早稲田大学現代政治経済研究所

Infinite Regresses Arising from Prediction/Decision Making in Games

Tai-Wei Hu and Mamoru Kaneko

No.E1308

Working Paper Series

Institute for Research in

Contemporary Political and Economic Affairs

Waseda University

169-8050 Tokyo,Japan

Infinite Regresses Arising from Prediction/Decision Making in Games *,†

Tai-Wei Hu‡and Mamoru Kaneko§

July 3, 2013

Working Paper No.E1308
Institute for Research in Contemporary Political and Economic Affairs,
Waseda University,

Abstract

We study infinite regresses arising from prediction/decision-making in an n-person game, taking beliefs and inferences explicitly into account. We adopt epistemic logic GL_ω, which is an infinitary extension of KD^n, in order to facilitate our discussions on infinite regresses. As a KD-type epistemic logic, we can distinguish between subjectivity and objectivity. In GL_ω, we formulate an infinite regress as a formula, and derive its basic properties. In particular, we show that when we add Axiom T (truthfulness) to GL_ω, our concept of infinite regress, which is defined in a subjective manner, collapses to the common knowledge concept, which is an objective concept. Then we give an epistemic axiomatization of the Nash (noncooperative) solution theory using our concept of infinite regress. Under the assumption of interchangeability, we obtain a full characterization of possible final decisions and predictions for the Nash theory.
Key words: Infinite Regress, Infinitary Epistemic Logic, Common Knowledge, Nash Equilibrium, Solution

---

*The authors are grateful for useful discussions with Nobu-Yuki Suzuki.

†The authors are partially supported by Grant-in-Aids for Scientific Research No.21243016, and No.2312002, Ministry of Education, Science and Culture.

‡MEDS Department, Northwestern University, Evanston, IL, (t-hu@kellogg.northwestern.edu)

§Faculty of Political Science and Economics, Waseda University, Tokyo 169-8-50, Japan (mkanekoepi@waseda.jp)

1. Introduction

1.1. An Infinite Regress Arising from Prediciton/Decision Making

A foundational study on human thinking may lead to some form of regresses and sometimes infinite regresses. A variety of such regresses and/or infinite arguments are discussed in Gratton [5]. Game theory have treated human thinking particularly in interactive situations with multiple players (persons), and has encountered regresses and infinite regresses specific to interactions of players. Nevertheless, the classical game theoretic language is incapable of studying such infinite regresses in a faithful manner. In this paper, we provide an apparatus suitable to such a study, and apply it to a specific game theory problem.

The problem we deal with may be regarded as an instance of a general notion of infinite regresses, but its basic context contains a specific structure in the sense of game theory. Thus, we should explain the nature of our problem. For the our development, we take three steps:

IR0: we discuss the infinite regresses arising from prediction/decision making in an interactive situation;

IR1: we formulate the concept of "infinite regress" in an infinitary epistemic logic GL_ω;

IR2: we apply this formalization to the game theoretic prediction/decision making in IR0.

Step IR0 may be described as the term "regress formula" due to Gratton [5]. Step IR1 is a general development of the specific formulae and their properties, and Step IR2 is an analysis of IR0 using the apparatus developed in IR1. We shall elaborate on these three steps below.

Game theory studies interactions between multiple players, and each player's payoff depends not only on his own decision, but also on other players' decisions. Therefore, in the *ex ante* decision making, meaning that players' decisions are made before actual plays, each player makes a decision together with predictions about the other players' decisions. Thus, in game theory, decision making is more accurately described as *prediction/decision making.*

Table 1.1

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	$(5,5)$	$(-2,-2)$	$(-2,-2)$
$\mathbf{s}_{12}$	$(-2,-2)$	$(7,-1)$	$(-1,7)$
$\mathbf{s}_{13}$	$(-2,-2)$	$(-1,7)$	$(7,-1)$

This prediction/decision-making involves an infinite regress of interpersonal beliefs, which we illustrate by using the 2-person game of Table 1.1: In this game, player 1's best

