WINPEC Working Paper Series No.E1721
October 2017

Characterizing Perfect and Proper Rationalizabilities by Lexicographic Epistemic Model with Incomplete Information

Shuige Liuy

Waseda INstitute of Political EConomy
Waseda University
Tokyo,Japan

Characterizing Perfect and Proper Rationalizabilities by Lexicographic Epistemic Model with Incomplete Information*

Shuige Liu[†]

October 15, 2017

Abstract

We characterize perfect and proper rationalizabilities of 2-person games in incomplete information framework. We define the lexicographic model with incomplete information, and show that a choice is perfectly (properly) rationalizable in the complete information framework if and only if there is a corresponding belief hierarchy within the incomplete information framework that expresses common full belief in caution, opponent's utilities that primarily belief in the original utility function (centered around and as close as possible to the original one and to that belief hierarchy that choice is optimal).

JEL Classification: C72

Keywords: *epistemic game theory*, *lexicographic belief*, *perfect rationalizability*, *proper rationalizability*, *incomplete information*

1. Introduction

The purpose of noncooperative game theory is to study an individual's decision making in an interrelated situation. Since in such a situation, one's payoff is not completely determined by his own choice, to make a decision he needs to form a belief on every other participant's choice, each other participant's belief on others' choices, and so on. Studying the structures of those belief hierarchies and choices supported by a belief hierarchy satisfying some particular conditions opened up a field called *epistemic game theory* (cf. Perea [14]).

In epistemic game theory, various concepts have been developed to describe some specific belief structure. One is *lexicographic belief* (Blume et al. [3], [4]). A lexicographic belief describes an player's subjective conjecture about the game situation by a sequence of probability distributions over choice and states, which is different from the adoption of a single probability distribution in classic probabilistic belief. The interpretation is that every choice-state pairs in the sequence is considered likely, while a pair appearing ahead in the sequence is deemed *infinitely more likely* than one appearing later. Based on lexicographic belief various concepts are developed. Perfect and proper rationalizabilities are two important and interrelated concepts among them.

*The author would like to thank Andrés Perea, Funaki Yukihiko, Christian Bach, János Flesch, Dmitriy Kvasov, Abraham Nayman for their valuable comments and encouragement. She thanks all teachers and students in the 4th Epicenter Spring Course on Epistemic Game Theory in Maastricht University to whom she owes inspiring teaching, stimulating discussions, and new ideas. She gratefully acknowledge the support of Grant-in-Aids for Young Scientists (B) of JSPS No.17K13707 and Grant for Special Research Project No. 2017K-016 of Waseda University.

[†]Faculty of Political Science and Economics, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-Ku, 169-8050, Tokyo, Japan (shuige_liu@aoni.waseda.jp).

Perfect rationalizability is originated in Selten [18]'s perfect equilibrium. It is defined and studied from epistemic viewpoint of based on lexicographic belief in Blume et al. [4], Brandenburger [8], Börgers [6], and Asheim and Dufwenberg [2]. Perfect rationalizability is based on two notions: *caution* and *primary belief in the opponents' rationality.* A lexicographic belief is said to be cautious if it does not exclude any choice of opponents; it is said to primarily believe in the opponents' rationality if in its first level belief it only deems possible those choice-state pairs where the choice is optimal based on the belief of the state.

Proper rationalizability is originated in Myerson [13]'s perfect equilibrium which is intended to be a refinement of perfect equilibrium. It is defined and studied in Schuhmacher [17] and Asheim [1] as an epistemic concept. Proper rationalizability shares the same notion of caution with perfect rationalizability while uses a stronger notion called *respecting the opponents' preferences* to replace primary belief in the opponents' rationality. A lexicographic belief is said to respect the opponents' preferences if a "better" choice is always located in front of a "worse" choice in the sequence.

We explain these two concepts by an example. Consider a game with the following utility function u_2 for player 2:

u_2	D	E	F
A	3	2	1
B	3	2	1
C	3	2	1

Caution requires that all choices of player 2 appears in player 1's belief. Note that for player 2, choice D is most preferred no matter what belief he holds. Therefore, a lexicographic belief of player 1 primarily believes in 2's rationality must put A in its first order. Also, since D is preferred to E and E is preferred to F for player 2, a lexicographic belief of player 1 respecting 2's preferences should deem D infinitely more likely than E and E infinitely more likely than F, that is, put D before E and E before F in the sequence.

One purpose for the development of lexicographic belief is to alleviate the tension between caution and rationality (Blume et al. [3], Brandenburger [8], Börgers [6], Börgers and Samuleson [7]). The definitions of perfect and proper rationalities above tried to solve this tension by sacrificing rationality in different ways. That is, though perfect rationality requires that the level-1 belief contains only rational choices and proper rationality requires that choices should be ordered according to the level of their "irrationality", both allow the appearance of irrational choices. However, there is another approach which can solve that tension without hurting rationality. That is, instead of considering the uncertainty about opponents' rationality in complete information framework, we can take the uncertainty of the objective game and consider states in incomplete information framework. Then the appearance of a non-optimal choice can be explained from the uncertainty about which is the "real" game situation. Both perfect and proper rationalizabity should also be definable in this incomplete information framework. This is the basic idea of this paper.

We use the above example to explain this idea. As mentioned before, caution requires D, E, F appear in player 1's belief, while only D is a rational choice for 2. In complete information framework, the appearance of E and F are explained as a possibility of player 2's irrationality. In incomplete information framework, that can be explained by that the possibility that "real" utility function of player 2 is not u_2 but v_2 or v_2' as follows:

v_2	D	E	F
A	2	2	1
B	2	2	1
C	2	2	1

,

v_2'	D	E	F
A	2	2	2
B	2	2	2
C	2	2	2

Choice E is optimal in v_2 and F is optimal in v_2'. In this manner, uncertainty in the rationality of the opponents in a complete information framework is transferred into uncertainty in oneself about the real situations in incomplete information framework. It can be seen that primary belief in the opponent's rationality in complete information framework is equivalent to that one deems u_2 infinitely more likely to be the real situation than v_2 and v_2', and respecting the opponent's preferences requires that those alternative utility functions should be ordered by their "similarity" to u_2.

In this paper, we study these equivalences formally. We discuss we characterize perfect and proper rationalizabilities of 2-person games in incomplete information frameworks. We define the lexicographic model with incomplete information, which is the counterpart of probabilistic epistemic model with incomplete information defined by Dekel and Fudenberg [10]. We show that a choice is a choice is perfectly (properly) rationalizable in the complete information framework if and only if there is a corresponding belief hierarchy within the incomplete information framework that expresses common full belief in caution, opponent's utilities that primarily belief in the original utility function (centered around and as close as possible to the original one and to that belief hierarchy that choice is optimal).

This paper is not the first one considering characterizing concepts in epistemic game theory within incomplete information framework. Perea and Roy [16] characterized ε-proper rationalizability in this approach by using probabilistic model. They showed that a state in a (probabilitstic) epistemic model with complete information expresses full common belief in caution and ε-trembling condition if and only if there is a state in model with incomplete information which expresses common belief in caution, ε-centered belief around u, and belief in rationality under closest utility function. Since proper rationalizability is the limit of ε-proper ones, the conditions in their characterizations are invaluable for us. Two conditions used in our characterizition of proper rationalizability, that is, caution and u-centered belief, are faithful translations of their definitions into lexicographic model. However, as show in Section 2.1, the most critical condition in their characterization, that is, belief in rationality under closest utility function, is unable to be adopted here since the nearest utility function does not exist in lexicographic models in general. We defined a weaker condition called "belief under a closer utility function" and show that the three conditions characterize proper rationalizability.

Also, there are essential differences between Perea and Roy [16]'s proof and ours. Equivalence of belief hieracrchies generated by states in different models and type morphism (Böge and Eisele [5], Heifetz and Samet [12], Perea and Kets [15]) play important roles in Perea and Roy [16]'s proof. The If part in their proof starts from the assumption of equivalence of belief hierarchies and shows that conditions on a state with incomplete information implies proper conditions of states in complete information sharing the same belief hierarchy. Their Only-if part uses a construction and it works through belief hierarchy and type morphism. In contrast, our proofs are based on constructing a specific correspondence from one model to the other. We show that some conditions in states of one model implies some proper one in those of its corresponding model. Equivalence of hierarchies can be implied from the construction. In Section 4 we show our construction can also be used in proving Perea and Roy [16]'s Theorem 6.1.

