

2. ε-Proper rationalizability

The concept of *proper rationalizability* has first been defined by Schuhmacher (1999), and has later been characterized in Asheim (2001) within a model with *lexicographic beliefs.* Here we will follow Schuhmacher's approach, who developed proper rationalizability by first defining ε-proper rationalizability for an arbitrary $\varepsilon > 0$, and then "taking the limit when ε tends to 0". In fact, in this paper we will focus on the concept of ε-proper rationalizability for a fixed but arbitrary ε with $0 < \varepsilon < 1$. Throughout the paper we will restrict our attention to the case of two players to keep our presentation as simple as possible. Everything we do in this paper can easily be generalized, however, to the case of more than two players.

2.1. Epistemic model

Consider a finite two-player static game $\Gamma = (C_i, u_i)_{i\in I}$ where $I = \{1, 2\}$ is the set of players, C_i is the finite set of choices for player i, and $u_i : C_1 \times C_2 \to \mathbb{R}$ is player i's utility function. We assume that player i holds a probabilistic belief about j's choices, a probabilistic belief about the possible probabilistic beliefs that j can hold about i's choices, and so on. Such belief hierarchies can be encoded within an epistemic model with types.

Definition 2.1 *(Epistemic model).* Consider a finite two-player static game $\Gamma = (C_i, u_i)_{i\in I}$. A finite epistemic model for Γ is a tuple $M^{co} = (T_i, b_i)_{i\in I}$ where

(a) T_i is a finite set of types, and

(b) b_i is a mapping that assigns to every $t_i \in T_i$ a probabilistic belief $b_i(t_i) \in \Delta(C_j \times T_j)$ on the opponent's choice-type pairs.

Here, the superscript *co* stands for "complete information", as to distinguish it from the epistemic model for games with *incomplete information* which will be introduced in Section 3. For every finite set X, we denote by $\Delta(X)$ the set of probability distributions on X. In Section 4 we show how to formally derive a full belief hierarchy for every type.

2.2. ε-Proper rationalizability

Consider a finite two-player static game $\Gamma = (C_i, u_i)_{i\in I}$, and a finite epistemic model $M^{co} = (T_i, b_i)_{i\in I}$. Fix a type $t_i \in T_i$ with belief $b_i(t_i) \in \Delta(C_j \times T_j)$.

Type t_i *deems possible* a type $t_j \in T_j$ if $b_i(t_i)(C_j \times \{t_j\}) > 0$. Let $T_j(t_i)$ be the set of types $t_j \in T_j$ that t_i deems possible.

Type t_i is *cautious* if for every $t_j \in T_j(t_i)$, and every $c_j \in C_j$, we have that $b_i(t_i)(c_j, t_j) > 0$. That is, type t_i takes into account all opponent's choices for every opponent's belief hierarchy he deems possible.

For every choice $c_i \in C_i$, let

$$u_i(c_i, t_i) := \sum_{(c_j,t_j)\in C_j\times T_j} b_i(t_i)(c_j, t_j) \cdot u_i(c_i, c_j)$$

be the expected utility for player i induced by the choice c_i and the probabilistic belief $b_i(t_i)$ on $C_j \times T_j$. Type t_i *prefers* choice c_i to choice c_i' if $u_i(c_i, t_i) > u_i(c_i', t_i)$.

Fix a number ε with $0 < \varepsilon < 1$. Type t_i satisfies the *ε-proper trembling condition* if for every $t_j \in T_j(t_i)$, and every $c_j, c_j' \in C_j$ with $u_j(c_j', t_j) < u_j(c_j, t_j)$, we have that

$$b_i(t_i)(c_j', t_j) \le \varepsilon \cdot b_i(t_i)(c_j, t_j).$$

That is, t_i deems inferior choices much less likely than superior choices for the opponent.

In words, we say that type t_i is *ε-properly rationalizable* if it expresses common belief in "caution and ε-proper trembling". To formally define this, let us first define the set of types $T^*(t_i)$ that "t_i reasons about". We recursively define sets $T_i^1(t_i), T_j^1(t_i), T_i^2(t_i), T_j^2(t_i), \ldots$ as follows:

$$T_i^1(t_i) := \{t_i\},$$
$$T_j^1(t_i) := T_j(t_i),$$
$$T_i^k(t_i) := \{t_i' \in T_i \mid t_i' \in T_i(t_j') \text{ for some } t_j' \in T_j^{k-1}(t_i)\},$$
$$T_j^k(t_i) := \{t_j' \in T_j \mid t_j' \in T_j(t_i') \text{ for some } t_i' \in T_i^k(t_i)\},$$

for every $k \geq 2$. Then, we define

$$T^*(t_i) := \bigcup_{k \in \mathbb{N}} \left[T_i^k(t_i) \cup T_j^k(t_i) \right],$$

representing the set of types that "t_i reasons about".

Definition 2.2 *(ε-Proper rationalizability).* Type t_i is ε-properly rationalizable if every type in $T^*(t_i)$ is cautious and satisfies the ε-proper trembling condition.

That is, type t_i is cautious and satisfies the ε-proper trembling condition, only deems possible types for j that are cautious and satisfy the ε-proper trembling condition, only deems possible types for j that only deem possible types for i that are cautious and satisfy the ε-proper trembling condition, and so on, *ad infinitum.*

3. Incomplete information

We will now propose an epistemic model for situations in which players are *uncertain* about the opponent's utility function, and define the conditions of *caution, ε-centered beliefs around u*, and *belief in opponent's rationality under closest utility function* within that framework.

3.1. Epistemic model

Consider a finite two-player static game form $G = (C_i)_{i \in I}$. That is, we only specify the choice sets, but not the utility functions, for the players. Suppose now that both players are *uncertain* about the opponent's utility function, that is, the game is with *incomplete information.* A belief hierarchy for a player must now also specify what this player believes about the opponent's utility function, what this player believes about the opponent's belief about his own utility function, and so on. Also such belief hierarchies can be encoded within an epistemic model with types, as we will see. To formally define this epistemic model, let us denote by V_i the set of all possible utility functions $v_i : C_1 \times C_2 \to \mathbb{R}$.

Definition 3.1 *(Epistemic model with incomplete information).* Consider a finite two-player static game form $G = (C_i)_{i \in I}$. A finite epistemic model for G with incomplete information is a tuple $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ where

(a) Θ_i is a finite set of types,

(b) w_i is a mapping that assigns to every $\theta_i \in \Theta_i$ a utility function $w_i(\theta_i) \in V_i$, and

(c) β_i is a mapping that assigns to every $\theta_i \in \Theta_i$ a probabilistic belief $\beta_i(\theta_i) \in \Delta(C_j \times \Theta_j)$.

Here, the superscript *in* stands for "incomplete information". As every type θ_i holds a belief about j's type, and each of j's types θ_j has a utility function $w_j(\theta_j)$, we can derive for every type θ_i the induced belief about j's utility function. In fact, for every type we can derive a full belief hierarchy on the players' choices *and utility functions.* In Section 4 we show how to derive, for every type $\theta_i \in \Theta_i$, an infinite belief hierarchy on the players' choices alone.

3.2. Caution

Consider an epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for the game form G, and a type $\theta_i \in \Theta_i$. Similarly to the previous section, we say that type θ_i is cautious if, for every opponent's belief hierarchy it takes into account, it deems possible every opponent's choice. To formally define this, we need some additional notation. For a given type $\theta_i \in \Theta_i$, we say that θ_i deems possible some type $\theta_j \in \Theta_j$ if $\beta_i(\theta_i)(C_j \times \{\theta_j\}) > 0$. We denote by $\Theta_j(\theta_i)$ the set of types that θ_i deems possible. For a given type $\theta_j \in \Theta_j$ and utility function v_j, let $\theta_j^{v_j}$ be the auxiliary type that has utility function v_j and holds exactly the same belief on $C_i \times \Theta_i$ as θ_j. Consequently, $\theta_j^{v_j}$ has exactly the same belief hierarchy on choice-utility pairs as θ_j, but differs only in the utility function. Formally, θ_i is *cautious* if, for every $\theta_j \in \Theta_j(\theta_i)$, and for every $c_j \in C_j$, there is some utility function $v_j \in V_j$ such that $\beta_i(\theta_i)(c_j, \theta_j^{v_j}) > 0$.

3.3. ε-Centered beliefs around u

Consider a game form $G = (C_i)_{i \in I}$ and an epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G. We will now put restrictions on the beliefs that types hold about the opponent's *utility function.*

Consider a pair $u = (u_i)_{i \in I}$ of utility functions, and some ε with $0 < \varepsilon < 1$. Informally, we say that a type θ_i has "ε-centered beliefs around u" if it deems an opponent's utility function v_j "much more likely" than some other utility function v'_j – where "much more likely" is measured by ε – whenever v_j is closer to u_j than v'_j is. To formally define this, we first need to define the distance between some utility function v_j and the true utility function u_j.

Consider the true utility function u_j for player j, and some other utility function v_j. We define the *distance* $d(v_j, u_j)$ by

$$d(v_j, u_j) := \left[\sum_{(c_1, c_2) \in C_1 \times C_2} \big(v_j(c_1, c_2) - u_j(c_1, c_2) \big)^2 \right]^{1/2}.$$

Mathematically, this is just the Euclidean distance between the real valued vectors v_j and u_j.

Definition 3.2 *(ε-Centered beliefs around u).* Consider a static game form $G = (C_i)_{i \in I}$, an epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G, and a pair $u = (u_i)_{i \in I}$ of utility functions. A type θ_i has ε-centered beliefs around u if for every $\theta_j \in \Theta_j(\theta_i)$, every $c_j, c'_j \in C_j$ and every $v_j, v'_j \in V_j$ with $\beta_i(\theta_i)(c_j, \theta_j^{v_j}) > 0$ and $\beta_i(\theta_i)(c'_j, \theta_j^{v'_j}) > 0$, it holds that

$$\beta_i(\theta_i)(c'_j, \theta_j^{v'_j}) \le \varepsilon \cdot \beta_i(\theta_i)(c_j, \theta_j^{v_j})$$

whenever $d(v_j, u_j) < d(v'_j, u_j)$.

Hence, θ_i must deem, for every fixed opponent's belief hierarchy $\theta_j \in \Theta_j(\theta_i)$, utility functions closer to u_j much more likely than those that are further away from u_j.

Note that the definition of "ε-centered beliefs around u" depends crucially on the specific distance function d. However, for establishing the epistemic characterization of ε-proper rationalizability – which is the main purpose of this paper – we could also have chosen a whole range of other distance functions instead. More specifically, we could have chosen any distance function d with the following three properties:

(a) there is a norm $\|\cdot\|$ such that $d(v_j, u_j) = \|v_j - u_j\|$ for all $v_j \in V_j$,

(b) $\|v_j\| = \|v'_j\|$ whenever v'_j can be obtained from v_j by a permutation of the coordinates, and

(c) $\left\| \frac{1}{2} v_j + \frac{1}{2} v'_j \right\| < \|v_j\|$ whenever $v_j \neq v'_j$ and $\|v_j\| = \|v'_j\|$.

But, to keep things as transparent as possible, we have chosen a particular, well-known distance function that satisfies these properties – the Euclidean distance.

Note also that the distance measure above is not invariant with respect to positive affine transformations of the utility functions. In Section 7 we discuss this conceptual issue in some more detail.

3.4. Belief in rationality under closest utility function

We next impose that a type, for a given opponent's belief hierarchy and choice, must always look for the opponent's utility function *closest to* u_j for which that choice is optimal. We say that the type *believes in the opponent's rationality under the closest utility function.*

Consider a type θ_i with utility function $w_i(\theta_i)$ and belief $\beta_i(\theta_i) \in \Delta(C_j \times \Theta_j)$. For every choice $c_i \in C_i$, let

$$w_i(\theta_i)(c_i, \theta_i) := \sum_{(c_j, \theta_j) \in C_j \times \Theta_j} \beta_i(\theta_i)(c_j, \theta_j) \cdot w_i(\theta_i)(c_i, c_j)$$

be the expected utility induced by the choice c_i, the belief $\beta_i(\theta_i)$, and the utility function $w_i(\theta_i)$. We say that choice c_i is optimal for θ_i if

$$w_i(\theta_i)(c_i, \theta_i) \ge w_i(\theta_i)(c'_i, \theta_i) \text{ for all } c'_i \in C_i.$$

We are now ready to formalize the condition described above.