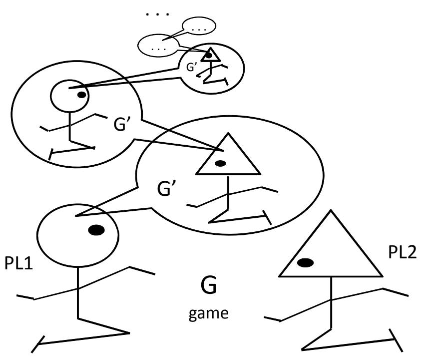

Figure 1.1: Regress from Prediction/Decision Making

strategy (in terms of payoff-maximization) depends on player 2's decision, but 2's decision depends upon 1's decision. Thus, this situation contains the structure of regression. It may lead either to an equilibrium (circular concept) or to an infinite regress. With an appropriate distinction between one's and the other's thinking, however, it would be faithful to have an infinite regress rather than an equilibrium: "Player 1 thinks about his decision, which requires player 2's thinking about 1's decision, which requires 1's thinking about 2' decision, ..." This process is illustrated in Fig.1.1.

In the game theory literature, however, an explicit treatment of this infinite regress is lacking. The closest concept is "common knowledge" due to Lewis [13] and Aumann [1]. Although a few papers consider epistemic aspects of solution concepts in game theory (Kaneko [9], Aumann and Brandenburger [2], Tan and Werlang [19]), existing approaches using common knowledge do not directly target the above infinite regress. In contrast, we explicitly formulate an infinite regress of interpersonal beliefs and apply the infinite regress to study the theory of *ex ante* prediction/decision-making proposed by Nash [17].

The above *ex ante* prediction/decision-making may be described by the following two statements:

$\mathrm{N1}^o$: player 1 choose his best strategy against all of his predictions about player 2's choice based on $\mathrm{N2}^o$;

$\mathrm{N2}^o$: player 2 choose his best strategy against all of his predictions about player 1's choice based on $\mathrm{N1}^o$.

A possible final decision for player 1 is (intended to be) determined by $\mathrm{N1}^o$, but needs a prediction about player 2's possible final decisions, which is determined by $\mathrm{N2}^o$ included in $\mathrm{N1}^o$. The symmetric form $\mathrm{N2}^o$ determines a decision for player 2 with a prediction about 1's decision. With the distinction between decisions and predictions, these lead to an infinite regress unless we stop at an arbitrary level.

Because the classical game theory language is incapable of having an explicit distinction between decisions and predictions, N1^o-N2 ^{o} is regarded as a circular definition of decisions and predictions, i.e., as an equilibrium. It is known that if we formulate N1^o-N2^o in the classical language, they characterize the Nash noncooperative theory due to Nash [17]. This is illustrated in Diagram 1.1, where the regress of interpersonal beliefs is hidden in a circular form and the hierarchy in Fig.1.1 is crushed into one layer.

Thanks to the recent development in epistemic logic (e.g., Fagin, *et al.* [4], Meyer-van der Hoek [15], Kaneko-Nagashima [8]), we can explicitly distinguish between decisions and predictions by means of belief operators in an adequately chosen epistemic logic system[1]. In the mind of player 1, he can think about his decision making based on N1^o, and about his prediction about 2's based on N2^o. This is illustrated in the leftmost column of Diagram 1.2. These are now separated and are the starting points in Diagram 1.2. Applying the same argument to the imagined player 2, we go to the top of the second column, and then go to its bottom, and so on. Here we meet an infinite regress consisting of predictions/decisions, rather than a circular argument.

Diagram 1.1

N1^o
$\downarrow\uparrow$
N2^o

Diagram 1.2

	N1^o		B_2B_1(N1^o)		$\cdots$
B_1	$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$
	B_2(N2^o)		$B_2B_1B_2$(N2^o)		$\cdots$

It is important to remark that this infinite regress is derived from *logical completion* in the argument of prediction/decision making, rather than from the claim that such a situation is well observed in reality. Because our target is an infinite regress of interpersonal beliefs, it is natural to use an infinitary logic system, which will be discussed in the next subsection.