Our results, as well as that of Perea and Roy [16] also provide insights in decision theory and general epistemology. They imply that any choice rationalizable in a complete information framework is also rationalizable in an incomplete one and vice versa. In other words, it is impossible to know the exact information structure of a game play by just looking at its outcome.

This paper is organized as follows. Section 2 gives definitions about lexicographic epistemic model with incomplete information and a survey of concepts in models with complete information. Section 3 gives the two characterization results and their proof. Section 4 gives some concluding remarks.

2. Models

2.1. Incomplete Information

In this subsection, we define the lexicographic model with incomplete information, which is the counterpart of probabilistic epistemic model with incomplete information defined by Dekel and Fudenberg [10]. We also define some concepts related to states in such a model. For a detailed introduction of lexicographic belief structure, see Perea [14], Chapter 4.

Definition 2.1 (Lexicographic epistemic model with incomplete information). Consider a finite 2-person static game form $G = (C_i)_{i \in I}$. For each $i \in I$, let V_i be the set of utility functions $v_i : C_1 \times C_2 \to \mathbb{R}$. A *finite lexicographic 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 each $\theta_i \in \Theta_i$ a utility function $w_i(\theta_i) \in V_i$, and

(c) β_i is a mapping that assigns to each $\theta_i \in \Theta_i$ a lexicographic belief over $\Delta(C_j \times \Theta_j)$, i.e., $\beta_i(\theta_i) = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})$ where $\beta_{ik} \in \Delta(C_j \times \Theta_j)$ for $k = 1, ..., K$.

Let $\theta_i \in \Theta_i$ with $\beta_i(\theta_i) = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})$. For $(c_j, \theta_j) \in C_j \times \Theta_j$, we say θ_i *deems* (c_j, θ_j) *possible* iff $\beta_{ik}(c_j, \theta_j) > 0$ for some $k \in \{1, ..., K\}$. We say θ_i *deems* $\theta_j \in \Theta_j$ *possible* iff θ_i deems (c_j, θ_j) possible for some $c_j \in C_j$. For each $\theta_i \in \Theta_i$, we use $\Theta_j(\theta_i)$ to denote the set of states in Θ_j deemed possible by θ_i. For each $\theta_i \in \Theta_i$ and $v_i \in V_i$, $\theta_i^{v_i}$ is the type satisfying that $w_i(\theta_i^{v_i}) = v_i$ and $\beta_i(\theta_i^{v_i}) = \beta_i(\theta_i)$. For $(c_j, \theta_j), (c'_j, \theta'_j) \in C_j \times \Theta_j$, we say that θ_i *deems* (c_j, θ_j) *infinitely more likely than* (c'_j, θ'_j) iff there exists $k \in \{0, ..., K-1\}$ such that the following two conditions are satisfied:

(a) $\beta_{it}((c_j, \theta_j)) = \beta_{it}((c'_j, \theta'_j)) = 0$ for $t = 1, ..., k$, and

(b) $\beta_{i,k+1}((c_j, \theta_j)) > 0$ and $\beta_{i,k+1}((c'_j, \theta'_j)) = 0$.

Definition 2.2 (Caution). $\theta_i \in \Theta_i$ is *cautious* iff for each $c_j \in C_j$ and each $\theta_j \in \Theta_j(\theta_i)$, there is some utility function $v_j \in V_j$ such that $\beta_i(\theta_i)$ deems $(c_j, \theta_j^{v_j})$ possible.

This is a faithful translation of Perea and Roy [16]'s definition of caution in probabilistic context (p.312) into lexicographic model. It is a parallel of the standard definition of a cautious type in lexicographic belief in complete information (cf. Perea [14], p.199) since both require that for each opponent's type θ_j deemed possible by θ_i, (c_j, θ_j) should also be deemed possible by θ_i for each $c_j \in C_j$; the only difference is that in incomplete information model we allow different utility function since c_j needs to be "supported" by some v_j (i.e., c_j needs to be optimal for the paired type).

For each $u_i, v_i \in V_i$, we define the distance $d(u_i, v_i)$ between u_i, v_i by $d(u_i, v_i) = [\Sigma_{c \in C}(u_i(c) - v_i(c))^2]^{1/2}$.

Definition 2.3 (Primary belief on u and u-centered belief). Consider a static game form $G = (C_i)_{i \in I}$, a lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ for G with incomplete information for G, and a pair $u = (u_i)_{i \in I}$ of utility functions.

(3.1) A type θ_i *primarily believes on* u if θ_i's level-1 belief only assigns positive probability to (c_j, θ_j) with $w_j(\theta_j) = u_j$.

(3.2) A type θ_i has *u-centered belief* if for any $c_j, c_{j'} \in C_j$, any $\theta_j \in \Theta_j$, and any $v_j, v'_j \in V_j$ such that $(c_j, \theta_j^{v_j})$ and $(c'_j, \theta_j^{v'_j})$ are deemed possible by θ_i, it holds that θ_i deems $(c_j, \theta_j^{v_j})$ infinitely more likely than $(c'_j, \theta_j^{v'_j})$ if and only if $d(v_j, u_j) < d(v'_j, u_j)$.

Definition 2.4 (3.2) is also a faithful translation of Perea and Roy [16]'s Definition 3.2 into lexicograhic model. It gives a regulation on the order of states in a lexicographic belief. That is, the farther a state's corresponding payoff function is from u, the later that state appears in the order of a lexicographic belief.

Definition 2.4 (Belief in rationality). Consider a static game form $G = (C_i)_{i\in I}$, a lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ for G with incomplete information. A type θ_i *believes in j's rationality* iff $\beta_i(\theta_i)$ deems (c_j, θ_j) possible only if c_j is optimal for θ_j.

The essential difference between our and Perea and Roy [16]'s conditions for characterization lies in the following definition.

Definition 2.5 (Belief under a closer utility function). Let $u = (u_i)_{i\in I}$ be a pair of utility functions. A type θ_i *has a belief under a closer utility function* iff $\beta_i(\theta_i)$ deems (c_j, θ_j) possible only if there is no $\theta'_j \in \Theta_j$ satisfying the following two conditions:

(a) $\beta_j(\theta_j) = \beta_j(\theta'_j)$ and $d(w_j(\theta'_j), u_j) \leq d(w_j(\theta_j), u_j)$

(b) $\beta_i(\theta_i)$ deems (c'_j, θ'_j) possible for some $c'_j \in C_j$ less preferred to c_j in u_j under $\beta_j(\theta_j)$.

We say a state *believes in rationality under a closer utility function* iff it has a belief in rationality and under a closer utility function. Definition 2.5 regulates on that what conditions are needed for a function (corresponding to a state) to be paired with a choice. It is requires here that a more "unrealistic" utility function (i.e., one that support c'_j) should be farther from u_i than a more "realistic" one (i.e., one that support c_j, for example, $w_j(\theta_j)$) in $\beta_i(\theta_i)$.

Definition 2.5 is a parallel to Perea and Roy [16]'s Definition 3.3, which requires that $w_j(\theta_j)$ to be the nearest utility function in V_j to u_j among those under which c_j is optimal. Lemma 5.5 in our Definition 2.5 is weaker than Perea and Roy [16]'s Definition 3.3. We adopt Definition 2.5 here since such a nearest utility function does not in general exist for lexicographic beliefs. That is, given $u_j \in V_j$, $c_j \in C_j$, and a lexicographic belief b_j, there may not exist $v_j \in V_j$ such that (1) c_j is optimal for v_j under b_j, and (2) there is no $v'_j \in V_j$ such that c_j is optimal for v'_j under b_j and $d(v'_j, u_j) < d(v_j, u_j)$. See the following example:

Example 2.1 (No nearest utility function). Consider the following game Γ :

u_1	D	E	F
A	1	1	1
B	1	1	0
C	1	0	1

Let $b_1 = (D, E, F)$. In u_1, A is optimal under b_1 but B is not. Now we show that there is no nearest utility function to u_1 under which B is optimal. Suppose there is such a function $v_1 \in V_1$. Let $d = d(v_1, u_1)$. It can be seen that $d > 0$. Consider the following v'_1 :

v'_1	D	E	F
A	1	1	1
B	$1 + \frac{d}{2}$	1	0
C	1	0	1

B is also optimal for v'_1 under b_1, while $d(v'_1, u_1) = \frac{d}{2} < d = d(v_1, u_1)$, which is a contradiction. Also, even though B is preferred to C under b_1, it can be seen that for each utility function v_1^B in which B is optimal under b_1, there is some $v_1^C \in V_1$ satisfying (1) C is optimal in v_1^C under b_1, and (2) $d(v_1^C, u_1) < d(v_1^B, u_1)$. Indeed, this can be done by letting $v_1^C(C, D) = 1 + d(v_1^B, u_1)/2$ and $v_1(c_1, c_2) = u_1(c_1, c_2)$ for all other choice pairs.