Definition 3.3 *(Belief in rationality under closest utility function).* Consider a static game form $G = (C_i)_{i \in I}$, an epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G, and a pair $u = (u_i)_{i \in I}$ of utility functions. A type θ_i believes in j's rationality under the closest utility function, if $\beta_i(\theta_i)$ only assigns positive probability to pairs (c_j, θ_j) where

(a) c_j is optimal for θ_j, and

(b) there is no $v_j \in V_j$ with $d(v_j, u_j) < d(w_j(\theta_j), u_j)$ such that c_j is optimal for $\theta_j^{v_j}$.

Hence, θ_i always looks for the utility function closest to u_j that rationalizes the choice c_j, for any given opponent's belief hierarchy θ_j and opponent's choice c_j. In Section 5 we will show that there is always a *unique* utility function v_j that rationalizes the choice c_j and that is closest to u_j.

3.5. Common belief in caution, centered beliefs and rationality

Consider a game form $G = (C_i)_{i \in I}$, an epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G, and a pair $u = (u_i)_{i \in I}$ of utility functions. We will focus on types $\theta_i \in \Theta_i$ that are not only cautious, hold ε-centered beliefs around u, and believe in the opponent's rationality under the closest utility function, but also express *common belief* in these three events. That is, types $\theta_i \in \Theta_i$ that also believe that j is cautious, that j has ε-centered beliefs around u, and that j believes in i's rationality under the closest utility function, and so on.

Similarly to the previous section, let $\Theta^*(\theta_i)$ be the set of types that θ_i reasons about.

Definition 3.4 *(Common belief in caution, centered beliefs and rationality).* Consider a game form $G = (C_i)_{i \in I}$, an epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G, and a pair $u = (u_i)_{i \in I}$ of utility functions. A type $\theta_i \in \Theta_i$ expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, if every type in $\Theta^*(\theta_i)$ is cautious, holds ε-centered beliefs around u, and believes in the opponent's rationality under the closest utility function.

A major difference with ε-proper rationalizability is thus that we require players to believe in the opponent's rationality – that is, to only deem possible opponent's choice-type pairs (c_j, θ_j) where c_j is actually optimal for θ_j. To make this possible, player i may believe that his opponent holds a utility function different from u_j, but still "as close as possible to u_j" in some sense. The concept of ε-proper rationalizability, in contrast, requires player i to believe that j's utility function is u_j – and no other – but at the same time allows player i to deem possible choice-type pairs (c_j, t_j) where c_j is not optimal for t_j.

In Section 6 we will see, however, that the concepts of "ε-proper rationalizability" and "common belief in caution, ε-centered beliefs around u, and rationality under the closest utility function" yields exactly the same belief hierarchies on choices for a given game $\Gamma = (C_i, u_i)_{i \in I}$. In that sense, "common belief in caution, ε-centered beliefs around u, and rationality under the closest utility function" may be viewed as an alternative epistemic characterization of ε-proper rationalizability. However, before we prove that result we first formally define the belief hierarchies on choices induced by types in an epistemic model, and establish some important lemmas that are needed for the proof of this characterization.

4. From types to belief hierarchies

In this section we show how to derive, for a given type within an epistemic model, the full belief hierarchy it induces on the players' choices. We first consider epistemic models with *complete* information, and subsequently we turn to epistemic models with *incomplete* information. This is essential for a formal statement of our characterization result in Section 6.

4.1. Complete information

Take a finite epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for the game $\Gamma = (C_i, u_i)_{i \in I}$. For every type $t_i \in T_i$ we can derive the belief about j's choices, by taking the marginal of $b_i(t_i)$ on C_j. We call this t_i's *first-order* belief. But we can also derive the belief it has about j's first-order beliefs, which we call t_i's *second-order* belief. In fact, we can derive for every type $t_i \in T_i$ the full belief hierarchy, consisting of a first-order belief, second-order belief, third-order belief, and so on. Formally, this works as follows.

For every type $t_i \in T_i$ we define the induced first-order belief $h_i^1(t_i) \in \Delta(C_j)$ by

$$h_i^1(t_i)(c_j) := b_i(t_i)(\{c_j\} \times T_j)$$

for every $c_j \in C_j$. Let

$$h_i^1(T_i) := \{h_i^1(t_i) \mid t_i \in T_i\}$$

be the set of first-order beliefs for player i induced by types in T_i.

Now, suppose that $m \geq 2$, and that the beliefs $h_i^{m-1}(t_i)$ and the sets $h_i^{m-1}(T_i)$ have been defined for both players i, and every type $t_i \in T_i$. For every $h_i^{m-1} \in h_i^{m-1}(T_i)$, let

$$T_i[h_i^{m-1}] := \{t_i \in T_i \mid h_i^{m-1}(t_i) = h_i^{m-1}\}.$$

We recursively define the beliefs $h_i^m(t_i)$ and the sets $h_i^m(T_i)$ as follows. For every type $t_i \in T_i$, let $h_i^m(t_i)$ be the m-th order belief on $C_j \times h_j^{m-1}(T_j)$ given by

$$h_i^m(t_i)(c_j, h_j^{m-1}) := b_i(t_i)(\{c_j\} \times T_j[h_j^{m-1}])$$

for every $c_j \in C_j$ and every $h_j^{m-1} \in h_j^{m-1}(T_j)$. By

$$h_i^m(T_i) := \{h_i^m(t_i) \mid t_i \in T_i\}$$

we denote the set of m-th order beliefs for player i induced by types in T_i.

Finally, for every type $t_i \in T_i$, we denote by

$$h_i(t_i) := (h_i^m(t_i))_{m \in \mathbb{N}}$$

the *belief hierarchy* on the players' choices induced by t_i.

4.2. Incomplete information

Consider a finite epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for the game form $G = (C_i)_{i \in I}$. In a similar way as for epistemic models with complete information, we can derive for every type $\theta_i \in \Theta_i$ the full belief hierarchy on the players' choices.

For every type $\theta_i \in \Theta_i$, let $h_i^1(\theta_i) \in \Delta(C_j)$ be the first-order belief given by

$$h_i^1(\theta_i)(c_j) := \beta_i(\theta_i)(\{c_j\} \times \Theta_j)$$

for all $c_j \in C_j$. Let

$$h_i^1(\Theta_i) := \{h_i^1(\theta_i) \mid \theta_i \in \Theta_i\}$$

be the set of first-order beliefs induced by types in Θ_i.

Let $m \geq 2$, and suppose that the beliefs $h_i^{m-1}(\theta_i)$ and the sets $h_i^{m-1}(\Theta_i)$ have been defined for both players i, and all types $\theta_i \in \Theta_i$. For every $h_i^{m-1} \in h_i^{m-1}(\Theta_i)$, let

$$\Theta_i[h_i^{m-1}] := \{\theta_i \in \Theta_i \mid h_i^{m-1}(\theta_i) = h_i^{m-1}\}.$$

For every type $\theta_i \in \Theta_i$, let $h_i^m(\theta_i) \in \Delta(C_j \times h_j^{m-1}(\Theta_j))$ be the m-th order belief given by

$$h_i^m(\theta_i)(c_j, h_j^{m-1}) := \beta_i(\theta_i)(\{c_j\} \times \Theta_j[h_j^{m-1}])$$

for all $c_j \in C_j$ and all $h_j^{m-1} \in h_j^{m-1}(\Theta_j)$. Let

$$h_i^m(\Theta_i) := \{h_i^m(\theta_i) \mid \theta_i \in \Theta_i\}$$

be the set of m-th order beliefs induced by types in Θ_i.

Finally, for every $\theta_i \in \Theta_i$ we denote by

$$h_i(\theta_i) := (h_i^m(\theta_i))_{m \in \mathbb{N}}$$

the *belief hierarchy* on the players' choices induced by θ_i.

5. Some preparatory results

In this section we will state five preparatory results that are needed to prove our characterization theorem. The proofs of these results can be found in the proofs section at the end of this paper.

For the first three results, fix a finite two-player static game $\Gamma = (C_i, u_i)_{i \in I}$, the corresponding game form $G = (C_i)_{i \in I}$, a finite epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for Γ, and a finite epistemic model with incomplete information $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ for G.

In our first preparatory result, we show that two types which induce the same m-th order belief on choices, also induce the same $(m-1)$-th order belief.

Lemma 5.1 *(Identical m-th order beliefs imply identical $(m-1)$-th order beliefs). Let $m \geq 2$. Then, if two types in $T_i \cup \Theta_i$ induce the same m-th order belief on choices, they also induce the same $(m-1)$-th order belief on choices.*

For our next preparatory result, let $T_i^*(t_i) := T^*(t_i) \cap T_i$ be the set of player i types that t_i reasons about, and let $T_j^*(t_i) := T^*(t_i) \cap T_j$ be the set of player j types that t_i reasons about. Similarly we define $\Theta_i^*(\theta_i)$ and $\Theta_j^*(\theta_i)$.

Lemma 5.2 *(Equivalent types deem possible equivalent opponent's types). Let $t_i^* \in T_i$ and $\theta_i^* \in \Theta_i$ be two types with $h_i(t_i^*) = h_i(\theta_i^*)$. Then,*

(a) for every $\theta_i \in \Theta_i^*(\theta_i^*)$ there is some $t_i \in T_i^*(t_i^*)$ with $h_i(\theta_i) = h_i(t_i)$, and for every $\theta_j \in \Theta_j^*(\theta_i^*)$ there is some $t_j \in T_j^*(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$, and

(b) for every $t_i \in T_i^*(t_i^*)$ there is some $\theta_i \in \Theta_i^*(\theta_i^*)$ with $h_i(t_i) = h_i(\theta_i)$, and for every $t_j \in T_j^*(t_i^*)$ there is some $\theta_j \in \Theta_j^*(\theta_i^*)$ with $h_j(t_j) = h_j(\theta_j)$.

Our third result provides necessary and sufficient conditions such that two types t_i^* and θ_i^* – one from a complete information model and the other from an incomplete information model – induce the same belief hierarchy on choices. This result plays a key role in the proof of our characterization theorem.

Lemma 5.3 *(Equivalent types lemma). Suppose that any two different types in M^{co} induce different belief hierarchies. Consider two types $t_i^* \in T_i$ and $\theta_i^* \in \Theta_i$.*
Then, $h_i(t_i^*) = h_i(\theta_i^*)$ if and only if there are mappings

$$f_i : \Theta_i^*(\theta_i^*) \to T_i^*(t_i^*) \text{ and } f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$$

with $f_i(\theta_i^*) = t_i^*$ such that

$$b_i(f_i(\theta_i))(c_j, t_j) = \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j)) \quad (1)$$

for all $\theta_i \in \Theta_i^*(\theta_i^*)$, all $t_j \in T_j^*(t_i^*)$ and all $c_j \in C_j$, and

$$b_j(f_j(\theta_j))(c_i, t_i) = \beta_j(\theta_j)(\{c_i\} \times f_i^{-1}(t_i)) \quad (2)$$

for all $\theta_j \in \Theta_j^*(\theta_i^*)$, all $t_i \in T_i^*(t_i^*)$ and all $c_i \in C_i$.

Here, by $f_j^{-1}(t_j)$ we denote the set $\{\theta_j \in \Theta_j^*(\theta_i^*) \mid f_j(\theta_j) = t_j\}$. Similarly for $f_i^{-1}(t_i)$.

In the literature, a combination of mappings (f_i, f_j) which satisfies the conditions (1) and (2) is called a *type morphism*. See, for instance, Böge and Eisele (1979), Mertens and Zamir (1985), Heifetz and Samet (1998), Friedenberg and Meier (2011) and Perea and Kets (2016). Heifetz and Samet (1998) prove – in a somewhat different setting than ours – that type morphisms preserve the belief hierarchies. Showing this result is actually part of our proof of Lemma 5.3.