1.2. Infinitary Logic Approach

We adopt the two research principles for the choice of a system of epistemic logic:

R1: it is capable to make a distinction between subjectivity and objectivity;

R2: the concept of infinite regress can be formulated in an explicit manner.

We adopt principle R1 because the concept of infinite regress described in Diagram 1.2 as well as in Fig.1.1 involves a subjective structure: The essential part of this structure requires a clear-cut distinction between subjectivity and objectivity.

[1] See Benthem, *et al* [22] and Benthem [21] for an overview on recent developments on logic and game theory.

Based on R1 and R2, we adopt an infinitary extension of epistemic logic KD^n, which is the system GL_ω given by Kaneko-Nagashima [8] (precisely speaking, Kaneko-Nagashima [8] gives an infinitary predicate version of $\mathrm{KD4}^n$). We drop both Axiom T (Truthfulness) and Axiom 4 (Positive Introspection) from our logical system to relativize and clarify their roles in our formulation of the infinite regress of beliefs.

In GL_ω, we formulate the formula expressing an infinite regress of the above sort. In the 2-person case, for two given formulae, (A_1, A_2), the infinite regress derived from (A_1, A_2) and from the viewpoint of player 1, is expressed as

$$\mathbf{Ir}_1(A_1, A_2) = \wedge\{A_1, \mathbf{B}_2(A_2), \mathbf{B}_2\mathbf{B}_1(A_1), ...\}, \tag{1.1}$$

where $\mathbf{B}_i$ is the belief operator for player i. Notice that by plugging $\mathrm{N1}^o$ and $\mathrm{N2}^o$ to A_1 and A_2, (1.1) becomes the conjunction of all the formulae in Diagram 1.2. A general formulation for the n-person case is given in Section 2. In our formulation, the concept of infinitary regress is a subjective notion in the sense that $\mathbf{Ir}_1(A_1, A_2)$ describes the content of the infinite regress occurring in the mind of player 1. We give Lemma 3.3 to explain this inner mental activities of an individual player i.

A study of the infinite regress formula $\mathbf{Ir}_1(A_1, A_2)$ is step IR1 in Section 1.1, rather than a study of infinite regress per se, since A_1 and A_2 are still arbitrary and may require no regression. Nevertheless, we should study general properties of the infinite regress formula in (1.1) to facilitate more developments.

Our approach clarifies subtle relationships between the infinite regress and common knowledge, together with the other epistemic axioms. First, we show that under Axiom T, our infinite regress formula $\mathbf{Ir}_i(A_1, A_2)$ collapses to the common knowledge formula $\mathbf{C}(A_1 \wedge A_2)$ (with respect to provability), which is an epistemic description of an objective situation (cf., Lewis [13], and also Fagin *et al.* [4], Myer-van der Hoek [15]). Thus, imposing Axiom T would remove the subjective nature of the infinite regress formula. We also show that when Axiom 4 is added to GL_ω, the infinite regress formula becomes equivalent to the individual belief of common knowledge, but not the *common beliefs.*

In step IR2, we apply our theory to the prediction/decision making for a game mentioned in Section 1.1. Specifically, the infinite regress in Diagram 1.2 is expressed as $\mathbf{Ir}_1(\mathrm{N1}^o, \mathrm{N2}^o)$. This allows a formal analysis of the regress $\mathbf{Ir}_1(\mathrm{N1}^o, \mathrm{N2}^o)$ and a full characterization of possible final decisions and predictions. The characterization leads to a particular formula taking the infinite regress of best responses. Under Axiom T, this formula becomes the common knowledge of the Nash equilibrium. Our result shows that our formulation of infinite regress completes the epistemic analysis of *ex ante* prediction/decision-making in the Nash theory. It is worth noting that the characterization requires a game theoretical assumption, *interchangeability,* due to Nash [17][2]. In

[2] Without interchangeability, we would meet an incompleteness result germinating from the "unsolvability" in the Nash [17] sense. Its full argument will be given in a separate paper.

this sense, our study of infinite regress is intrinsic only not in an epistemic logic but also in game theory.