Example 2.1 shows that the relation between preferences of choices and the distance of their corresponding utility functions are more complicated in lexicographic beliefs. That is why we adopt Definition 2.5 here. The following lemma guarantees the existence of utility functions satisfying conditions in Definition 2.5. It plays the role of Lemma 5.5 in Perea and Roy [16] in our characterizations.

Lemma 2.1 (Weakened choice ranking lemma). Let $u_i \in V_i$, $b_i = (b_{i1}, b_{i2}, ..., b_{iK})$ be a lexicographic belief, and $C_i = \{c_i^1, ..., c_i^N\}$. Without loss of generality, we assume that $u_i(c_i^1, b_i) \geq u_i(c_i^2, b_i) \geq ... \geq u_i(c_i^N, b_i)$. Then there is a sequence $v_i^1, ..., v_i^N \in V_i$ satisfying

(a) $v_i^1 = u_i$,

(b) For each $n = 1, ..., N$, c_i^n is optimal for v_i^n under b_i, and

(c) For each $n = 1, ..., N-1$, $d(v_i^n, u_i) = d(v_i^{n+1}, u_i)$ whenever $u_i(c_i^n, b_i) = u_i(c_i^{n+1}, b_i)$, and $d(v_i^n, u_i) < d(v_i^{n+1}, u_i)$ whenever $u_i(c_i^n, b_i) > u_i(c_i^{n+1}, b_i)$.

Proof. We construct such a sequence by induction. First, let $v_i^1 = u_i$. For v_i^2, if $u_i(c_i^2, b_i) = u_i(c_i^1, b_i)$, then we just let $v_i^2 = v_i^1$. If $u_i(c_i^2, b_i) < u_i(c_i^1, b_i)$, then make some change in $\{u_i(c_i^1, c_j) : c_j \in \text{supp} b_{i1}\}$ to define $\{v_i^2(c_i^1, c_j) : c_j \in \text{supp} b_{i1}\}$ satisfying that $v_i^2(c_i^1, b_{i1}) < v_i^1(c_i^N, b_{i1})$ while keep other payoffs the same. It can be seen that c_i^2 is optimal for v_i^2 under b_i and $d(v_i^2, u_i) > 0 = d(v_i^1, u_i)$.

Now suppose for some $n \in \{1, ..., N-1\}$ we have defined $v_i^1, ..., v_i^n$ satisfying (a)-(c). We show how to define v_i^{n+1}. If $u_i(c_i^n, b_i) = u_i(c_i^{n+1}, b_i)$, we let $v_i^{n+1} = v_i^n$. Suppose that $u_i(c_i^n, b_i) > u_i(c_i^{n+1}, b_i)$. Let ${}_1^n c_i, ..., {}_1^N c_i$ be an order from the most preferred choice to the least preferred on in C_i under v_i^n with $c_i^{n+1} ={}_1^k c_i$ and $v_i^n({}_1^{k-1} c_i, b_i) > v_i^n({}_1^k c_i, b_i)$. We make some change in $\{v_i^n({}_1^t c_i, c_j) : t = 1, ..., k-1,\ c_j \in \text{supp} b_{i1}\}$ to define $\{v_i^{n+1}({}_1^t c_i, c_j) : t = 1, ..., k-1,\ c_j \in \text{supp} b_{i1}\}$ satisfying that $v_i^{n+1}({}_1^t c_i, b_{i1}) < v_i^n({}_1^N c_i, b_{i1})$ for all $t = 1, ..., k-1$ while keep other payoffs the same. It can be seen that c_i^{n+1} is optimal for v_i^{n+1} under b_i and, since $d(v_i^{n+1}, v_i^n) > 0$, $d(v_i^{n+1}, u_i) = d(v_i^{n+1}, v_i^n) + d(v_i^n, u_i) > d(v_i^n, u_i)$. By induction, we can construct a sequence $v_i^1, ..., v_i^N \in V_i$ satisfying (a)-(c). //

The basic idea behind this inductive construction is depicted as follows.

$$\left(c_i^1, c_i^2, c_i^3, ..., c_i^N\right) \ \underrightarrow{v_i^2} \ \left(c_i^2, c_i^3, ..., c_i^N, c_i^1\right) \ \underrightarrow{v_i^3} \ \left(c_i^3, ..., c_i^N, c_i^1, c_i^2\right) \ ... \ \underrightarrow{v_i^N} \ \left(c_i^N, c_i^1, ..., c_i^{N-1}\right)$$

We move choice one by one to the end of the sequence. When there is a tie, we just move the whole equivalence class to the end of the sequence. We use the following example to show how this construction works.

Example 2.1 (Continued). Consider u_1 in Example 2.1. Under $b_1 = (D, E, F)$, A is preferred to B and B is preferred to C in u_1. We can define v_1^1, v_1^2, v_1^3 as follows:

$u_1 = v_1^1$	D	E	F
A	1	1	1
B	1	1	0
C	1	0	1

$\longrightarrow$

v_1^2	D	E	F
A	$\frac{1}{2}$	1	1
B	1	1	0
C	1	0	1

$\longrightarrow$

v_1^3	D	E	F
A	$\frac{1}{2}$	1	1
B	$\frac{1}{4}$	1	0
C	1	0	1

It should be noted that $v_1^1, ..., v_1^N$ satisfying (a) - (c) are not unique, which will be seen in Example 3.1 and 3.4. Also, such a sequence can be seen as defined to equivalent classes of indifferent choices under a belief rather than to individual choices. This is important in constructing correspondence between states of complete and incomplete information models.

2.2. Complete Information

In this subsection we give a survey of lexicographic epistemic model with complete information. For details, see Perea [14], Chapters 5-6.

Definition 2.5 (Epistemic model). Consider a finite two-player static game $\Gamma = (C_i, u_i)_{i \in I}$. A finite *lexicographic 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 lexicographic belief over $\Delta(C_j \times T_j)$, i.e., $b_i(t_i) = (b_{i1}, b_{i2}, ..., b_{iK})$ where $b_{ik} \in \Delta(C_j \times T_j)$ for $k = 1, ..., K$.

For each $t_i \in T_i$, we denote by $T_j(t_i)$ the set of all t_j's deemed possible by t_i.

Definition 2.6 (Caution) Type $t_i \in T_i$ is *cautious* iff whenever it deems an $t_j \in T_j$ possible, then for every choice $c_j \in C_j$ it deems (c_j, t_j) possible.

Definition 2.7 (Primary belief in the opponent's rationality) Let $t_i \in T_i$ with $b_i(t_i) = (b_{i1}, b_{i2}, ..., b_{iK})$. t_i *primarily believes in* j*'s rationality* iff t_i's primary belief b_{i1} only assigns positive probability to those (c_j, t_j) where c_j is rational for t_j.

Definition 2.8 (Respect the opponent's preferences) Let $t_i \in T_i$ with $b_i(t_i) = (b_{i1}, b_{i2}, ..., b_{iK})$. t_i *respects the opponent's preferences* iff for any $t_j \in T_j(t_i)$ and any $c_j, c_j' \in C_j$ where t_j prefers c_j to c_j', t_i deems (c_j, t_j) infinitely more likely than (c_j', t_j).

Definition 2.9 (Common full belief in some property) Let P be an arbitrary property of lexicographic beliefs.

(a) $t_i \in T_i$ *expresses* 1*-fold full belief in* P iff t_i only deems possible j's types that satisfy P;

(b) For each $n \in \mathbb{N}$, $t_i \in T_i$ *expresses* $(n+1)$*-fold full belief in* P iff t_i only deems possible j's types that express n-fold full belief in P.

t_i *expresses common full belief in* P iff it expresses n-fold full belief in P for each $n \in \mathbb{N}$.

Definition 2.10 (Perfect and proper rationalizabilities). Given a lexicographic model $M^{co} = (T_i, b_i)_{i \in I}$ for a game $\Gamma = (C_i, u_i)_{i \in I}$. $c_i \in C_i$ is *perfectly rationalizable* iff it is optimal to some $t_i \in T_i$ which expresses common full belief in caution and primary belief in rationality. c_i is *properly rationalizable* iff it is optimal to some $t_i \in T_i$ which expresses common full belief in caution and respect of preferences.