In our fourth preparatory result, we show that for every choice c_i and every belief $b_i \in \Delta(C_j)$, there is a *unique* utility function closest to u_i that rationalizes this choice c_i.

Lemma 5.4 *(Unique closest utility function that rationalizes a choice). For player i, consider a utility function $u_i \in V_i$, a choice $c_i \in C_i$, and a probabilistic belief $b_i \in \Delta(C_j)$. Then, there is a unique utility function $v_i \in V_i$ such that (a) choice c_i is optimal for the utility function v_i and the belief b_i, and (b) there is no other utility function $v_i' \in V_i$ with $d(v_i', u_i) < d(v_i, u_i)$ such that choice c_i is optimal for the utility function v_i' and the belief b_i.*

On the basis of this lemma we may define, for every choice $c_i \in C_i$ and every belief $b_i \in \Delta(C_j)$, the utility function $v_i[c_i, b_i] \in V_i$ as the *unique* utility function such that (a) c_i is optimal for the belief b_i and the utility function $v_i[c_i, b_i]$, and (b) there is no other utility function $v_i' \in V_i$ with $d(v_i', u_i) < d(v_i[c_i, b_i], u_i)$ such that c_i is optimal for b_i and v_i'.

Our last preparatory result links the ranking of two choices c_i and c_i' under the utility function u_i and the belief b_i, to the distance that the corresponding utility functions $v_i[c_i, b_i]$ and $v_i[c_i', b_i]$ have to u_i.

Lemma 5.5 *(Choice ranking lemma). Fix a belief $b_i \in \Delta(C_j)$. Then, for every two choices $c_i, c_i' \in C_i$ we have that $u_i(c_i, b_i) > u_i(c_i', b_i)$, if and only if, $d(v_i[c_i, b_i], u_i) < d(v_i[c_i', b_i], u_i)$.*

Here, $u_i(c_i, b_i)$ denotes the expected utility generated by the choice c_i, the belief b_i, and the utility function u_i. As we already argued in the introduction, this Choice Ranking Lemma is fundamental for our characterization result in the next section. On the basis of this lemma, a "choice c_i' being inferior to c_i" in the complete information setting, can be translated into "the closest utility function rationalizing c_i' being further away than the closest utility function rationalizing c_i" in the incomplete information setting. This equivalence then implies that ε-proper trembling in the complete information setting, stating that inferior choices must be deemed much less likely than superior choices, corresponds in the incomplete information setting to ε-centered beliefs around u and belief in the opponent's rationality under the closest utility function, stating that choices must be rationalized by the closest possible utility function and that more distant utility functions must be deemed much less likely than less distant utility functions. This, in turn, leads to the epistemic characterization of ε-proper rationalizability in the following section.

6. Characterization result

So far we have introduced two different concepts for static games – "ε-proper rationalizability" for a context with complete information, and "common belief in caution, ε-centered beliefs around u and rationality under the closest utility function" for a context with incomplete information. In this section we will prove that both concepts yield precisely the same belief hierarchies on choices. More formally, we prove the following theorem.

Theorem 6.1 *(Characterization result). Consider a finite two-player static game $\Gamma = (C_i, u_i)_{i\in I}$, the corresponding game form $G = (C_i)_{i\in I}$, the corresponding utility pair $u = (u_i)_{i\in I}$, a finite epistemic model $M^{co} = (T_i, b_i)_{i\in I}$ for Γ, and a type $t_i^* \in T_i$. Suppose that any two different types in M^{co} induce different belief hierarchies on choices.*

Then, t_i^* is ε-properly rationalizable, if and only if, there is some finite epistemic model with incomplete information $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ for G, and some type $\theta_i^* \in \Theta_i$, such that

(a) θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, and

(b) $h_i(\theta_i^*) = h_i(t_i^*)$.

That is, for every ε-properly rationalizable type within the complete information setting we can find a type within the incomplete information setting that generates exactly the same belief hierarchy on choices, and which expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function. The other direction, however, is also true: if we can find a type θ_i^* within the incomplete information setting that generates exactly the same belief hierarchy on choices as t_i^*, and which expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, then t_i^* must be ε-properly rationalizable. The above theorem thus provides a characterization of ε-properly rationalizable belief hierarchies within an incomplete information setting.

Proof. (If) Consider some type $t_i^* \in T_i$. Suppose that there is some finite epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ for G, and some type $\theta_i^* \in \Theta_i$, such that θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, and $h_i(\theta_i^*) = h_i(t_i^*)$. We show that t_i^* is ε-properly rationalizable. To that purpose, we show that every $t \in T^*(t_i^*)$ is cautious and satisfies ε-proper trembling.

Since $h_i(\theta_i^*) = h_i(t_i^*)$, we know by Lemma 5.3 that there are mappings $f_i : \Theta_i^*(\theta_i^*) \to T_i^*(t_i^*)$ and $f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$ with $f_i(\theta_i^*) = t_i^*$ such that

$$b_i(f_i(\theta_i))(c_j, t_j) = \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j)) \tag{3}$$

for all $\theta_i \in \Theta_i^*(\theta_i^*)$, all $t_j \in T_j^*(t_i^*)$ and all $c_j \in C_j$, and

$$b_j(f_j(\theta_j))(c_i, t_i) = \beta_j(\theta_j)(\{c_i\} \times f_i^{-1}(t_i)) \tag{4}$$

for all $\theta_j \in \Theta_j^*(\theta_i^*)$, all $t_i \in T_i^*(t_i^*)$ and all $c_i \in C_i$.

For every $\theta_i \in \Theta_i^*(\theta_i^*)$ we have, by construction, that $\Theta^*(\theta_i) \subseteq \Theta^*(\theta_i^*)$. By applying Lemma 5.3 to θ_i and $f_i(\theta_i)$, while restricting the mappings f_i and f_j to $\Theta^*(\theta_i) \subseteq \Theta^*(\theta_i^*)$, it follows that $h_i(f_i(\theta_i)) = h_i(\theta_i)$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$. In a similar fashion it follows that $h_j(f_j(\theta_j)) = h_j(\theta_j)$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$.

Now, take some arbitrary $t_i \in T_i^*(t_i^*)$. Then, by Lemma 5.2, part (b), we know that there is some $\theta_i \in \Theta_i^*(\theta_i^*)$ with $h_i(\theta_i) = h_i(t_i)$. As $h_i(f_i(\theta_i)) = h_i(\theta_i) = h_i(t_i)$, and any two different types in $T_i^*(t_i^*)$ generate different belief hierarchies, it must be that $f_i(\theta_i) = t_i$. So, we have found a $\theta_i \in \Theta_i^*(\theta_i^*)$ with $f_i(\theta_i) = t_i$. As θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, and $\theta_i \in \Theta_i^*(\theta_i^*)$, we know that θ_i is cautious, has ε-centered beliefs around u and believes in j's rationality under the closest utility function. We will show that $t_i = f_i(\theta_i)$ is cautious and satisfies ε-proper trembling.

Caution. To prove that t_i is cautious, we must show that for every $t_j \in T_j(t_i)$, and every $c_j \in C_j$, we have $b_i(t_i)(c_j, t_j) > 0$. Take some $t_j \in T_j(t_i)$. Then, in particular, $t_j \in T_j^*(t_i^*)$, as $t_i \in T_i^*(t_i^*)$. As $t_j \in T_j(t_i)$, we have $b_i(t_i)(C_j \times \{t_j\}) > 0$. But then, by condition (3),

$$\beta_i(\theta_i)(C_j \times f_j^{-1}(t_j)) = b_i(f_i(\theta_i))(C_j \times \{t_j\}) = b_i(t_i)(C_j \times \{t_j\}) > 0.$$

Hence, there is some $\theta_j \in f_j^{-1}(t_j)$ such that $\beta_i(\theta_i)(C_j \times \{\theta_j\}) > 0$. So, $\theta_j \in \Theta_j(\theta_i)$. Since θ_i is cautious, there is for every $c_j \in C_j$ some utility function v_j with $\beta_i(\theta_i)(c_j, \theta_j^{v_j}) > 0$. Note that $h_j(\theta_j^{v_j}) = h_j(\theta_j)$. Since $\theta_j \in f_j^{-1}(t_j)$, we have that $f_j(\theta_j) = t_j$, and hence $h_j(\theta_j) = h_j(t_j)$. We also know that $h_j(f_j(\theta_j^{v_j})) = h_j(\theta_j^{v_j}) = h_j(\theta_j) = h_j(t_j)$. Since any two different types in T_j induce different belief hierarchies, we must necessarily have that $f_j(\theta_j^{v_j}) = t_j$ as well. Hence $\theta_j^{v_j} \in f_j^{-1}(t_j)$, which implies that $\beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j)) > 0$. So, for every $c_j \in C_j$ we have that $\beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j)) > 0$. But then, by condition (3), for every $c_j \in C_j$ we have

$$b_i(t_i)(c_j, t_j) = b_i(f_i(\theta_i))(c_j, t_j) = \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j)) > 0.$$

Since this holds for every $t_j \in T_j(t_i)$, it follows that t_i is cautious.

ε-Proper trembling. We next show that t_i satisfies ε-proper trembling. That is, we must show for every $t_j \in T_j(t_i)$, and every two choices c_j, c'_j with $u_j(c'_j, t_j) < u_j(c_j, t_j)$ that $b_i(t_i)(c'_j, t_j) \leq \varepsilon \cdot b_i(t_i)(c_j, t_j)$. Remember that $t_i = f_i(\theta_i)$ for some $\theta_i \in \Theta_i^*(\theta_i^*)$.

By condition (3),

$$\beta_i(\theta_i)(\{c'_j\} \times f_j^{-1}(t_j)) = b_i(f_i(\theta_i)(c'_j, t_j) = b_i(t_i)(c'_j, t_j) > 0,$$

hence there must be some $\theta_j \in f_j^{-1}(t_j)$ such that $\beta_i(\theta_i)(c'_j, \theta_j) > 0$.

Take some arbitrary $\theta_j \in f_j^{-1}(t_j)$ with $\beta_i(\theta_i)(c'_j, \theta_j) > 0$. Let b_j be the belief that t_j has about i's choices. Since $\theta_j \in f_j^{-1}(t_j)$, we have that $f_j(\theta_j) = t_j$, and hence θ_j has the same belief on i's choices as t_j. We thus conclude that θ_j has belief b_j as well. Since θ_i believes in j's rationality under the closest utility function, c'_j must be optimal for type θ_j, and $d(v_j[c'_j, b_j], u_j) = d(w_j(\theta_j), u_j)$.

Remember that $u_j(c'_j, t_j) < u_j(c_j, t_j)$, which implies that $u_j(c'_j, b_j) < u_j(c_j, b_j)$. By Lemma 5.5, we conclude that

$$d(v_j[c_j, b_j], u_j) < d(v_j[c'_j, b_j], u_j) = d(w_j(\theta_j), u_j).$$

That is, there must be some utility function v_j with $d(v_j, u_j) < d(w_j(\theta_j), u_j)$ such that c_j is optimal for the belief b_j and the utility function v_j. As θ_j has the belief b_j on i's choices, there must be some utility function v_j with $d(v_j, u_j) < d(w_j(\theta_j), u_j)$ such that c_j is optimal for $\theta_j^{v_j}$.

Since $\theta_i \in \Theta_i^*(\theta_i^*)$, we know, by assumption, that θ_i is cautious, holds ε-centered beliefs around u and believes in j's rationality under the closest utility function. As θ_i is cautious, there must be some utility function v_j such that θ_i deems possible the pair $(c_j, \theta_j^{v_j})$. As θ_i believes in j's rationality under the closest utility function, type θ_i only deems possible a choice-type pair $(c_j, \theta_j^{v_j})$ if there is no v''_j with $d(v''_j, u_j) < d(v_j, u_j)$ such c_j is optimal for v''_j and b_j. This implies that there is some utility function v_j with $d(v_j, u_j) < d(w_j(\theta_j), u_j)$, such that θ_i deems possible the pair $(c_j, \theta_j^{v_j})$. As θ_i has ε-centered beliefs around u, we must have that $\beta_i(\theta_i)(c'_j, \theta_j) \leq \varepsilon \cdot \beta_i(\theta_i)(c_j, \theta_j^{v_j})$.