1.3. Three Remarks on Our Approach

First, we emphasize that GL_ω is taken as a proof-theoretic system, though it is Kripke-complete. A reason for this choice is also compatible with principle R2; the logical inferences are an engine for prediction/decision making. In general, we take the proof-theoretic approach for positive statements, and the semantic approach for negative (unprovable) statements by giving counter models.

In related strands of literatures, model-theoretic approaches in a broad sense are adopted. In logic, the Kripke semantics itself is regarded as targets as well as representing tools, while syntactical axiomatizations are also presented (cf., Fagin *et al.* [4], Meyer-van der Hoek [15]). In related game theory literature such as Aumann-Brandenburger [2] and Tan-Werlang [19], the main focus is the study of the "states of mind" of the players, though common knowledge as an objective description is the main theoretical concept. In contrast to those papers, we adopt a syntactical approach since our focus is on players' mental activities, i.e., logical inferences.

Finally, we remark that we can formulate our problem in a fixed-point point extension of KD^n. This is less direct than the infinitary logic approach since the concept of infinite regress is, by definition, of infinite nature. In Section 3, however, we will indicate how the fixed-point approach can be formulated. Also, we note that our infinitary system GL_ω is a small infinitary extension of KD^n, rather than a large infinitary extension such as ones in Karp [12] and in Heifetz [7].

The paper is organized as follows: Section 2 gives a Hilbert-style formulation of GL_ω and some basic lemmas. The infinite regress formula $\mathbf{Ir}_i(A_1, ..., A_n)$ and the common knowledge formula $\mathbf{C}(A)$ are formulated there. Section 3 gives lemmas summarizing basic properties of infinite regress formulae. Section 4 gives two theorems on the relationships between our infinite regress and common knowledge formulae. Section 5 formulates axioms for prediction/decision making and the final axiomatic forms in terms of infinite regresses, together with their analysis and presenting the characterization theorems. Section 6 gives conclusions and some remarks. Section 7 gives proofs of the two theorems in Section 5.

2. Epistemic Logic

We give an infinitary language for our epistemic logic, and formulate the concept of an infinite regress. To analyze its behavior, we adopt infinitary epistemic logic GL_ω, and give a concise introduction to it.

2.1. Infinitary Language

We start with the following list of primitive symbols:

propositional variables: $\mathbf{p}_0, \mathbf{p}_1 ...$;

logical connective symbols: $\neg$ (not), $\supset$ (implies), $\wedge$ (and), $\vee$ (or)[3];

unary belief operators: $\mathbf{B}_1(\cdot)$, ..., $\mathbf{B}_n(\cdot)$;

parentheses: (,); *comma*: ,; and *braces*: {, }.

The last list of parentheses, commas, and braces will be used since set-theoretical expressions are allowed in our definition of formulae.

Diagram 2.1

$\mathcal{P}_0 = PV$		$\mathcal{P}_1$	$\rightarrow$	$\mathcal{P}_2$	$\rightarrow$	$\mathcal{P}_3$	$\rightarrow \cdots$
$\downarrow$		$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$
$\mathcal{F}(\mathcal{P}_0) = \emptyset$		$\mathcal{F}(\mathcal{P}_1)$		$\mathcal{F}(\mathcal{P}_2)$		$\mathcal{F}(\mathcal{P}_2)$	

We define the set $\mathcal{P}_k$ of *formulae* and a certain set $\mathcal{F}(\mathcal{P}_k)$ of countable subsets of $\mathcal{P}_k$ for k $(0 \leq k < \omega)$ by the following induction. This definition has two induction steps: For each k, we define $\mathcal{P}_{k+1}$ and $\mathcal{F}(\mathcal{P}_{k+1})$ from $\mathcal{P}_k$ and $\mathcal{F}(\mathcal{P}_k)$; and each definition is based on some induction. These steps are are illustrated in Diagram 2.1. Formulae $\mathcal{P}_1$ are all finitary, but for $k > 1$, $\mathcal{P}_k$ includes infinitary formulae.