3. Characterizations

3.1. Statements and Examples

Theorem 3.1 (Characterization of perfect rationality). Consider a finite 2-person static game $\Gamma = (C_i, u_i)_{i \in I}$, the corresponding game form $G = (C_i)_{i \in I}$, a finite lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for Γ.

Then, $c_i^* \in C_i$ is perfectly rationalizable in M^{co} if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G and some $\theta_i^* \in \Theta_i$ with $w_i(\theta_i^*) = u_i$ such that

(a) θ_i^* expresses common full belief in caution, primary belief on u, and rationality, and

(b) c_i^* is optimal to θ_i^*.

Theorem 3.2 (Characterization of proper rationality). Consider a finite 2-person static game $\Gamma = (C_i, u_i)_{i \in I}$, the corresponding game form $G = (C_i)_{i \in I}$, a finite lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for Γ.

Then, $c_i^* \in C_i$ is properly rationalizable in M^{co} if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ for G and some $\theta_i^* \in \Theta_i$ with $w_i^*(\theta_i^*) = u_i$ such that

(a) θ_i^* expresses common full belief in caution, u-centered belief, and rationality under a closer utility function

(b) c_i^* is optimal to θ_i^*.

To show these statements, we need to construct a correspondence between states of complete and incomplete information models. Before we go to the formal proof, we use the following three examples to show the intuition.

Example 3.1 (for Theorem 3.1). Consider the following game Γ (Perea [14], p.188):

$u_1 \backslash u_2$	C	D
A	$1,0$	$0,1$
B	$0,0$	$0,1$

and the lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ with complete information for Γ where $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, and

$$b_1(t_1) = ((D, t_2), (C, t_2)),\ b_2(t_2) = ((A, t_1), (B, t_1)).$$

It can be seen that t_1 expresses common full belief in caution and primary belief in rationality. We can construct a corresponding lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information where $\Theta_1 = \{\theta_{11}, \theta_{12}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}\}$, and

$$\begin{aligned} w_1(\theta_{11}) &= u_1,\ \beta_1(\theta_{11}) = ((D, \theta_{21}), (C, \theta_{22})), \\ w_1(\theta_{12}) &= v_1,\ \beta_1(\theta_{12}) = ((D, \theta_{21}), (C, \theta_{22})), \\ w_2(\theta_{21}) &= u_2,\ \beta_2(\theta_{21}) = ((A, \theta_{11}), (B, \theta_{12})), \\ w_2(\theta_{22}) &= v_2,\ \beta_2(\theta_{22}) = ((A, \theta_{11}), (B, \theta_{12})), \end{aligned} \tag{3.1}$$

where

v_1	C	D
A	0.5	0
B	0.5	0

,

v_2	C	D
A	0.5	0.5
B	0.5	0.5

It can be seen that θ_{11} expresses common full belief in caution, primary belief on u, and rationality. Also, θ_{11} generates the same belief hierarchy as t_1 does.

Example 3.2 (for Theorem 3.2). Consider the following game Γ (Perea [14], p.211):

$u_1 \backslash u_2$	D	E	F
A	$0, 5$	$1, 2$	$1, 1$
B	$1, 3$	$0, 4$	$1, 1$
C	$1, 3$	$1, 2$	$0, 3$

Consider the lexicographic model $M^{co} = (T_i, b_i)_{i \in I}$ for Γ where $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, and

$$b_1(t_1) = ((D, t_2), (F, t_2), (E, t_2)),\ b_2(t_2) = ((C, t_1), (B, t_1), (A, t_1)).$$

It can be seen that t_1 expresses common full belief in caution and respect of preferences. We can construct a corresponding lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for $G = (C_i)_{i \in I}$ where $\Theta_1 = \{\theta_{11}, \theta_{12}, \theta_{13}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}, \theta_{23}\}$, and

$$\begin{aligned} w_1(\theta_{11}) &= u_1,\ \beta_1(\theta_{11}) = ((D, \theta_{21}), (F, \theta_{22}), (E, \theta_{23})), \\ w_1(\theta_{12}) &= v_1,\ \beta_1(\theta_{12}) = ((D, \theta_{21}), (F, \theta_{22}), (E, \theta_{23})), \\ w_1(\theta_{13}) &= v_1',\ \beta_1(\theta_{13}) = ((D, \theta_{21}), (F, \theta_{22}), (E, \theta_{23})), \\ w_2(\theta_{21}) &= u_2,\ \beta_2(\theta_{21}) = ((C, \theta_{11}), (B, \theta_{12})), (A, \theta_{13})), \\ w_2(\theta_{22}) &= v_2,\ \beta_2(\theta_{22}) = ((C, \theta_{11}), (B, \theta_{12})), (A, \theta_{13})), \\ w_2(\theta_{23}) &= v_2',\ \beta_2(\theta_{23}) = ((C, \theta_{11}), (B, \theta_{12})), (A, \theta_{13})). \end{aligned}$$

where

v_1	D	E	F
A	0	1	1
B	1.1	0	1
C	1	1	0

,

v_1'	D	E	F
A	1	1	1
B	1	0	1
C	1	1	0

,

v_2	D	E	F
A	5	2	1
B	3	4	1
C	3	2	3.1

,

v_2'	D	E	F
A	5	2	1
B	3	4	1
C	2	3	3

,

It can be seen that θ_{21} expresses common belief in caution, u-centered belief, and rationality under a closer utility function and generates the same belief hierarchy as t_2.

We can also use this example to compare Theorems 3 2 and 3. Consider the lexicographic epistemic model $(T_i', b_i')_{i\in I}$ for Γ where $T_1' = \{t_1'\}$, $T_2' = \{t_2'\}$, and

$$b_1'(t_1') = ((A, t_2'), (B, t_2'), (C, t_2')),\ b_2'(t_2') = ((C, t_1'), (A, t_1'), (B, t_1')).$$

It can be seen that t_1' is expresses common full belief in caution and primary belief in rationality. We can construct the corresponding lexicographic epistemic model $M^{in} = (\Theta_i', w_i', \beta_i')_{i\in I}$ with incomplete information for $G = (C_i)_{i\in I}$ where $\Theta_1' = \{\theta_{11}', \theta_{12}', \theta_{13}'\}$, $\Theta_2 = \{\theta_{21}', \theta_{22}', \theta_{23}'\}$, and

$$\begin{aligned}
w_1'(\theta_{11}') &= u_1,\ \beta_1'(\theta_{11}') = ((D, \theta_{21}'), (E, \theta_{22}'), (F, \theta_{23}')),\\
w_1'(\theta_{12}') &= v_1',\ \beta_1'(\theta_{12}') = ((D, \theta_{21}'), (E, \theta_{22}'), (F, \theta_{23}')),\\
w_1'(\theta_{13}') &= v_1,\ \beta_1'(\theta_{13}') = ((D, \theta_{21}'), (E, \theta_{22}'), (F, \theta_{23}')),\\
w_2'(\theta_{21}') &= u_2,\ \beta_2'(\theta_{21}') = ((C, \theta_{11}'), (A, \theta_{12}')), (B, \theta_{13}')),\\
w_2'(\theta_{22}') &= v_2',\ \beta_2'(\theta_{22}') = ((C, \theta_{11}'), (A, \theta_{12}')), (B, \theta_{13}')),\\
w_2'(\theta_{23}') &= v_2,\ \beta_2'(\theta_{23}') = ((C, \theta_{11}'), (A, \theta_{12}')), (B, \theta_{13}')),
\end{aligned}$$

It can be seen that θ_{21}' expresses common full belief in caution, primary belief on u, and rationality, and generates the same belief hierarchy as t_2' does. On the other hand, it can be seen that t_2' does not respect 1's preferences, since under type t_1', B should be preferred to A, while t_2' deems A infinitely more likely than B. This is shown in the violation of rationality under a closer utility function of θ_{21}'. Indeed, $\beta_1'(\theta_{12}') = \beta_1'(\theta_{12}')$ and $d(w_1'(\theta_{12}'), u_1) = d(v_1', u_1) = 1 > d(w_1'(\theta_{13}'), u_1) = d(v_1, u_1) = 0.1$ while θ_{21}' deems (A, θ_{12}') infinitely more likely than (B, θ_{13}').