We thus have shown that there is some utility function v_j such that for every $\theta_j \in f_j^{-1}(t_j)$ with $\beta_i(\theta_i)(c'_j, \theta_j) > 0$ we have $\beta_i(\theta_i)(c'_j, \theta_j) \leq \varepsilon \cdot \beta_i(\theta_i)(c_j, \theta_j^{v_j})$. Note that we can choose the v_j independent of the specific $\theta_j \in f_j^{-1}(t_j)$ with $\beta_i(\theta_i)(c'_j, \theta_j) > 0$ as any two types in $f_j^{-1}(t_j)$ induce the same belief hierarchy on choices, namely $h_i(t_i)$. So, by condition (3),

$$\begin{aligned}
b_i(t_i)(c'_j, t_j) &= b_i(f_i(\theta_i))(c'_j, t_j) \\
&= \beta_i(\theta_i)(\{c'_j\} \times f_j^{-1}(t_j)) \\
&= \sum_{\theta_j \in f_j^{-1}(t_j)} \beta_i(\theta_i)(c'_j, \theta_j) \\
&= \sum_{\theta_j \in f_j^{-1}(t_j) : \beta_i(\theta_i)(c'_j, \theta_j) > 0} \beta_i(\theta_i)(c'_j, \theta_j) \\
&\leq \sum_{\theta_j \in f_j^{-1}(t_j) : \beta_i(\theta_i)(c'_j, \theta_j) > 0} \varepsilon \cdot \beta_i(\theta_i)(c_j, \theta_j^{v_j}) \\
&= \varepsilon \cdot \sum_{\theta_j \in f_j^{-1}(t_j) : \beta_i(\theta_i)(c'_j, \theta_j) > 0} \beta_i(\theta_i)(c_j, \theta_j^{v_j}) \\
&\leq \varepsilon \cdot \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j)) \\
&= \varepsilon \cdot b_i(f_i(\theta_i))(c_j, t_j) \\
&= \varepsilon \cdot b_i(t_i)(c_j, t_j),
\end{aligned}$$

and hence $b_i(t_i)(c'_j, t_j) \leq \varepsilon \cdot\ b_i(t_i)(c_j, t_j)$. Here, the first inequality follows from the fact that $\beta_i(\theta_i)(c'_j, \theta_j) \leq \varepsilon \cdot \beta_i(\theta_i)(c_j, \theta_j^{v_j})$ for all $\theta_j \in f_j^{-1}(t_j)$ with $\beta_i(\theta_i)(c'_j, \theta_j) > 0$. The second inequality follows from the observation that $\theta_j^{v_j} \in f_j^{-1}(t_j)$ whenever $\theta_j \in f_j^{-1}(t_j)$ with $\beta_i(\theta_i)(c'_j, \theta_j) > 0$. We thus conclude that $b_i(t_i)(c'_j, t_j) \leq \varepsilon \cdot\ b_i(t_i)(c_j, t_j)$ whenever $u_j(c'_j, t_j) < u_j(c_j, t_j)$, and hence t_i satisfies ε-proper trembling.

We have thus shown that every $t_i \in T_i^*(t_i^*)$ is cautious and satisfies ε-proper trembling. In exactly the same way, it can be shown that also every $t_j \in T_j^*(t_i^*)$ is cautious and satisfies ε-proper trembling. This implies that every $t \in T^*(t_i^*)$ is cautious and satisfies ε-proper trembling. Hence, t_i^* is ε-properly rationalizable, which was to show.

(Only if) Take now some type $t_i^* \in T_i$ which is ε-properly rationalizable. We will construct a finite epistemic model with incomplete information $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ for G, and some type $\theta_i^* \in \Theta_i$, such that (a) θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, and (b) $h_i(\theta_i^*) = h_i(t_i^*)$.

We construct $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ in the following way. Let

$$\Theta_i := \{\theta_i[c_i, t_i] \mid c_i \in C_i,\ t_i \in T_i^*(t_i^*)\},$$
$$\Theta_j := \{\theta_j[c_j, t_j] \mid c_j \in C_j,\ t_j \in T_j^*(t_i^*)\}.$$

That is, we construct a type $\theta_i[c_i, t_i]$ for every player i choice c_i and every player i type $t_i \in T_i^*(t_i^*)$ that t_i^* reasons about. Similarly for the player j types.

For every type $\theta_i[c_i, t_i]$, let $w_i(\theta_i[c_i, t_i])$ be the unique utility function such that (a) c_i is optimal for type t_i under the utility function $w_i(\theta_i[c_i, t_i])$, and (b) there is no other utility function v_i with $d(v_i, u_i) < d(w_i(\theta_i[c_i, t_i]), u_i)$ under which c_i is optimal for t_i. The fact that this utility function exists, and that it is unique, follows from Lemma 5.4. Similarly for the player j types.

Finally, the belief $\beta_i(\theta_i[c_i, t_i])$ of type $\theta_i[c_i, t_i]$ on $C_j \times \Theta_j$ is defined by

$$\beta_i(\theta_i[c_i, t_i])(c_j, \theta_j[c_j', t_j]) := \begin{cases} b_i(t_i)(c_j, t_j), & \text{if } c_j = c_j' \\ 0, & \text{if } c_j \neq c_j' \end{cases} \tag{5}$$

for all $(c_j, \theta_j[c_j', t_j]) \in C_j \times \Theta_j$. Similarly, the belief $\beta_j(\theta_j[c_j, t_j])$ of type $\theta_j[c_j, t_j]$ on $C_i \times \Theta_i$ is defined by

$$\beta_j(\theta_j[c_j, t_j])(c_i, \theta_i[c_i', t_i]) := \begin{cases} b_j(t_j)(c_i, t_i), & \text{if } c_i = c_i' \\ 0, & \text{if } c_i \neq c_i' \end{cases} \tag{6}$$

for all $(c_i, \theta_i[c_i', t_i]) \in C_i \times \Theta_i$. This completes the construction of the epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$.

Define $\theta_i^* := \theta_i[c_i^*, t_i^*]$ for some fixed but arbitrary choice c_i^*. We show that θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function, and that $h_i(\theta_i^*) = h_i(t_i^*)$.

Let us start by showing that $h_i(\theta_i^*) = h_i(t_i^*)$. Here, we use Lemma 5.3. So, we will define functions $f_i : \Theta_i^*(\theta_i^*) \to T_i^*(t_i^*)$ and $f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$ with the desired properties. For every $\theta_i[c_i, t_i] \in \Theta_i^*(\theta_i^*)$, define

$$f_i(\theta_i[c_i, t_i]) := t_i. \tag{7}$$

Similarly, for every $\theta_j[c_j, t_j] \in \Theta_j^*(\theta_i^*)$, define

$$f_j(\theta_j[c_j, t_j]) := t_j. \tag{8}$$

Then, obviously, f_i is a mapping from $\Theta_i^*(\theta_i^*)$ to $T_i^*(t_i^*)$, and f_j is a mapping from $\Theta_j^*(\theta_i^*)$ to $T_j^*(t_i^*)$. As $\theta_i^* = \theta_i[c_i^*, t_i^*]$, we also have that $f_i(\theta_i^*) = t_i^*$.

We now prove that the mappings f_i and f_j satisfy condition (1) in Lemma 5.3. Take some type $\theta_i[c_i, t_i] \in \Theta_i^*(\theta_i^*)$, some choice $c_j \in C_j$ and some type $t_j \in T_j$. Then, by construction,

$$\begin{aligned} b_i(f_i(\theta_i[c_i, t_i]))(c_j, t_j) &= b_i(t_i)(c_j, t_j) \\ &= \sum_{c_j' \in C_j} \beta_i(\theta_i[c_i, t_i])(c_j, \theta_j[c_j', t_j]) \\ &= \beta_i(\theta_i[c_i, t_i])(\{c_j\} \times f_j^{-1}(t_j)), \end{aligned}$$

which implies condition (1). Here, the second equality follows from (5), whereas the third equality follows from (8).

We have thus shown that the mappings f_i and f_j satisfy condition (1) in Lemma 5.3. In a similar way, it can be shown that f_i and f_j satisfy condition (2). As $f_i(\theta_i^*) = t_i^*$, it follows from Lemma 5.3 that $h_i(\theta_i^*) = h_i(t_i^*)$, as we wanted to show.

We finally show that θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function. To that purpose, we show that every $\theta \in \Theta^*(\theta_i^*)$ is cautious, has ε-centered beliefs around u and believes in the opponent's rationality under the closest utility function. Take first some player i type $\theta_i[c_i, t_i] \in \Theta_i(\theta_i^*)$.

Caution. Take some $\theta_j[c_j, t_j] \in \Theta_j(\theta_i[c_i, t_i])$ and some $c_j' \in C_j$. We show that there is some utility function v_j such that θ_i deems possible $(c_j', \theta_j[c_j, t_j]^{v_j})$.

Since $\theta_j[c_j, t_j] \in \Theta_j(\theta_i[c_i, t_i])$, there is some $c_j'' \in C_j$ such that

$$\beta_i(\theta_i[c_i, t_i])(c''_j, \theta_j[c_j, t_j]) > 0.$$

Hence, by (5), $c''_j = c_j$ and

$$b_i(t_i)(c_j, t_j) = \beta_i(\theta_i[c_i, t_i])(c_j, \theta_j[c_j, t_j]) > 0,$$

which implies that $t_j \in T_j(t_i)$.

As $t_i \in T_i^*(t_i^*)$ and t_i^* is ε-properly rationalizable, we know that t_i is cautious. Hence, for every $c'_j \in C_j$, type t_i deems possible the choice-type pair (c'_j, t_j). Take some arbitrary $c'_j \in C_j$. Then, $b_i(t_i)(c'_j, t_j) > 0$, and hence, by (5),

$$\beta_i(\theta_i[c_i, t_i])(c'_j, \theta_j[c'_j, t_j]) > 0.$$

Let $v_j := w_j(\theta_j[c'_j, t_j])$. Then, it may be verified by (6) that $\theta_j[c'_j, t_j] = \theta_j[c_j, t_j]^{v_j}$, and hence

$$\beta_i(\theta_i[c_i, t_i])(c'_j, \theta_j[c_j, t_j]^{v_j}) > 0.$$

That is, $\theta_i[c_i, t_i]$ deems possible the pair $(c'_j, \theta_j[c_j, t_j]^{v_j})$.

So, for every $\theta_j[c_j, t_j] \in \Theta_j(\theta_i[c_i, t_i])$, and every $c'_j \in C_j$, there is some utility function v_j such that θ_i deems possible $(c'_j, \theta_j[c_j, t_j]^{v_j})$. Consequently, $\theta_i[c_i, t_i]$ is cautious.

ε-centered beliefs around u. Suppose that $\theta_i[c_i, t_i]$ deems possible two choice-type pairs $(c_j, \theta_j^{v_j})$ and $(c'_j, \theta_j^{v'_j})$, where $\theta_j^{v_j}$ and $\theta_j^{v'_j}$ only differ in the utility function, and $d(v_j, u_j) < d(v'_j, u_j)$. We must show that

$$\beta_i(\theta_i[c_i, t_i])(c'_j, \theta_j^{v'_j}) \leq \varepsilon \cdot \beta_i(\theta_i[c_i, t_i])(c_j, \theta_j^{v_j}).$$

Since $\theta_i[c_i, t_i]$ deems possible the choice-type pairs $(c_j, \theta_j^{v_j})$ and $(c'_j, \theta_j^{v'_j})$, we conclude from (5) that there must be some types $t_j, t'_j \in T_j(t_i)$ such that

$$\theta_j^{v_j} = \theta_j[c_j, t_j] \text{ and } \theta_j^{v'_j} = \theta_j[c'_j, t'_j].$$

Moreover, as $h_j(\theta_j^{v_j}) = h_j(\theta_j^{v'_j})$, $h_j(\theta_j^{v_j}) = h_j(t_j)$ and $h_j(\theta_j^{v'_j}) = h_j(t'_j)$, it follows that $h_j(t_j) = h_j(t'_j)$. But then, $t_j = t'_j$, as by assumption any two different types in T_j induce different belief hierarchies on choices. So, we conclude that

$$\theta_j^{v_j} = \theta_j[c_j, t_j] \text{ and } \theta_j^{v'_j} = \theta_j[c'_j, t_j].$$

By construction of the type $\theta_j[c_j, t_j]$, we know that c_j is optimal for $\theta_j[c_j, t_j]$, and there is no other utility function v''_j with $d(v''_j, u_j) < d(v_j, u_j)$ such that c_j is optimal for $\theta_j[c_j, t_j]^{v''_j}$. Now, let b_j be the belief that $\theta_j[c_j, t_j]$ has about i's choices. Then, it follows that $v_j = v_j[c_j, b_j]$, where $v_j[c_j, b_j]$ is defined as in Lemma 5.5. As $\theta_j[c'_j, t_j]$ has the same belief b_j about i's choices, it can be shown in the same way that $v'_j = v_j[c'_j, b_j]$. Since we assume that $d(v_j, u_j) < d(v'_j, u_j)$, we conclude that

$$d(v_j[c_j, b_j], u_j) < d(v_j[c'_j, b_j], u_j).$$

But then, by Lemma 5.5, it follows that $u_j(c'_j, b_j) < u_j(c_j, b_j)$. As type t_j has exactly this belief b_j and the utility function u_j, we obtain that t_j prefers c_j to c'_j.