The base case for $k = 0$ is as follows:

(o): $\mathcal{P}_0 := PV$ is the set of all propositional variables, and $\mathcal{F}(\mathcal{P}_0) = \emptyset$.

Now we suppose that $\mathcal{P}_k$ and $\mathcal{F}(\mathcal{P}_k)$ are given. Then we define $\mathcal{P}_{k+1}$ by the following finite induction:

(i): $\mathcal{P}_k \cup \{(\wedge\Phi), (\vee\Phi) : \Phi \in \mathcal{F}(\mathcal{P}_k)\} \subseteq \mathcal{P}_{k+1}$;

(iia): if A, C are formulae $\mathcal{P}_{k+1}$, so are $(A \supset C)$, $(\neg A)$, $\mathbf{B}_1(A), ..., \mathbf{B}_n(A)$;

(iib): if Φ is a finite nonempty subset of $\mathcal{P}_{k+1}$, so are $(\wedge\Phi)$ and $(\vee\Phi)$.

Now we define the set $\mathcal{F}(\mathcal{P}_{k+1})$: We first define *permissible sequences* from $\mathcal{P}_{k+1}$ as follows: Let $D_1(p_1, ..., p_l), ..., D_l(p_1, ..., p_l)$ be formulae in $\mathcal{P}_1$ with a specification of propositional variables $p_1, ..., p_l$ (l is a natural number) and $C_1, ..., C_l$ formulae in $\mathcal{P}_{k+1}$. Then, the sequences $\langle C_t^\nu : \nu \geq 0 \rangle, t = 1, ..., l$ are $(k+1)$-permissible if they are generated by the following recursion:

A0: $C_t^0 = C_t$ for $t = 1, ..., l$; and for $\nu \geq 0$, $C_t^{\nu+1} = D_t(C_1^\nu, ..., C_l^\nu)$ for $t = 1, ..., l$.

Then we add the following step:

[3]Since we adopt classical logic as the base logic, we can abbreviate some of those connectives. Since, however, our aim is to study logical inferences for decision making rather than semantical contents, we use a full system.

A1: If $\langle C_t^\nu : \nu \geq 0\rangle, t = 1, ..., l$ are $(k+1)$-permissible sequences and if $F(q_1, ..., q_l)$ in $\mathcal{P}_1$, then $\langle F(C_1^\nu, ..., C_l^\nu) : \nu \geq 0\rangle$ is $(k+1)$-permissible.

Then, we say that Φ is a $(k+1)$-*permissible set* iff Φ is a nonempty finite subset of $\mathcal{P}_{k+1}$ or $\Phi = \{C^\nu : \nu \geq 0\}$ for some a $(k+1)$-sequence $\langle C^\nu : \nu \geq 0\rangle$. We denote, by $\mathcal{F}(\mathcal{P}_{k+1})$, the set of all $(k+1)$-permissible sets.

We define the set of all formulae by $\mathcal{P}_\omega = \bigcup_{k<\omega} \mathcal{P}_k$, and also the set of *permissible sets* by $\mathcal{F}_\omega = \cup_{k<\omega}\mathcal{F}(\mathcal{P}_k)$.

In the following, we may write $A \wedge B, A \vee B$ and $A \vee B \vee C$ for $\wedge\{A, B\}, \vee\{A, B\}$ and $\vee\{A, B, C\}$, etc., when these are easier. We also write $A \equiv B$ for $(A \supset B) \wedge (B \supset A)$. We often abbreviate the parentheses $(,)$ when it causes no confusion, and we may use different parentheses such as $[,]$ and $\langle,\rangle$. Also, we say that a formula A is *non-epistemic* iff no belief operators $\mathrm{B}_i(\cdot), i \in N$ occurs in A.

In $\mathcal{P}_\omega$, we formulate infinite regress, common knowledge, and common belief.