3.2. Proof of Theorem 3.1

To show the Only-if part of Theorem 3.1, we construct the following mapping from finite lexicographic epistemic models with complete information to those with incomplete information. Let $\Gamma = (C_i, u_i)_{i\in I}$ and $M^{co} = (T_i, b_i)_{i\in I}$ be a finite lexicographic epistemic model for Γ. We first define states in a incomplete information model in the following two steps:

Step 1. For each $i \in I$ and $t_i \in T_i$, define $\Pi_i(t_i) = (C_{i1}, ..., C_{i\ell})$ to be a partition of C_i satisfying that (1) for each $r = 1, ..., \ell$ and each $c_i, c_i' \in C_{ir}$, $u_i(c_i, b_i(t_i)) = u_i(c_i', b_i(t_i))$, and (2) for each $r = 1, ..., \ell - 1$ and $c_{ir} \in C_{ir}$, $c_{i,r+1} \in C_{i,r+1}$, $u_i(c_{ir}, b_i(t_i)) > u_i(c_{i,r+1}, b_i(t_i))$. That is, $\Pi_i(t_i)$ is a sequence of equivalent classes of choices in C_i ordered from the most to the least preferred under $b_i(t_i)$. By lemma 2.1, for each C_{ir} there is some $v_{ir}(t_i) \in V_i$ such that each choice in C_{ir} is optimal under $v_{ir}(t_i)$, and $0 = d(v_{i1}(t_i), u_i) < d(v_{i2}(t_i), u_i) < ... < d(v_{i\ell}(t_i), u_i)$.

Step 2. We define $\Theta_i(t_i) = \{\theta_{i1}(t_i), ..., \theta_{i\ell}(t_i)\}$ where for each $r = 1, ..., \ell$, the state $\theta_{ir}(t_i)$ satisfies that (1) $w_i(\theta_{ir}(t_i)) = v_{ir}(t_i)$, and (2) $\beta_i(\theta_{ir}(t_i))$ has the same distribution on choices at each level as $b_i(t_i)$ does, and for each (c_j, θ_j) corresponding to (c_j, t_j) in $b_i(t_i)$, $w_j(\theta_j)$ is the utility function $\{v_{jr}(t_j)\}$ for which c_j is the optimal under $b_i(t_i)$; in other words, if $c_j \in C_{jr} \in \Pi_j(t_j)$ for some r, then $\theta_j = \theta_{jr}(t_j)$.

Let $\Theta_i = \cup_{t_i\in T_i}\Theta_i(t_i)$. In this manner we construct a finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ with incomplete information for G. It should be noted that so far we did not put any restriction on t_i. In the following example we show how this construction works.

Example 3.4 (1). Consider the game Γ in Example 3.1:

$u_1 \backslash u_2$	C	D
A	$1,0$	$0,1$
B	$0,0$	$0,1$

and the lexicographic epistemic model $M^{co} = (T_i, b_i)_{i\in I}$ Γ where $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, and

$$b_1(t_1) = ((D, t_2), (C, t_2)),\ b_2(t_2) = ((A, t_1), (B, t_1)).$$

We show how to construct a corresponding model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$. First, by Step 1 it can be seen that $\Pi_1(t_1) = (\{A\}, \{B\})$ and $\Pi_2(t_1) = (\{D\}, \{C\})$. We can define $v_{11}(t_1) = u_1$, where A is optimal under $b_1(t_1)$, and $v_{12}(t_1)$ as follows, where B is optimal under $b_1(t_1)$. Similarly, we can define $v_{21}(t_2) = u_2$ and $v_{22}(t_2)$ as follows:

$v_{12}(t_1)$	C	D
A	0	-1
B	0	0

,

$v_{22}(t_2)$	C	D
A	0	-1
B	0	1

Then we go to Step 2. It can be seen that $\Theta_1(t_1) = \{\theta_{11}(t_1), \theta_{12}(t_1)\}$, where

$$\begin{aligned} w_1(\theta_{11}(t_1)) &= v_{11}(t_1),\ \beta_1(\theta_{11}(t_1)) = ((D, \theta_{21}(t_2)), (C, \theta_{22}(t_2))), \\ w_1(\theta_{12}(t_1)) &= v_{12}(t_1),\ \beta_1(\theta_{12}(t_1)) = ((D, \theta_{21}(t_2)), (C, \theta_{22}(t_2))). \end{aligned}$$

Also, $\Theta_2(t_2) = \{\theta_{21}(t_2), \theta_{22}(t_2)\}$, where

$$\begin{aligned} w_2(\theta_{21}(t_2)) &= v_{21}(t_2),\ \beta_2(\theta_{21}(t_2)) = ((A, \theta_{11}(t_1)), (B, \theta_{12}(t_1))), \\ w_2(\theta_{22}(t_2)) &= v_{22}(t_2),\ \beta_2(\theta_{22}(t_2)) = ((A, \theta_{11}(t_1)), (B, \theta_{12}(t_1))). \end{aligned}$$

(2) Since it can be seen that both t_1 and t_2 expresses common full belief in caution and primary belief in rationality, to show that our construction also works for general cases, consider the following states in a complete information model:

$$b_1(t_1') = ((C, t_2'), (D, t_2')),\ b_2(t_2') = ((A, t_1'), (B, t_1')).$$

First, by Step 1 it can be seen that $\Pi_1(t_1') = (\{A\}, \{B\})$ and $\Pi_2(t_1') = (\{D\}, \{C\})$. We can define $v_{11}(t_1') = u_1$, where A is optimal under $b_1(t_1')$, and $v_{12}(t_1')$ as follows, where B is optimal under $b_1(t_1')$. Similarly, we can define $v_{21}(t_2') = u_2$ and $v_{22}(t_2')$ as follows:

$v_{12}(t_1')$	C	D
A	-1	0
B	0	0

,

$v_{22}(t_2')$	C	D
A	0	-1
B	0	1

Then we go to Step 2. It can be seen that $\Theta_1(t_1') = \{\theta_{11}(t_1'), \theta_{12}(t_1')\}$, where

$$\begin{aligned} w_1(\theta_{11}(t_1')) &= v_{11}(t_1'),\ \beta_1(\theta_{11}(t_1')) = ((C, \theta_{22}(t_2')), (D, \theta_{21}(t_2'))), \\ w_1(\theta_{12}(t_1')) &= v_{12}(t_1'),\ \beta_1(\theta_{12}(t_1')) = ((C, \theta_{22}(t_2')), (D, \theta_{21}(t_2'))). \end{aligned}$$

Also, $\Theta_2(t_2') = \{\theta_{21}(t_2'), \theta_{22}(t_2')\}$, where

$$\begin{aligned} w_2(\theta_{21}(t_2)) &= v_{21}(t_2'),\ \beta_2(\theta_{21}(t_2')) = ((A, \theta_{11}(t_1')), (B, \theta_{12}(t_1'))), \\ w_2(\theta_{22}(t_2')) &= v_{22}(t_2'),\ \beta_2(\theta_{22}(t_2')) = ((A, \theta_{11}(t_1')), (B, \theta_{12}(t_1'))). \end{aligned}$$

It can be seen that, by construction, each t_i shares the same belief hierarchy (as can be defined in a similar way as in Perea and Roy [16]) with any $\theta_i \in \Theta_i(t_i)$. For each $t_i \in T_i$, we use $\theta_i(t_i)$ to denote the state in $\Theta_i(t_i)$ which has utility function u_i. It is clear that any c_i optimal to t_i is also optimal to $\theta_i \in \Theta_i(t_i)$ with $w_i(\theta_i) = u_i$. Then, to show the Only-if part of Theorem 3.1, we show that if t_i expresses common full belief in caution and primary belief in rationality, then $\theta_i(t_i)$ expresses common belief in caution, primary belief on u, and rationality. We need the following lemmas.

Lemma 3.1(Redundancy). Let $M^{co} = (T_i, b_i)_{i\in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ be constructed from M^{co} by the two steps above. Then for each $t_i \in T_i$ and each $\theta_i, \theta_i' \in \Theta_i(t_i)$, $\beta_i(\theta_i) = \beta_i(\theta_i')$.

Lemma 3.1 follows directly by the construction, so we eliminate its proof. It implies that the difference between any two states in a $\Theta_i(t_i)$ is their corresponding utility functions. This redundancy is critical in our proof.

Lemma 3.2 (Caution). Let $M^{co} = (T_i, b_i)_{i\in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ be constructed from M^{co} by the two steps above. If $t_i \in T_i$ expresses common full belief in caution, so does each $\theta_i \in \Theta_i(t_i)$.