Since $t_i \in T_i^*(t_i^*)$, and t_i^* is ε-properly rationalizable, we know that t_i satisfies ε-proper trembling. As $t_j \in T_j(t_i)$, and t_j prefers c_j to c'_j, we conclude that $b_i(t_i)(c'_j, t_j) \leq \varepsilon \cdot b_i(t_i)(c_j, t_j)$. It then follows by (5) that

$$\begin{aligned}
\beta_i(\theta_i[c_i, t_i])(c'_j, \theta_j[c'_j, t_j]) &= b_i(t_i)(c'_j, t_j) \\
&\leq \varepsilon \cdot b_i(t_i)(c_j, t_j) \\
&= \varepsilon \cdot \beta_i(\theta_i[c_i, t_i])(c_j, \theta_j[c_j, t_j]).
\end{aligned}$$

As $\theta_j^{v_j} = \theta_j[c_j, t_j]$ and $\theta_j^{v'_j} = \theta_j[c'_j, t_j]$, it follows that

$$\beta_i(\theta_i[c_i, t_i])(c'_j, \theta_j^{v'_j}) \leq \varepsilon \cdot \beta_i(\theta_i[c_i, t_i])(c_j, \theta_j^{v_j}),$$

which was to show. As such, we conclude that $\theta_i[c_i, t_i]$ has ε-centered beliefs around u.

Belief in j's rationality under the closest utility function. Suppose that $\theta_i[c_i, t_i]$ deems possible a choice-type pair $(c_j, \theta_j[c'_j, t_j])$. That is,

$$\beta_i(\theta_i[c_i, t_i])(c_j, \theta_j[c'_j, t_j]) > 0.$$

By (5), it follows that $c_j = c'_j$, and hence $(c_j, \theta_j[c'_j, t_j]) = (c_j, \theta_j[c_j, t_j])$. By construction, the utility function $w_j(\theta_j[c_j, t_j])$ is chosen such that (a) c_j is optimal for type t_j under the utility function $w_j(\theta_j[c_j, t_j])$, and (b) there is no other utility function v_j with $d(v_j, u_j) < d(w_j(\theta_j[c_j, t_j]), u_j)$ under which c_j is optimal for t_j. Since, by (6), $\theta_j[c_j, t_j]$ has the same belief on C_i as t_j, it follows that c_j is optimal for type $\theta_j[c_j, t_j]$. Hence, $\theta_i[c_i, t_i]$ only deems possible choice-type pairs $(c_j, \theta_j[c_j, t_j])$ where (a) c_j is optimal for $\theta_j[c_j, t_j]$, and (b) there is no other utility function v_j with $d(v_j, u_j) < d(w_j(\theta_j[c_j, t_j]), u_j)$ such that c_j is optimal for $\theta_j[c_j, t_j]^{v_j}$. Hence, $\theta_i[c_i, t_i]$ believes in j's rationality under the closest utility function.

Summarizing, we see that every $\theta_i[c_i, t_i] \in \Theta_i^*(\theta_i^*)$ is cautious, has ε-centered beliefs around u, and believes in j's rationality under the closest utility function. As the same can be shown for every $\theta_j[c_j, t_j] \in \Theta_j^*(\theta_i^*)$, we conclude that every $\theta \in \Theta^*(\theta_i^*)$ is cautious, has ε-centered beliefs around u and believes in the opponent's rationality under the closest utility function. Hence, θ_i^* expresses common belief in caution, ε-centered beliefs around u and rationality under the closest utility function. Since $h_i(\theta_i^*) = h_i(t_i^*)$, the proof is complete. $\square$

7. Concluding remarks

7.1. Characterization of proper rationalizability

In this paper we have given an epistemic characterization of ε-proper rationalizability, for a fixed ε, within an incomplete information setting. The crucial conditions we use are *ε-centered beliefs around u* and *belief in the opponent's rationality under the closest utility function.* A natural question is whether similar conditions can be used to characterize *proper rationalizability,* which essentially corresponds to ε-proper rationalizability as ε tends to zero.

Asheim (2001) has shown that proper rationalizability can be characterized by means of *lexicographic beliefs,* and that the ε-proper trembling condition, as ε tends to zero, can be mimicked by the requirement that better choices must be deemed *infinitely more likely* than inferior choices within the lexicographic belief. On the basis of this insight, we conjecture that proper rationalizability can be characterized within an incomplete information setting by using *lexicographic beliefs.* More precisely, we could introduce a lexicographic notion of *centered beliefs around u*, stating that closer utility functions must be deemed *infinitely more likely* than more distant utility functions. We believe that (common belief in) this condition, together with appropriately adapted definitions of caution and belief in the opponent's rationality under the closest utility function, would characterize proper rationalizability. This investigation, however, is beyond the scope of this paper, and is left for future research.

7.2. Distance measure

The distance measure between utility functions that we use in this paper (see Section 3.2) has its problems, since it is not invariant with respect to positive affine transformations of the utility functions.[1] If we view a utility function as the numerical representation of a preference relation over lotteries, then ideally the distance measure should be invariant with respect to such positive affine transformations, as these would not change the underlying preference relation. A possible remedy could be to assume that all utility functions are normalized in a uniform way, for instance by requiring that the minimum utility is 0 and the maximum utility is 1 (provided the underlying preference relation is non-trivial). This would remove the invariance problem, but would cause other, technical problems. For instance, the space of normalized utility functions would no longer be convex – a property that is important for showing that there is always a unique utility function that rationalizes a certain choice and that is closest to u. We conjecture that our characterization would still hold by restricting to normalized utility functions, but we are short of a formal proof at this moment. We hope to be able to clarify this issue in the near future.

7.3. Beliefs about utilities versus beliefs about preferences

In this paper, when investigating the context with incomplete information, we assume that players hold beliefs about the specific opponent's utility function. Similarly to the problem above, one could argue that only the belief about the opponent's underlying preference relation over lotteries should matter, and not so much the belief about the precise numerical representation of it. It would thus be desirable to have a model in which we describe the players' beliefs about the opponent's preference relations rather than their beliefs about utilities. At this moment we have no satisfactory resolution to this problem, but hope to address this issue in the future.

[1] We thank an anonymous referee for pointing this out to us.

8. Proofs

Proof of Lemma 5.1. We prove the statement by induction on m. Consider first the case where $m = 2$.

Take two types in $T_i \cup \Theta_i$ that induce the same second-order belief on choices. Suppose that the first type is in T_i – call it t_i – and the second type is in Θ_i – call it θ_i. So, $h_i^2(t_i) = h_i^2(\theta_i)$. We show that t_i and θ_i also induce the same first-order belief on choices – that is, $h_i^1(t_i) = h_i^1(\theta_i)$.

For every $c_j \in C_j$,

$$\begin{aligned}
h_i^1(t_i)(c_j) &= b_i(t_i)(\{c_j\} \times T_j) \\
&= \sum_{h_j^1 \in h_j^1(T_j) \cup h_j^1(\Theta_j)} b_i(t_i)(\{c_j\} \times T_j[h_j^1]) \\
&= \sum_{h_j^1 \in h_j^1(T_j) \cup h_j^1(\Theta_j)} h_i^2(t_i)(c_j, h_j^1) \\
&= \sum_{h_j^1 \in h_j^1(T_j) \cup h_j^1(\Theta_j)} h_i^2(\theta_i)(c_j, h_j^1) \\
&= \sum_{h_j^1 \in h_j^1(T_j) \cup h_j^1(\Theta_j)} \beta_i(\theta_i)(\{c_j\} \times \Theta_j[h_j^1]) \\
&= \beta_i(\theta_i)(\{c_j\} \times \Theta_j) \\
&= h_i^1(\theta_i)(c_j),
\end{aligned}$$

which implies that $h_i^1(t_i) = h_i^1(\theta_i)$. Here, the fourth equality follows from the assumption that $h_i^2(t_i) = h_i^2(\theta_i)$. The proof for the case when both types are from T_i, or when both types are from Θ_i, is very similar, and is therefore omitted.

Take now some $m \geq 3$, and suppose that the statement is true for $m-1$, for both players i. Consider some type $t_i \in T_i$ and some type $\theta_i \in \Theta_i$ with $h_i^m(t_i) = h_i^m(\theta_i)$. We show that $h_i^{m-1}(t_i) = h_i^{m-1}(\theta_i)$.

To prove this, we first show that

$$h_j^{m-1}(T_j(t_i)) = h_j^{m-1}(\Theta_j(\theta_i)), \tag{9}$$

and, for every $h_j^{m-1} \in h_j^{m-1}(T_j)$,

$$h_j^{m-2}(T_j[h_j^{m-1}]) = h_j^{m-2}(\Theta_j[h_j^{m-1}]) = \{h_j^{m-2}\} \tag{10}$$

for some $h_j^{m-2} \in h_j^{m-2}(T_j)$. Here, $h_j^{m-1}(T_j(t_i)) = \{h_j^{m-1}(t_j) \mid t_j \in T_j(t_i)\}$ and $h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}(t_j) \mid t_j \in T_j[h_j^{m-1}]\}$, and similarly for Θ_j.

We first show (9). By definition, for every $h_j^{m-1} \in h_j^{m-1}(T_j)$,

$$\begin{aligned}
h_i^m(t_i)(C_j \times \{h_j^{m-1}\}) &= b_i(t_i)(C_j \times T_j[h_j^{m-1}]) \\
&= b_i(t_i)(C_j \times \{t_j \in T_j \mid h_j^{m-1}(t_j) = h_j^{m-1}\}) \\
&= b_i(t_i)(C_j \times \{t_j \in T_j(t_i) \mid h_j^{m-1}(t_j) = h_j^{m-1}\}),
\end{aligned} \tag{11}$$

where the third equality follows from the fact that $b_i(t_i)$ only assigns positive probability to types in $T_j(t_i)$. In fact, $b_i(t_i)$ assigns positive probability *precisely* to those types that are in $T_j(t_i)$. Hence, it follows from (11) that $h_i^m(t_i)(C_j \times \{h_j^{m-1}\}) > 0$ if and only if there is some $t_j \in T_j(t_i)$ with $h_j^{m-1}(t_j) = h_j^{m-1}$, which is the case precisely when $h_j^{m-1} \in h_j^{m-1}(T_j(t_i))$.

In a similar way, it follows that $h_i^m(\theta_i)(C_j \times \{h_j^{m-1}\}) > 0$ if and only if $h_j^{m-1} \in h_j^{m-1}(\Theta_j(\theta_i))$. Since, by the induction assumption, $h_i^m(t_i) = h_i^m(\theta_i)$, it follows that $h_j^{m-1}(T_j(t_i)) = h_j^{m-1}(\Theta_j(\theta_i))$, and hence (9) holds.