Proof. We show this statement by induction. First we show that if t_i is cautious, then each $\theta_i \in \Theta_i(t_i)$ is also cautious. Let $c_j \in C_j$ and $\theta_j \in \Theta_j(\theta_i)$. By construction, it can be seen that $t_j \in T_j$ corresponding to θ_j (i.e., $\theta_j \in \Theta_j(t_j)$) is in $T_j(t_i)$. Since t_i is cautious, t_i deems (c_j, t_j) possible. Consider the pair (c_j, θ'_j) in $\beta_i(\theta_i)$ corresponding to (c_j, t_j). Since both θ_j and θ'_j are in $\Theta_j(t_j)$, it follows from Lemma 3.1 that $\beta_i(\theta_i) = \beta_i(\theta'_i)$. Hence it follows that $(c_j, \theta_j^{w_j(\theta'_j)})$ is deemed possible by θ_i. Here we have shown that θ_i is cautious.

Suppose we have shown that, for each $i \in I$, if t_i expresses n-fold full belief in caution then each $\theta_i \in \Theta_i(t_i)$ also expresses n-fold belief in caution. Now suppose that t_i expresses $(n+1)$-fold full belief in caution, i.e., it only deems possible j's types t_j's that express n-fold full belief in caution. It follows that each $\theta_j \in \Theta_j(t_j)$ also express n-fold full belief in caution. By construction, it follows that each $\theta_i \in \Theta_i(t_i)$ also expresses $(n+1)$-fold belief in caution. //

Lemma 3.3 (Rationality, and primary belief in rationality $\rightarrow$ Primary belief on u). Let $M^{co} = (T_i, b_i)_{i\in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ be constructed from M^{co} by the two steps above. Then

(a) each $\theta_i \in \Theta_i(t_i)$ expresses common full belief in primary belief on u and rationality, and

(b) if $t_i \in T_i$ expresses common full belief in primary belief in rationality, then each $\theta_i \in \Theta_i(t_i)$ expresses common full belief in primary belief on u.

Proof. (a) holds by construction. We show (b) statement by induction. First we show that if t_i primarily believes in j's rationality, then each $\theta_i \in \Theta_i(t_i)$ expresses full belief in primary belief on u_j. Let (c_j, θ_j) be a pair deemed possible in the level-1 belief of θ_i. Consider its corresponding (c_j, t_j) in level-1 belief of t_i. Since t_i primarily believes in j's rationality, c_j is optimal for t_i. It follows that $c_j \in C_{j1} \in \Pi_j(t_j)$. By Lemma 2.1 and construction, it follows that $w_j(\theta_j) = u_j$. Here we have shown that θ_i primarily believes on u_j.

Suppose we have shown that, for each $i \in I$, if t_i expresses n-fold full belief in primary belief in rationality then each $\theta_i \in \Theta_i(t_i)$ also expresses n-fold belief in primary belief on u. Now suppose that t_i expresses $(n+1)$-fold full belief in primary belief in rationality, i.e., it only deems possible j's types t_j's that express n-fold full belief in primary belief in rationality. It follows that each $\theta_j \in \Theta_j(t_j)$ also express n-fold full belief in primary belief on u. By construction, it follows that each $\theta_i \in \Theta_i(t_i)$ also expresses $(n+1)$-fold belief in primary belief on u. //

Proof of the Only-if part of Theorem 3.1. Let $M^{co} = (T_i, b_i)_{i\in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ be constructed from M^{co} by the two steps above, $c_i^* \in C_i$ be properly rationlizable, and $t_i^* \in T_i$ be a state supporting c_i. Let $\theta_i^* = \theta_i(t_i)$. By definition, $w_i(\theta_i^*) = u_i$, and c_i^* is optimal to θ_i^*. Also, it follows from Lemmas 3.2-3 that θ_i^* expresses common belief in caution, primary belief on u, and rationality. //

To show the If part, we need a mapping from models with incomplete information to those with complete information. Consider a finite 2-person static game $\Gamma = (C_i, u_i)_{i\in I}$, the corresponding game form $G = (C_i)_{i\in I}$, and a finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ for G. We construct $M^{co} = (T_i, b_i)_{i\in I}$ for Γ as follows. For each $\theta_i \in \Theta_i$, we define $E_i(\theta_i) = \{\theta'_i \in \Theta_i : \beta_i(\theta'_i) = \beta(\theta_i)\}$. In this manner, Θ_i is partitioned into some equivalent classes $\mathbb{E}_i = \{E_{i1}, ..., E_{i\ell}\}$. For each $E_i \in \mathbb{E}_i$ we assign a symbol $t_i(E_i)$ to represent a state. We define $b_i(t_i(E_i))$ to be a lexicographic belief on $\Delta(C_j \times T_j)$ which has the same distribution of choices at each level as any $\theta_i \in E_i$, and for each (c_j, t_j) corresponding to (c_j, θ_j) in $\beta_i(\theta_i)$,

$t_j = t_j(E_j(\theta_j))$.

Let $T_i = \{t_i(E_i)\}_{E_i \in \mathbb{E}_i}$. In this manner we construct a finite lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ with complete information for Γ from M^{in}. In the following example we show how this construction works.

Example 3.4 (Continued). Consider the game in Example 3.1:

$u_1 \backslash u_2$	A	B
A	$1, 0$	$0, 1$
B	$0, 0$	$0, 1$

and the lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ for $G = (C_i)_{i \in I}$ where $\Theta_1 = \{\theta_{11}, \theta_{12}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}\}$, and

$$\begin{aligned} w_1(\theta_{11}) &= u_1,\ \beta_1(\theta_{11}) = ((B, \theta_{21}), (A, \theta_{22})), \\ w_1(\theta_{12}) &= v_1,\ \beta_1(\theta_{12}) = ((B, \theta_{21}), (A, \theta_{22})), \\ w_2(\theta_{21}) &= u_2,\ \beta_2(\theta_{21}) = ((A, \theta_{11}), (B, \theta_{12})), \\ w_2(\theta_{22}) &= v_2,\ \beta_2(\theta_{22}) = ((A, \theta_{11}), (B, \theta_{12})). \end{aligned}$$

It can be seen that $\mathbb{E}_1 = \{\{\theta_{11}, \theta_{12}\}\}$ since $\beta_1(\theta_{11}) = \beta_1(\theta_{12})$. Similarly, $\mathbb{E}_2 = \{\{\theta_{21}, \theta_{22}\}\}$. Corresponding to those equivalent classes we have $t_1(\{\theta_{11}, \theta_{12}\})$ and $t_2(\{\theta_{21}, \theta_{22}\})$, and

$$\begin{aligned} b_1(t_{11}(\{\theta_{11}, \theta_{12}\})) &= ((B, t_2(\{\theta_{21}, \theta_{22}\})), (A, t_2(\{\theta_{21}, \theta_{22}\}))), \\ b_2(t_2(\{\theta_{21}, \theta_{22}\})) &= ((A, t_1(\{\theta_{11}, \theta_{12}\})), (B, t_1(\{\theta_{11}, \theta_{12}\}))). \end{aligned}$$

It can be seen that this construction can be seen as a reversion of the previous one. That is, given $M^{co} = (T_i, b_i)_{i \in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ be constructed from M^{co} by the previous two steps. Then $\mathbb{E}_i = \{\Theta_i(t_i)\}_{t_i \in T_i}$, and $t_i(\Theta_i(t_i)) = t_i$. We have the following lemmas.

Lemma 3.4 (Caution). Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ and $M^{co} = (T_i, b_i)_{i \in I}$ be constructed from M^{in} by the above approach. If $\theta_i \in \Theta_i$ expresses common full belief in caution, so does each $t_i(E_i(\theta_i))$.

Proof. We show this statement by induction. First we show that if θ_i is cautious, then $t_i(E_i(\theta_i))$ is also cautious. Let $c_j \in C_j$ and $t_j \in T_j(t_i(E_i(\theta_i)))$. By construction t_j corresponds to some $E_j \in \mathbb{E}_j$, and there is some $\theta_j \in E_j$ which is deemed possible by θ_i. Since θ_i is cautious, there is some θ'_j with $\beta_j(\theta'_j) = \beta_j(\theta_j)$, i.e., $\theta'_j \in E_j$, such that (c_j, θ'_j) is deemed possible by θ_i. By construction it follows that (c_j, t_j) is deemed possible by t_i.