We now prove (10). We first show that $h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}$ for some $h_j^{m-2} \in h_j^{m-2}(T_j)$. Take two types $t_j, t_j' \in T_j[h_j^{m-1}]$. That is, $h_j^{m-1}(t_j) = h_j^{m-1}(t_j') = h_j^{m-1}$. Then, by the induction assumption, it follows that $h_j^{m-2}(t_j) = h_j^{m-2}(t_j')$. So, all types in $T_j[h_j^{m-1}]$ induce the same $(m-2)$-th order belief, which we call h_j^{m-2}. So, $h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}$.

Next we show that $h_j^{m-2}(\Theta_j[h_j^{m-1}]) = \{h_j^{m-2}\}$ as well. Take some $\theta_j \in \Theta_j[h_j^{m-1}]$ and some $t_j \in T_j[h_j^{m-1}]$. As $h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}$, it follows that $h_j^{m-2}(t_j) = h_j^{m-2}$. Since $h_j^{m-1}(\theta_j) = h_j^{m-1} = h_j^{m-1}(t_j)$, it follows by the induction assumption that $h_j^{m-2}(\theta_j) = h_j^{m-2}(t_j) = h_j^{m-2}$. So, we may conclude that $h_j^{m-2}(\Theta_j[h_j^{m-1}]) = \{h_j^{m-2}\}$. We have thus shown (10).

We now prove that $h_i^{m-1}(t_i) = h_i^{m-1}(\theta_i)$. By (10) we know that for every $h_j^{m-1} \in h_j^{m-1}(T_j)$ there is some $h_j^{m-2} \in h_j^{m-2}(T_j)$ with $h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}$. Consequently, for every $h_j^{m-2} \in h_j^{m-2}(T_j)$,

$$T_j[h_j^{m-2}] = \bigcup_{h_j^{m-1} \in h_j^{m-1}(T_j): h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}} T_j[h_j^{m-1}],$$

and hence

$$T_j[h_j^{m-2}] \cap T_j(t_i) = \bigcup_{h_j^{m-1} \in h_j^{m-1}(T_j(t_i)): h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}} T_j[h_j^{m-1}] \cap T_j(t_i). \tag{12}$$

So, for every $c_j \in C_j$ and $h_j^{m-2} \in h_j^{m-2}(T_j)$,

$$\begin{aligned}
h_i^{m-1}(t_i)(c_j, h_j^{m-2}) &= b_i(t_i)(\{c_j\} \times T_j[h_j^{m-2}]) \\
&= b_i(t_i)(\{c_j\} \times (T_j[h_j^{m-2}] \cap T_j(t_i))) \\
&= \sum_{h_j^{m-1} \in h_j^{m-1}(T_j(t_i)): h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}} b_i(t_i)(\{c_j\} \times (T_j[h_j^{m-1}] \cap T_j(t_i))) \\
&= \sum_{h_j^{m-1} \in h_j^{m-1}(T_j(t_i)): h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}} b_i(t_i)(\{c_j\} \times T_j[h_j^{m-1}]) \\
&= \sum_{h_j^{m-1} \in h_j^{m-1}(T_j(t_i)): h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}} h_i^m(t_i)(c_j, h_j^{m-1}) \\
&= \sum_{h_j^{m-1} \in h_j^{m-1}(T_j(t_i)): h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}} h_i^m(\theta_i)(c_j, h_j^{m-1}) \\
&= \sum_{h_j^{m-1} \in h_j^{m-1}(\Theta_j(\theta_i)): h_j^{m-2}(\Theta_j[h_j^{m-1}]) = \{h_j^{m-2}\}} h_i^m(\theta_i)(c_j, h_j^{m-1}) \\
&= h_i^{m-1}(\theta_i)(c_j, h_j^{m-2}),
\end{aligned}$$

which implies that $h_i^{m-1}(t_i) = h_i^{m-1}(\theta_i)$. Here, the second equality follows from the fact that $b_i(t_i)$ only assigns positive probability to types in $T_j(t_i)$. The third equality follows from (12). The fourth equality follows, again, from the fact that $b_i(t_i)$ only assigns positive probability to types in $T_j(t_i)$. The fifth equality follows from the definition of $h_i^m(t_i)$. The sixth equality follows from the assumption that $h_i^m(t_i) = h_i^m(\theta_i)$. The seventh equality follows from (9), that $h_j^{m-1}(T_j(t_i)) = h_j^{m-1}(\Theta_j(\theta_i))$, and from (10) which implies that $h_j^{m-2}(T_j[h_j^{m-1}]) = \{h_j^{m-2}\}$ if and only if $h_j^{m-2}(\Theta_j[h_j^{m-1}]) = \{h_j^{m-2}\}$. The eighth equality follows from mimicking the first five equalities, in reverse order, to

$$\sum_{h_j^{m-1} \in h_j^{m-1}(\Theta_j(\theta_i)): h_j^{m-2}(\Theta_j[h_j^{m-1}]) = \{h_j^{m-2}\}} h_i^m(\theta_i)(c_j, h_j^{m-1}).$$

In a similar way, it can be shown that any two types in T_i, or any two types in Θ_i, which induce the same m-th order belief, also induce the same $(m-1)$-th order belief.

By induction on m, the proof is complete. $\square$

Proof of Lemma 5.2. We only prove (a), as the proof for (b) is very similar. Remember that

$$\begin{aligned}
\Theta_i^*(\theta_i^*) &= \bigcup_{k \in \mathbb{N}} \Theta_i^k(\theta_i^*), \ \Theta_j^*(\theta_i^*) = \bigcup_{k \in \mathbb{N}} \Theta_j^k(\theta_i^*), \\
T_i^*(t_i^*) &= \bigcup_{k \in \mathbb{N}} T_i^k(t_i^*), \ T_j^*(t_i^*) = \bigcup_{k \in \mathbb{N}} T_j^k(t_i^*).
\end{aligned}$$

We prove, by induction on k, that for every $\theta_i \in \Theta_i^k(\theta_i^*)$ there is some $t_i \in T_i^k(t_i^*)$ with $h_i(\theta_i) = h_i(t_i)$, and for every $\theta_j \in \Theta_j^k(\theta_i^*)$ there is some $t_j \in T_j^k(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$.

From Lemma 5.1 we know that any two types in $T_i \cup \Theta_i$ which induce the same m-th order belief, also induce the same first-order, second order, ... , $(m-1)$-th order belief on choices. As M^{co} and M^{in} contain only finitely many types, there must be some $m \in \mathbb{N}$ such that, for every $t_i \in T_i$ and $\theta_i \in \Theta_i$,

$$h_i(t_i) = h_i(\theta_i) \ \text{ if and only if } h_i^m(t_i) = h_i^m(\theta_i). \tag{13}$$

So, it is sufficient to show that for every $\theta_i \in \Theta_i^k(\theta_i^*)$ there is some $t_i \in T_i^k(t_i^*)$ with $h_i^m(\theta_i) = h_i^m(t_i)$, and for every $\theta_j \in \Theta_j^k(\theta_i^*)$ there is some $t_j \in T_j^k(t_i^*)$ with $h_j^m(\theta_j) = h_j^m(t_j)$.

Consider first the case where $k = 1$. By definition, $\Theta_i^1(\theta_i^*) = \{\theta_i^*\}$ and $T_i^1(t_i^*) = \{t_i^*\}$. As, by assumption, $h_i(\theta_i^*) = h_i(t_i^*)$, the statement holds for $\Theta_i^1(\theta_i^*)$ and $T_i^1(t_i^*)$.

Now, turn to $\Theta_j^1(\theta_i^*)$ and $T_j^1(t_i^*)$. By definition,

$$\Theta_j^1(\theta_i^*) = \{\theta_j \in \Theta_j \mid \beta_i(\theta_i^*)(C_j \times \{\theta_j\}) > 0,$$

and

$$T_j^1(t_i^*) = \{t_j \in T_j \mid b_i(t_i^*)(C_j \times \{t_j\}) > 0.$$

Take some arbitrary $\theta_j \in \Theta_j^1(\theta_i^*)$. Then, $\beta_i(\theta_i^*)(C_j \times \{\theta_j\}) > 0$. Now, choose m as in (13). Then,

$$\begin{aligned} h_i^{m+1}(\theta_i^*)(C_j \times \{h_j^m(\theta_j)\}) &= \beta_i(\theta_i^*)(C_j \times \Theta_j[h_j^m(\theta_j)]) \\ &\geq \beta_i(\theta_i^*)(C_j \times \{\theta_j\}) > 0, \end{aligned}$$

where the inequality follows from the fact that $\theta_j \in \Theta_j[h_j^m(\theta_j)]$. As $h_i(\theta_i^*) = h_i(t_i^*)$, we must have that

$$h_i^{m+1}(\theta_i^*)(C_j \times \{h_j^m(\theta_j)\}) = h_i^{m+1}(t_i^*)(C_j \times \{h_j^m(\theta_j)\}),$$

and hence $h_i^{m+1}(t_i^*)(C_j \times \{h_j^m(\theta_j)\}) > 0$. Therefore,

$$\begin{aligned} h_i^{m+1}(t_i^*)(C_j \times \{h_j^m(\theta_j)\}) &= b_i(t_i^*)(C_j \times T_j[h_j^m(\theta_j)]) \\ &= b_i(t_i^*)(C_j \times \{t_j \in T_j \mid h_j^m(t_j) = h_j^m(\theta_j)\}) \\ &= b_i(t_i^*)(C_j \times \{t_j \in T_j^1(t_i^*) \mid h_j^m(t_j) = h_j^m(\theta_j)\}) > 0, \end{aligned}$$

where the third equality follows from the fact that $b_i(t_i^*)$ only assigns positive probability to j's types in $T_j^1(t_i^*)$. Hence, there must be some $t_j \in T_j^1(t_i^*)$ with $h_j^m(t_j) = h_j^m(\theta_j)$. By (13) it then follows that $h_j(t_j) = h_j(\theta_j)$.

So, we see that for every $\theta_j \in \Theta_j^1(\theta_i^*)$ there is some $t_j \in T_j^1(t_i^*)$ with $h_j(t_j) = h_j(\theta_j)$. This completes the induction start, with $k = 1$.

Take now some $k \geq 2$, and suppose, by the induction assumption, that for every $\theta_i \in \Theta_i^{k-1}(\theta_i^*)$ there is some $t_i \in T_i^{k-1}(t_i^*)$ with $h_i(\theta_i) = h_i(t_i)$, and for every $\theta_j \in \Theta_j^{k-1}(\theta_i^*)$ there is some $t_j \in T_j^{k-1}(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$. We prove that for every $\theta_i \in \Theta_i^k(\theta_i^*)$ there is some $t_i \in T_i^k(t_i^*)$ with $h_i(\theta_i) = h_i(t_i)$, and for every $\theta_j \in \Theta_j^k(\theta_i^*)$ there is some $t_j \in T_j^k(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$.

Take first some $\theta_i^k \in \Theta_i^k(\theta_i^*)$. Then, there is some $\theta_j^{k-1} \in \Theta_j^{k-1}(\theta_i^*)$ such that $\beta_j(\theta_j^{k-1})(C_i \times \{\theta_i^k\}) > 0$. Choose m as in (13). Then,

$$\begin{aligned} h_j^{m+1}(\theta_j^{k-1})(C_i \times \{h_i^m(\theta_i^k)\}) &= \beta_j(\theta_j^{k-1})(C_i \times \Theta_i[h_i^m(\theta_i^k)]) \\ &\geq \beta_j(\theta_j^{k-1})(C_i \times \{\theta_i^k\}) > 0, \end{aligned}$$

where the inequality follows from the fact that $\theta_i^k \in \Theta_i[h_i^m(\theta_i^k)]$. Since $\theta_j^{k-1} \in \Theta_j^{k-1}(\theta_i^*)$, we know by the induction assumption that there must be some $t_j^{k-1} \in T_j^{k-1}(t_i^*)$ with $h_j(\theta_j^{k-1}) = h_j(t_j^{k-1})$. As such,

$$h_j^{m+1}(\theta_j^{k-1})(C_i \times \{h_i^m(\theta_i^k)\}) = h_j^{m+1}(t_j^{k-1})(C_i \times \{h_i^m(\theta_i^k)\}),$$

which implies that $h_j^{m+1}(t_j^{k-1})(C_i \times \{h_i^m(\theta_i^k)\}) > 0$. Therefore,

$$\begin{aligned} h_j^{m+1}(t_j^{k-1})(C_i \times \{h_i^m(\theta_i^k)\}) &= b_j(t_j^{k-1})(C_i \times T_i[h_i^m(\theta_i^k)]) \\ &= b_j(t_j^{k-1})(C_i \times \{t_i \in T_i \mid h_i^m(t_i) = h_i^m(\theta_i^k)\}) \\ &= b_j(t_j^{k-1})(C_i \times \{t_i \in T_i^k(t_i^*) \mid h_i^m(t_i) = h_i^m(\theta_i^k)\}) > 0, \end{aligned}$$

where the third equality follows from the fact that $b_j(t_j^{k-1})$ only assigns positive probability to i's types in $T_i^k(t_i^*)$, since $t_j^{k-1} \in T_j^{k-1}(t_i^*)$. Hence, there must be some $t_i^k \in T_i^k(t_i^*)$ with $h_i^m(t_i^k) = h_i^m(\theta_i^k)$. By (13) it follows that $h_i(t_i^k) = h_i(\theta_i^k)$.

Hence, we can find for every $\theta_i^k \in \Theta_i^k(\theta_i^*)$ some $t_i^k \in T_i^k(t_i^*)$ with $h_i(t_i^k) = h_i(\theta_i^k)$.

In a similar fashion, we can then show that for every $\theta_j \in \Theta_j^k(\theta_i^*)$ there is some $t_j \in T_j^k(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$. This completes the induction step.

By induction on k, we can thus conclude that for every $\theta_i \in \Theta_i^*(\theta_i^*)$ there is some $t_i \in T_i^*(t_i^*)$ with $h_i(\theta_i) = h_i(t_i)$, and for every $\theta_j \in \Theta_j^*(\theta_i^*)$ there is some $t_j \in T_j^*(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$. This completes the proof. □

Proof of Lemma 5.3. **(If)** Suppose first that there are mappings $f_i : \Theta_i^*(\theta_i^*) \to T_i^*(t_i^*)$ and $f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$ with $f_i(\theta_i^*) = t_i^*$ which satisfy the conditions (1) and (2). We show that $h_i(t_i^*) = h_i(\theta_i^*)$. In fact, we will show that $h_i(\theta_i) = h_i(f_i(\theta_i))$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$ and $h_j(\theta_j) = h_j(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$.

In order to show the latter, we prove, by induction on m, that $h_i^m(\theta_i) = h_i^m(f_i(\theta_i))$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$ and $h_j^m(\theta_j) = h_j^m(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$.

Consider first the case $m = 1$. Take some $\theta_i \in \Theta_i^*(\theta_i^*)$. Then, by definition, $h_i^1(\theta_i)$ and $h_i^1(f_i(\theta_i))$ are both in $\Delta(C_j)$. Moreover, for every $c_j \in C_j$,

$$\begin{aligned} h_i^1(\theta_i)(c_j) &= \beta_i(\theta_i)(\{c_j\} \times \Theta_j) \\ &= \beta_i(\theta_i)(\{c_j\} \times \Theta_j^*(\theta_i^*)) \\ &= \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(T_j^*(t_i^*)) \\ &= b_i(f_i(\theta))(\{c_j\} \times T_j^*(t_i^*)) \\ &= b_i(f_i(\theta))(\{c_j\} \times T_j) \\ &= h_i^1(f_i(\theta_i))(c_j), \end{aligned}$$

which implies that $h_i^1(\theta_i) = h_i^1(f_i(\theta_i))$. Here, the second equality follows from the observation that θ_i only deems possible j's types in $\Theta_j^*(\theta_i^*)$, as $\theta_i \in \Theta_i^*(\theta_i^*)$. The third equality follows from the assumption that $f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$. The fourth equality follows from condition (1). The fifth equality follows from the observation that $f_i(\theta_i)$ only deems possible j's types in $T_j^*(t_i^*)$, as $f_i(\theta_i) \in T_i^*(t_i^*)$.

In a similar way we can prove that $h_j^1(\theta_j) = h_j^1(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$.

Consider now some $m \geq 2$, and assume that $h_i^{m-1}(\theta_i) = h_i^{m-1}(f_i(\theta_i))$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$ and $h_j^{m-1}(\theta_j) = h_j^{m-1}(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$. Take some $\theta_i \in \Theta_i^*(\theta_i^*)$. Then, $h_i^m(\theta_i) \in \Delta(C_j \times h_j^{m-1}(\Theta_j^*(\theta_i^*)))$ and $h_i^m(f_i(\theta_i)) \in \Delta(C_j \times h_j^{m-1}(T_j^*(t_i^*)))$.

We will now show that $h_i^m(\theta_i) = h_i^m(f_i(\theta_i))$. For every $c_j \in C_j$ and every $h_j^{m-1} \in h_j^{m-1}(\Theta_j^*(\theta_i^*))$ we have

$$\begin{aligned} h_i^m(\theta_i)(c_j, h_j^{m-1}) &= \beta_i(\theta_i)(\{c_j\} \times \Theta_j[h_j^{m-1}]) \\ &= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j \mid h_j^{m-1}(\theta_j) = h_j^{m-1}\}) \\ &= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid h_j^{m-1}(\theta_j) = h_j^{m-1}\}) \\ &= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid h_j^{m-1}(f_j(\theta_j)) = h_j^{m-1}\}) \\ &= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid f_j(\theta_j) \in T_j[h_j^{m-1}]\}) \\ &= \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(T_j[h_j^{m-1}])) \\ &= b_i(f_i(\theta_i))(\{c_j\} \times T_j[h_j^{m-1}]) \\ &= h_i^m(f_i(\theta_i))(c_j, h_j^{m-1}), \end{aligned}$$

which implies that $h_i^m(\theta_i) = h_i^m(f_i(\theta_i))$. Here, the third equality follows from the fact that θ_i only assigns positive probability to j's types in $\Theta_j^*(\theta_i^*)$. The fourth equality follows from the induction assumption that $h_j^{m-1}(\theta_j) = h_j^{m-1}(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$. The seventh equality follows from condition (1).

Hence, we have shown that $h_i^m(\theta_i) = h_i^m(f_i(\theta_i))$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$. In a similar way, it can be shown that $h_j^m(\theta_j) = h_j^m(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$.

By induction on m, we may conclude that $h_i(\theta_i) = h_i(f_i(\theta_i))$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$ and $h_j(\theta_j) = h_j(f_j(\theta_j))$ for all $\theta_j \in \Theta_j^*(\theta_i^*)$. In particular, since $f_i(\theta_i^*) = t_i^*$, we may conclude that $h_i(\theta_i^*) = h_i(t_i^*)$, which was to show.

(Only if) Suppose now that $h_i(\theta_i^*) = h_i(t_i^*)$. We prove that there are mappings $f_i : \Theta_i^*(\theta_i^*) \to T_i^*(t_i^*)$ and $f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$ with $f_i(\theta_i^*) = t_i^*$ which satisfy the conditions (1) and (2).

As $h_i(\theta_i^*) = h_i(t_i^*)$, we know by Lemma 5.2 that for every $\theta_i \in \Theta_i^*(\theta_i^*)$ there is some $t_i \in T_i^*(t_i^*)$ with $h_i(\theta_i) = h_i(t_i)$, and for every $\theta_j \in \Theta_j^*(\theta_i^*)$ there is some $t_j \in T_j^*(t_i^*)$ with $h_j(\theta_j) = h_j(t_j)$. That is, we can find a mapping f_i: $\Theta_i^*(\theta_i^*) \to T_i^*(t_i^*)$ such that $h_i(\theta_i) = h_i(f_i(\theta_i))$ for all $\theta_i \in \Theta_i^*(\theta_i^*)$, and we can find a mapping $f_j : \Theta_j^*(\theta_i^*) \to T_j^*(t_i^*)$ such that $h_j(\theta_j) = h_j(f_j(\theta_j))$

for all $\theta_j \in \Theta_j^*(\theta_j^*)$. Moreover, as $h_i(\theta_i^*) = h_i(t_i^*)$, and any two different types in M^{co} induce different belief hierarchies, we necessarily have that $f_i(\theta_i^*) = t_i^*$.

We will now prove that these mappings f_i and f_j satisfy the conditions (1) and (2).

We will first prove condition (1). From Lemma 5.1 we know that any two types in $T_i \cup \Theta_i$ which induce the same m-th order belief, also induce the same first-order, second order, ... , $(m-1)$-th order belief on choices. As M^{co} and M^{in} contain only finitely many types, there must be some $m \in \mathbb{N}$ such that, for every two types $r_i, r_i' \in T_i \cup \Theta_i$,

$$h_i(r_i) = h_i(r_i') \text{ if and only if } h_i^m(r_i) = h_i^m(r_i'), \tag{14}$$

and similarly for player j. As, by assumption, any two different types in M^{co} induce different belief hierarchies, it holds that $h_i^m(t_i) \neq h_i^m(t_i')$ for any two different types $t_i, t_i' \in T_i$, and similarly for player j. Or, in other words, $T_i[h_i^m(t_i)] = \{t_i\}$ for all $t_i \in T_i$, where

$$T_i[h_i^m(t_i)] = \{t_i' \in T_i \mid h_i^m(t_i') = h_i^m(t_i)\},$$

and similarly for player j. Take some $\theta_i \in \Theta_i^*(\theta_i^*)$ and some $t_j \in T_j^*(t_i^*)$. Then, for every $c_j \in C_j$, we have that

$$\begin{aligned}
b_i(f_i(\theta))(c_j, t_j) &= b_i(f_i(\theta))(\{c_j\} \times T_j[h_j^m(t_j)]) \\
&= h_i^{m+1}(f_i(\theta))(c_j, h_j^m(t_j)) \\
&= h_i^{m+1}(\theta_i)(c_j, h_j^m(t_j)) \\
&= \beta_i(\theta_i)(\{c_j\} \times \Theta_j[h_j^m(t_j)]) \\
&= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j \mid h_j^m(\theta_j) = h_j^m(t_j)\}) \\
&= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid h_j^m(\theta_j) = h_j^m(t_j)\}) \\
&= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid h_j(\theta_j) = h_j(t_j)\}) \\
&= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid h_j(f_j(\theta_j)) = h_j(t_j)\}) \\
&= \beta_i(\theta_i)(\{c_j\} \times \{\theta_j \in \Theta_j^*(\theta_i^*) \mid f_j(\theta_j) = t_j\}) \\
&= \beta_i(\theta_i)(\{c_j\} \times f_j^{-1}(t_j))
\end{aligned}$$

which establishes condition (1). Here, the first equality follows from the fact that $T_j[h_j^m(t_j)] = \{t_j\}$, as we have seen above. The second equality follows from the definition of $h_i^{m+1}(f_i(\theta))$. The third equality follows from the fact that we have chosen f_i such that $h_i(f_i(\theta)) = h_i(\theta)$. The fourth equality follows from the definition of $h_i^{m+1}(\theta_i)$. The sixth equality follows from the fact that θ_i only assigns positive probability to j's types in $\Theta_j^*(\theta_i^*)$. The seventh equality follows from (14), which implies that $h_j^m(\theta_j) = h_j^m(t_j)$ if and only if $h_j(\theta_j) = h_j(t_j)$. The eighth equality follows from the fact that $h_j(\theta_j) = h_j(f_j(\theta_j))$, by construction of f_j. The ninth equality follows from the assumption that any two different types in T_j have different belief hierarchies.

Hence, condition (1) holds. In a similar way, one can prove that also condition (2) holds. The proof is hereby complete. $\square$

Proof of Lemma 5.4. Let V_i be the set of all utility functions for player i. Take an arbitrary utility function $v_i^* \in V_i$ such that choice c_i is optimal for the utility function v_i^* and the belief b_i. Let $M := d(v_i^*, u_i)$. Then, the set

$$V_i^* := \{v_i \in V_i \mid c_i \text{ optimal for } v_i \text{ and } b_i, \text{ and } d(v_i, u_i) \leq M\}$$

is closed and bounded, and hence compact. As the distance function $d(\cdot, u_i)$ is continuous on V_i, it follows from Weierstrass' theorem that $d(\cdot, u_i)$ takes a minimum on V_i^*. That is, there is some $v_i \in V_i^*$ with $d(v_i, u_i) \leq d(v_i', u_i)$ for all $v_i' \in V_i^*$.