Suppose we have shown that, for each $i \in I$, if θ_i expresses n-fold full belief in caution then $t_i(E_i(\theta_i))$ also expresses n-fold belief in caution. Now suppose that θ_i expresses $(n+1)$-fold full belief in caution, i.e., it only deems possible j's types θ_j's that express n-fold full belief in caution. It follows that $t_j(E_j(\theta_j))$ also express n-fold full belief in caution. By construction, it follows that $t_i(E_i(\theta_i))$ also expresses $(n+1)$-fold belief in caution. //

Lemma 3.5 (Primary belief on u + rationality$\rightarrow$ Primary belief in rationality). Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ and $M^{co} = (T_i, b_i)_{i \in I}$ be constructed from M^{in} by the above approach. If $\theta_i \in \Theta_i$ expresses common full belief in primary belief on u and rationality, then $t_i(E_i(\theta_i))$ expresses common full belief in primary belief in rationality.

Proof. We show this statement by induction. First we show that if θ_i primarily believes on u and j's rationality, then $t_i(E_i(\theta_i))$ primarily believes in j's rationality. Let (c_j, t_j) be a choice-type pair which is deemed possible in $t_i(E_i(\theta_i))$'s level-1 belief. By construction t_j corresponds to some $E_j \in \mathbb{E}_j$, and for some $\theta_j \in E_j$, (c_j, θ_j) is deemed possible in θ_i's level-1 belief. Since θ_i primarily believes on u, it follows that $w_j(\theta_j) = u_j$. Also, since θ_i believes j's rationality, it follows that c_j is optimal for u_j under the belief on choices of $\beta_j(\theta_j)$, i.e., that of $b_i(t_j)$. Therefore c_j is optimal for t_j. Here we have shown that $t_i(E_i(\theta_i))$ primarily believes in j's rationality.

Suppose we have shown that, for each $i \in I$, if θ_i expresses n-fold full belief in primary belief on u and rationality then $t_i(E_i(\theta_i))$ also expresses n-fold belief in primary belief in rationality. Now suppose that θ_i expresses $(n+1)$-fold full belief in primary belief on u and rationality, i.e., it only deems possible j's types θ_j's that express n-fold full belief in primary belief on u and rationality. It follows that $t_j(E_j(\theta_j))$ also express n-fold full belief in primary belief in rationality. By construction, it follows that $t_i(E_i(\theta_i))$ also expresses $(n+1)$-fold belief in primary belief in rationality. //

Proof of the If part of Theorem 3.1. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ and $M^{co} = (T_i, b_i)_{i\in I}$ be constructed from M^{in} by the above approach, $c_i^* \in C_i$ be optimal to some θ_i^* with $w_i(\theta_i^*) = u_i$ which expresses common belief in caution, primary belief on u, and rationality. Consider $t_i(E_i(\theta_i^*))$. By construction, c_i is optimal to $t_i(E_i(\theta_i^*))$. Also, it follows from Lemmas 3.4-5 that $t_i(E_i(\theta_i^*))$ expresses common full belief in caution and primary belief in rationality. Hence c_i^* is perfectly rationalizable in Γ. //

3.3. Proof of Theorem 3.2

To show the Only-if part of Theorem 3.2, we need the following lemma.

Lemma 3.6 (Respect of preferences $\rightarrow$ u-centered belief). Let $M^{co} = (T_i, b_i)_{i\in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ be constructed from M^{co} by the two steps in Section 3.1. If $t_i \in T_i$ expresses common full belief in respect of preferences, then each $\theta_i \in \Theta_i(t_i)$ expresses full belief in u-centered belief.

Proof. We show this statement by induction. First we show that if t_i respects j's preferences, then each $\theta_i \in \Theta_i(t_i)$ expresses u_j-centered belief. Let $c_j, c_j' \in C_j$, $\theta_j \in \Theta_j$, and $v_j, v_j' \in V_j$ such that $(c_j, \theta_j^{v_j})$ and $(c_j', \theta_j^{v_j'})$ be deemed possible by θ_i with $d(v_j, u_j) < d(v_j', u_j)$. By construction, it follows that (1) for $t_j \in T_j$ with $\theta_j^{v_j}, \theta_j^{v_j'} \in \Theta_j(t_j)$, t_i deems both (c_j, t_j) and (c_j', t_j) possible, and (2) $u_i(c_j, b_i(t_i)) > u_i(c_j', b_i(t_i))$. Since t_i respects j's preferences, t_i deems (c_j, t_j) infinitely more likely than (c_j', t_j). By construction, it follows that θ_i deems $(c_j, \theta_j^{v_j})$ more likely than $(c_j', \theta_j^{v_j'})$. Here we have shown that θ_i expresses u_j-centered belief.

Suppose we have shown that, for each $i \in I$, if t_i expresses n-fold full belief in respect of preferences then each $\theta_i \in \Theta_i(t_i)$ also expresses n-fold u-centered belief. Now suppose that t_i expresses $(n+1)$-fold full belief in respect of preferences, i.e., it only deems possible j's types t_j's that express n-fold full belief in respect of preferences. It follows that each $\theta_j \in \Theta_j(t_j)$ also express n-fold full belief in u-centered belief. By construction, it follows that each $\theta_i \in \Theta_i(t_i)$ also expresses $(n+1)$-fold belief in u-centered belief. //

Proof of the Only-if part of Theorem 3.2. Let $M^{co} = (T_i, b_i)_{i\in I}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ be constructed from M^{co} by the two steps Section 3.1, $c_i \in C_i$ be perfectly rationlizable, and $t_i^* \in T_i$ be a state supporting c_i. Let $\theta_i^* = \theta_i(t_i)$. By construction, θ_i^* believes in rationality under the closer utility function, $w_i(\theta_i^*) = u_i$, and c_i is optimal to θ_i^*. Also, it follows from Lemmas 3.2-3 and 6 that θ_i^* expresses common belief in caution, u-centered belief, and rationality under the closer utility function. //

To show the If part, we still use the construction from M^{in} to M^{co} defined in the previous subsection. We need the following lemma.

Lemma 3.7 (u-centered belief + rationality under the closer utility function $\rightarrow$ respect of preferences). Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in I}$ and $M^{co} = (T_i, b_i)_{i\in I}$ be constructed from M^{in} by the approach in Section 3.1. If $\theta_i \in \Theta_i$ expresses common full belief in u-centered belief and rationality under a closer utility function, then $t_i(E_i(\theta_i))$ expresses common full belief in respect of preferences.

Proof. We show this statement by induction. First we show that if θ_i has u-centered belief and believes in j's rationality under the closest utility function, then $t_i(E_i(\theta_i))$ respects j's preferences. Let $c_j, c_j' \in C_j$ and $t_j \in T_j(t_i(E_i(\theta_i)))$ where both (c_j, t_j) and (c_j', t_j) are deemed possible by $t_i(E_i(\theta_i))$ [1]and t_j prefers c_j to c_j'. By construction t_j corresponds to some $E_j \in \mathbb{E}_j$, and there are $\theta_j, \theta_j' \in E_j$ such that θ_i deems (c_j, θ_j) and (c_j', θ_j') possible. Suppose that θ_i deems (c_j', θ_j') infinitely more or the same likely than (c_j, θ_j). Since θ_i has u-centered belief, $d(w_j(\theta_j'), u_j) \leq d(w_j(\theta_j), u_j)$. Also, since θ_i believes j's rationality, c_j' is optimal to θ_j'. However, since c_j' is less preferred to c_j in u_j under $b_j(t_j)$ (i,e, by construction, $\beta_j(\theta_j)$), it is a contradiction to that θ_i believes in rationality under the closer utility function. Therefore θ_i deems (c_j, θ_j) infinitely more likely than (c_j', θ_j'), and, consequently $t_i(E_i(\theta_i))$ deems (c_j, t_j) infinitely more likely than (c_j', t_j). Here we have shown that $t_i(E_i(\theta_i))$ respects j's preferences.

Suppose we have shown that, for each $i \in I$, if θ_i expresses n-fold full belief in u-centered belief and rationality under a closer utility function, then $t_i(E_i(\theta_i))$ also expresses n-fold belief in respect of preferences. Now suppose that θ_i expresses $(n+1)$-fold full belief in primary belief on u and rationality, i.e., it only deems possible j's types θ_j's that express n-fold full belief in u-centered belief and rationality under a closer utility function. It follows that $t_j(E_j(\theta_j))$ also express n-fold full belief in primary belief in respect preferences. By construction, it follows that $t_i(E_i(\theta_i))$ also expresses $(n+1)$-fold belief in respect preferences. //

Proof of the If part of Theorem 3.2. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ and $M^{co} = (T_i, b_i)_{i \in I}$ be constructed from M^{in} by the above approach, $c_i^* \in C_i$ be optimal to some θ_i^* with $w_i(\theta_i^*) = u_i$ which expresses common belief in caution, u-centered belief, and rationality under a closer utility function. Consider $t_i(E_i(\theta_i^*))$. By construction, c_i is optimal to $t_i(E_i(\theta_i^*))$. Also, it follows from Lemmas 3.4 and 7 that $t_i(E_i(\theta_i^*))$ expresses common full belief in caution and respect of preference. Hence c_i^* is properly rationalizable in Γ. //

4. Concluding Remarks

4.1. Expression parallel to Perea and Roy [16]'s Theorem 6.1

Theorems 3.1 and 2 can be written as faithful parallels to Perea and Roy [16]'s Theorem 6.1 as follows, which focus on equivalence between states in complete and incomplete information models.

Theorem 4.1 (Characterizing perfect rationalizability). Consider a finite 2-person static game $\Gamma = (C_i, u_i)_{i \in I}$, the corresponding game form $G = (C_i)_{i \in I}$, the corresponding $u = (u_i)_{i \in I}$, a finite lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for Γ, and a type $t^* \in T_i$.

Then, $t_i^* \in T_i$ expresses common full belief in caution and primary belief in rationality if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ with incomplete information for G and some $\theta_i^* \in \Theta_i$ such that

(a) θ_i^* expresses common belief in caution, primary belief on u, and rationality, and

(b) t_i^* and θ_i^* has the same belief hierarchy.[2]

Theorem 4.2 (Characterizing proper rationalizability). Consider a finite 2-person static game $\Gamma = (C_i, u_i)_{i \in I}$, the corresponding game form $G = (C_i)_{i \in I}$, the corresponding $u = (u_i)_{i \in I}$, a finite lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for Γ, and a type $t^* \in T_i$.

Then, $t_i^* \in T_i$ expresses common full belief in caution and respect of preferences if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in I}$ for G and some $\theta_i^* \in \Theta_i$

[1]In our characterization, that both (c_j, t_j) and (c_j', t_j) are deemed possible by $t_i(E_i(\theta_i))$ is guaranteed by caution. Since here we did not assume that θ_i is cautious, we need this assumption.

[2]Belief hierarchy here can be defined in a similar way as in Perea and Roy [16], Section 4.

such that

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

(b) t_i^* and θ_i^* has the same belief hierarchy.

We adopt the expressions in Theorems 3.1 and 2 because (1) we want to take the viewpoint of choices since we believe that reasoning about choices is the starting point of epistemic game theory, and (2) the coincidence of belief hierarchy seems to be straightforward from our construction, hence we think it is not needed to mention it independent in the statements.

4.2. Extending to n-person cases

Though both Perea and Roy [16] and this note focus on 2-person games, there is no fundamental difficulty in extending those results into n-person cases. The only problem here is the distance between utility functions. In a 2-person game, a belief $\beta_i(\theta_i)$ of i needs only to consider distribution on $\Delta(C_j \times \Theta_j)$; a "cell" in $\beta_i(\theta_i)$ is just a pair (c_j, θ_j), and its location in b_i can be connected directly to the distance $d(w_j(\theta_j), u_j)$. On the other hand, in an n-person a "cell" of a belief contains more than one pair, e.g., something like $\langle (c_1, \theta_1), (c_2, \theta_2), ..., (c_{i-1}, \theta_{i-1}), (c_{i+1}, \theta_{i+1}), ..., (c_n, \theta_n) \rangle$, and more than one distance, e.g.,, $d(w_1(\theta_1), u_1), ..., d(w_n(\theta_n), u_n)$ here. Then how to connect the location of this cell and those distances? One direct way is to consider each player one-by-one. In this way, all definitions and conditions in 2-person cases can be directly applied to n-person cases, and the results in Perea and Roy [16] and this note still holds.

Another way is to define a distance on the whole cell. For example, to define $d(v_1, ..., v_n; u_1, ..., u_n) = \min_i\{d(v_i, u_i)\}$ seems still have the results in 2-person cases hold in general. However, it should be noted that not very distance works. For example, $\Sigma_i d(v_i, u_i)$ does not work since it does not differentiate distances of utility functions among different players. Further work is expected in this direction.

4.3. Using our construction to show Perea and Roy [16]'s Theorem 6.1

Our proofs are based on construction of a specific correspondence between two models. It can be seen that this construction can be translated directly into probabilistic context and be used to show Perea and Roy [16]'s Theorem 6.1. Further, it can be seen that, by using our Lemma 2.1, belief in rationality under closest utility function in Perea and Roy [16] can be replaced by the weaker one (Definition 2.5) here.

4.4. Characterizing other rationalities

As Perea and Roy [16] did, this characterizes some rationalities by epistemic model with incomplete information. It is wondered whether there is some other rationalizability concepts which can also be characterized in this way. However, it should be noted that using incomplete information model provides more insights when there is tension between caution and rationality, and the "levels" of irrationality matters (or, in words of conditional probability, when we need to consider conditional probabilities under impossible events. Cf. Blume et al. [3], [4], Halpern [11]). That is why characterization of ε-proper rationalizability in Perea and Roy [16] and that of proper rationalizability here interesting since they show that the "levels" of irrationality (i.e., the order of preferences) in complete information models can be replaced by distances of "unrealistic" payoff functions from the original one in incomplete information models with rationality.

In this sense, we believe that some rationalizabiliy concepts may not be so suitable to be characterized by incomplete information models, even though theoretically it can be. For example, states assuming rationality (Brandenburger et al. [9]) seems not suitable since the "levels" of irrationality does not matter there. Actually, for each t_i cautious, rational, and assuming the opponent's rationality, we can construct a state θ_i as a faithful "copy" of t_i, which automatically satisfies caution, rationality, and assumption of opponent's rationality.

References

[1] Asheim, G.B., (2001). Proper rationalizability in lexicographic beliefs. *International Journal of Game Theory* **30**, 453-478.

[2] Asheim, G.B., Dufwenberg, M. (2003). Admissibility and common belief. *Games and Economic Behavior* **42**, 208-234.

[3] Blume, L., Brandenburger, A., Dekel, E. (1991). Lexicographic probabilities and choice under uncertainty. *Econometrica* 59, 61-79.

[4] Blume, L., Brandenburger, A., Dekel, E. (1991). Lexicographic probabilities and equilibrium refinements. *Econometrica* 59, 81-98.

[5] Böge, W., Eisele, T. (1979). On solutions of Bayesian games. *International Journal of Game Theory* **8**, 193-215.

[6] Börgers, T. (1994). Weak dominance and approximate common knowledge. *Journal of Economic Theory* **64**, 265-276.

[7] Börgers, T. and Samuelson, L. (1992). "Cautious" utility maximization and iterated weak dominance. *International Journal of Game Theory* **21**, 13-25.

[8] Brandenburger, A. (1992). Lexicographic probabilities and iterated admissibility. In *Economic Analysis of Markets and Games*, ed. P. Dasgupta, et al. MIT Press, 282-290.

[9] Brandenburger, A., Friedenberge, A., Keisler, J. (2008). Admissibility in games. *Econometrica* **76**, 307-352.

[10] Dekel, E., Fudenberg, D. (1990). Rational behavior with payoff uncertainty. *Journal of Economic Theory* **52**, 243-267.

[11] Halpern, J.Y. (2010). Lexicographic probability, conditional probability, and nonstandard probability. *Games and Economic Behavior* **68**, 155-179.

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

[13] Myerson, R.B. (1978). Refinements of Nash equilibrium concept. *International Journal of Game Theory* **7**, 73-80.

[14] Perea, A. (2012). *Epistemic Game Theory: Reasoning and Choice*, Cambridge University Press.

[15] Perea, A., Kets, W. (2016). When do types induce the same belief hierarchy? *Games* **7**, 28. http://dx.doi.org/10.3390/g7040028.

[16] Perea, A., Roy, S. (2017) A new epistemic characterization of ε-proper rationalizability. *Games and Economic Behavior* **104**: 309-328.

[17] Schuhmacher, F. (1999). Proper rationalizability and backward induction. *International Journal of Game Theory* **28**, 599-615.

[18] Selten, R. (1978). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **4**, 25-55.