We now show that there is only *one* $v_i \in V_i^*$ with this property. Suppose, on the contrary, that were would be two different utility functions $v_i, \hat{v}_i \in V_i^*$ with

$$d(v_i, u_i) = d(\hat{v}_i, u_i) \leq d(v_i', u_i) \text{ for all } v_i' \in V_i^*. \tag{15}$$

Then, it may easily be verified that also $\tilde{v}_i := \frac{1}{2}v_i + \frac{1}{2}\hat{v}_i$ is in V_i^*, and that $d(\tilde{v}_i, u_i) < d(v_i, u_i) = d(\hat{v}_i, u_i)$. This, however, contradicts (15). Hence, we conclude that there is a *unique* $v_i \in V_i^*$ with $d(v_i, u_i) \leq d(v_i', u_i)$ for all $v_i' \in V_i^*$. But then, this is also the unique $v_i \in V_i$ such that (a) c_i is optimal for the utility function v_i and the belief b_i, and (b) there is no other utility function $v_i' \in V_i$ with $d(v_i', u_i) < d(v_i, u_i)$ such that choice c_i is optimal for the utility function v_i' and the belief b_i. $\square$

Proof of Lemma 5.5. (Only if) Suppose first that $u_i(c_i, b_i) > u_i(c_i', b_i)$. We prove that $d(v_i[c_i, b_i], u_i) < d(v_i[c_i', b_i], u_i)$.

Consider the utility function $v_i' := v_i[c_i', b_i]$. We show that there is a utility function v_i with $d(v_i, u_i) = d(v_i', u_i)$ such that $v_i(c_i, b_i) > v_i(c_i'', b_i)$ for every $c_i'' \in C_i \backslash \{c_i\}$.

Remember that we fixed the choices c_i and c_i'. We define the utility function v_i by

$$v_i(c_i, c_j) := u_i(c_i, c_j) + v_i'(c_i', c_j) - u_i(c_i', c_j) \text{ for all } c_j \in C_j,$$
$$v_i(c_i', c_j) := u_i(c_i', c_j) + v_i'(c_i, c_j) - u_i(c_i, c_j) \text{ for all } c_j \in C_j,$$
$$v_i(c_i'', c_j) := v_i'(c_i'', c_j) \text{ for all } c_i'' \in C_i \backslash \{c_i, c_i'\} \text{ and all } c_j \in C_j.$$

Then, by construction,

$$\begin{aligned} d(v_i, u_i)^2 &= \sum_{c_i'' \in C_i} \sum_{c_j \in C_j} \left(v_i(c_i'', c_j) - u_i(c_i'', c_j)\right)^2 \\ &= \sum_{c_j \in C_j} \left(v_i(c_i, c_j) - u_i(c_i, c_j)\right)^2 + \sum_{c_j \in C_j} \left(v_i(c_i', c_j) - u_i(c_i', c_j)\right)^2 \\ &\quad + \sum_{c_j \in C_j} \sum_{c_i'' \in C_i \backslash \{c_i, c_i'\}} \left(v_i(c_i'', c_j) - u_i(c_i'', c_j)\right)^2 \\ &= \sum_{c_j \in C_j} \left(v_i'(c_i', c_j) - u_i(c_i', c_j)\right)^2 + \sum_{c_j \in C_j} \left(v_i'(c_i, c_j) - u_i(c_i, c_j)\right)^2 \\ &\quad + \sum_{c_j \in C_j} \sum_{c_i'' \in C_i \backslash \{c_i, c_i'\}} \left(v_i'(c_i'', c_j) - u_i(c_i'', c_j)\right)^2 \\ &= d(v_i', u_i)^2, \end{aligned}$$

which implies that $d(v_i, u_i) = d(v_i', u_i)$.

We next show that $v_i(c_i, b_i) > v_i(c_i'', b_i)$ for every $c_i'' \in C_i \backslash \{c_i\}$. Take some $c_i'' \in C_i \backslash \{c_i\}$. We distinguish two cases.

Case 1. If $c_i'' = c_i'$. Then,

$$\begin{aligned} v_i(c_i, b_i) - v_i(c_i', b_i) &= \left[u_i(c_i, b_i) + v_i'(c_i', b_i) - u_i(c_i', b_i)\right] - \left[u_i(c_i', b_i) + v_i'(c_i, b_i) - u_i(c_i, b_i)\right] \\ &= 2\left[u_i(c_i, b_i) - u_i(c_i', b_i)\right] + \left[v_i'(c_i', b_i) - v_i'(c_i, b_i)\right]. \end{aligned}$$

Recall our assumption that $u_i(c_i, b_i) > u_i(c_i', b_i)$. Moreover, since $v_i' = v_i[c_i', b_i]$, choice c_i' is optimal for b_i and v_i', and hence $v_i'(c_i', b_i) \geq v_i'(c_i, b_i)$. We thus conclude that $v_i(c_i, b_i) - v_i(c_i', b_i) > 0$, and hence $v_i(c_i, b_i) > v_i(c_i', b_i)$.

Case 2. If $c_i'' \in C_i \backslash \{c_i, c_i'\}$. Then,

$$\begin{aligned} v_i(c_i, b_i) - v_i(c_i'', b_i) &= \left[u_i(c_i, b_i) + v_i'(c_i', b_i) - u_i(c_i', b_i)\right] - v_i'(c_i'', b_i) \\ &= \left[u_i(c_i, b_i) - u_i(c_i', b_i\right] + \left[v_i'(c_i', b_i) - v_i'(c_i'', b_i)\right]. \end{aligned}$$

Recall that $u_i(c_i, b_i) > u_i(c_i', b_i)$. Moreover, as c_i' is optimal for b_i and v_i', it follows that $v_i'(c_i', b_i) \geq v_i'(c_i'', b_i)$. We thus conclude that $v_i(c_i, b_i) - v_i(c_i'', b_i) > 0$, and hence $v_i(c_i, b_i) > v_i(c_i'', b_i)$.

Summarizing, we see that $v_i(c_i, b_i) > v_i(c_i'', b_i)$ for all $c_i'' \in C_i \backslash \{c_i\}$. Since $u_i(c_i, b_i) > u_i(c_i', b_i)$, we know that c_i' is not optimal for b_i and u_i, which implies that $d(v_i', u_i) > 0$. As $d(v_i, u_i) = d(v_i', u_i) > 0$, and $v_i(c_i, b_i) > v_i(c_i'', b_i)$ for all $c_i'' \in C_i \backslash \{c_i\}$, we can find a utility function $\hat{v}_i$ with $d(\hat{v}_i, u_i) < d(v_i', u_i)$ such that c_i is optimal for b_i and $\hat{v}_i$. This implies that

$$d(v_i[c_i, b_i], u_i) \leq d(\hat{v}_i, u_i) < d(v_i', u_i) = d(v_i[c_i', b_i], u_i),$$

which was to show.

(If) Suppose now that $d(v_i[c_i, b_i], u_i) < d(v_i[c_i', b_i], u_i)$. We show that $u_i(c_i, b_i) > u_i(c_i', b_i)$.

Let $v_i := v_i[c_i, b_i]$. We construct the utility function v_i' by

$$v_i'(c_i, c_j) := u_i(c_i, c_j) + v_i(c_i', c_j) - u_i(c_i', c_j) \text{ for all } c_j \in C_j,$$
$$v_i'(c_i', c_j) := u_i(c_i', c_j) + v_i(c_i, c_j) - u_i(c_i, c_j) \text{ for all } c_j \in C_j,$$
$$v_i'(c_i'', c_j) := v_i(c_i'', c_j) \text{ for all } c_i'' \in C_i \backslash \{c_i, c_i'\} \text{ and all } c_j \in C_j.$$

Then, in the same way as above, it can be shown that $d(v_i', u_i) = d(v_i, u_i)$. Since $v_i = v_i[c_i, b_i]$, and $d(v_i[c_i', b_i], u_i) > d(v_i[c_i, b_i], u_i)$, the choice c_i' cannot be optimal for the belief b_i and the utility function v_i'. Hence, there must be some $c_i'' \in C_i \backslash \{c_i'\}$ such that $v_i'(c_i'', b_i) > v_i'(c_i', b_i)$. We distinguish two cases.

Case 1. If $c''_i = c_i$. Then, $v'_i(c_i, b_i) - v'_i(c'_i, b_i) > 0$. Hence,

$$\begin{aligned} v'_i(c_i, b_i) - v'_i(c'_i, b_i) &= \left[u_i(c_i, b_i) + v_i(c'_i, b_i) - u_i(c'_i, b_i)\right] - \left[u_i(c'_i, b_i) + v_i(c_i, b_i) - u_i(c_i, b_i)\right] \\ &= 2\left[u_i(c_i, b_i) - u_i(c'_i, b_i)\right] + \left[v_i(c'_i, b_i) - v_i(c_i, b_i)\right] > 0. \end{aligned}$$

As $v_i = v_i[c_i, b_i]$, we know that c_i is optimal for b_i and v_i, and hence $v_i(c'_i, b_i) - v_i(c_i, b_i) \leq 0$. We thus must have that $u_i(c_i, b_i) - u_i(c'_i, b_i) > 0$, implying that $u_i(c_i, b_i) > u_i(c'_i, b_i)$.

Case 2. If $c''_i \in C_i \backslash \{c_i, c'_i\}$. Then, $v'_i(c''_i, b_i) - v'_i(c'_i, b_i) > 0$. Hence,

$$\begin{aligned} v'_i(c''_i, b_i) - v'_i(c'_i, b_i) &= v_i(c''_i, b_i) - \left[u_i(c'_i, b_i) + v_i(c_i, b_i) - u_i(c_i, b_i)\right] \\ &= \left[u_i(c_i, b_i) - u_i(c'_i, b_i)\right] + \left[v_i(c''_i, b_i) - v_i(c_i, b_i)\right] > 0. \end{aligned}$$

Since $v_i = v_i[c_i, b_i]$, choice c_i is optimal for b_i and v_i, and hence $v_i(c''_i, b_i) - v_i(c_i, b_i) \leq 0$. It thus follows that $u_i(c_i, b_i) - u_i(c'_i, b_i) > 0$, which means that $u_i(c_i, b_i) > u_i(c'_i, b_i)$.

Summarizing, we can thus conclude that $u_i(c_i, b_i) > u_i(c'_i, b_i)$, which was to show. This completes the proof. □

References

Asheim, G.B., 2001. Proper rationalizability in lexicographic beliefs. Int. J. Game Theory 30, 453–478.
Böge, W., Eisele, T., 1979. On solutions of Bayesian games. Int. J. Game Theory 8, 193–215.
Dekel, E., Fudenberg, D., 1990. Rational behavior with payoff uncertainty. J. Econ. Theory 52, 243–267.
Friedenberg, A., Meier, M., 2011. On the relationship between hierarchy and type morphisms. Econ. Theory 46, 377–399.
Heifetz, A., Samet, D., 1998. Topology-free typology of beliefs. J. Econ. Theory 82, 324–341.
Mertens, J.-F., Zamir, S., 1985. Formulation of Bayesian analysis for games with incomplete information. Int. J. Game Theory 14, 1–29.
Myerson, R., 1978. Refinement of the Nash equilibrium concept. Int. J. Game Theory 7, 73–80.
Perea, A., 2012. Epistemic Game Theory: Reasoning and Choice. Cambridge University Press.
Perea, A., Kets, W., 2016. When do types induce the same belief hierarchy? Games 7, 28. http://dx.doi.org/10.3390/g7040028.
Schuhmacher, F., 1999. Proper rationalizability and backward induction. Int. J. Game Theory 28, 599–615.
Tan, T., Werlang, S.R.C., 1988. The Bayesian foundations of solution concepts of games. J. Econ. Theory 45, 370–391.