Infinite Regress (in the mind of player i): Let $D_i(p_1, ..., p_n) = \wedge_{j\neq i}\mathbf{B}_j(p_j)$, $i \in N$, where $p_1, ..., p_n$ are propositional variables. Let $\mathbf{A} = (A_1, ..., A_n)$ be an n-tuple of formulae in $\mathcal{P}_\omega$. Then we form the permissible sequences $\langle C_i^0, C_i^1, C_i^2, ...\rangle, i \in N$ as follows:

$$C_i^0 = A_i,\ C_i^{\nu+1} = D_i(C_1^\nu, ..., C_n^\nu) = \wedge_{j\neq i}\mathbf{B}_j(C_j^\nu) \text{ for } i \in N. \tag{2.1}$$

Then, we define $\mathbf{Ir}_i(\mathbf{A}) := \wedge\{C_i^\nu : \nu \geq 0\}$, which we call the *infinite regress* for player i from the *reference list* $\mathbf{A} = (A_1, ..., A_n)$. It should be noted that $\mathbf{Ir}_i(\mathbf{A})$ is occurring in the mind of player i, i.e., in the scope of $\mathbf{B}_i$. This is crucial for our consideration of the Nash solution theory in Section 5, and will be elaborated more in Sections 4 and 5. The sequence in (2.1) is illustrated graphically for $n = 2$ from with $\mathbf{A} = (A_1, A_2)$ as follows:

Diagram 2.2

$C_1^0 = A_1$		$C_1^1 = \mathbf{B}_2(C_2^0) = \mathbf{B}_2(A_2)$		$\cdots$
	$\nearrow$		$\nearrow$	
$C_2^0 = A_2$		$C_2^1 = \mathbf{B}_1(C_1^0) = \mathbf{B}_1(A_1)$		$\cdots$

The first raw row of Diagram 2.2 corresponds to the sequence following the South-ward and Northeast-ward arrows in Diagram 1.2.

We denote each C_i^ν by $\mathbf{Ir}_i^\nu(\mathbf{A})$; and hence, $\mathbf{Ir}_i(\mathbf{A}) := \wedge\{\mathbf{Ir}_i^\nu(\mathbf{A}) : \nu \geq 0\}$; this expression will be used later. In this definition, the infinite regress $\mathbf{Ir}_i(\mathbf{A})$ is assumed to occur in the mind of player i; and if we refer to this from the outside observer's perspective, we should use $\mathbf{B}_i(\mathbf{Ir}_i(\mathbf{A}))$. In fact, Lemma 3.3 guarantees that we can move out and in the scope of $\mathbf{B}_i(\cdot)$. This abbreviation simplifies our expressions a lot, but we should not forget the abbreviation.

Common Knowledge: Let $D^c(p) = \wedge_{i\in N}\mathbf{B}_i(p)$, where p is a propositional variable.

Given $A \in \mathcal{P}_\omega$, the *common knowledge* of A is defined as the infinite conjunction $\wedge\{\mathbf{C}^\nu(A) : \nu \geq 0\}$, where $\langle \mathbf{C}^\nu(A) : \nu \geq 0\rangle$ is generated by

$$\mathbf{C}^0(A) = A \text{ and } \mathbf{C}^{\nu+1}(A) = \wedge_{i\in N}\mathbf{B}_i(\mathbf{C}^\nu(A)) \text{ for all } \nu \geq 0. \tag{2.2}$$

We write $\mathbf{C}(A) := \wedge\{\mathbf{C}^\nu(A) : \nu \geq 0\}$. We will discuss the relationship between the infinite regress $\mathbf{Ir}_i(\mathbf{A})$ and common knowledge $\mathbf{C}(A)$ in Section 4[4].

The infinite regress $\mathbf{Ir}_i(\mathbf{A})$ is given subjectively in the mind of player i, while the common knowledge $\mathbf{C}(A)$ is formulated objectively. This unparallelism will be explained in Section 4.

2.2. Logical Axioms and Inference Rules for $\mathbf{GL}_\omega$

The base logic of GL_ω is an infinitary classical logic, formulated by the five axiom schemata and three inference rules: for all formulae A, B, C and permissible set Φ,

L1: $A \supset (B \supset A)$;

L2: $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;

L3: $(\neg A \supset \neg B) \supset ((\neg A \supset B) \supset A)$;

L4: $\wedge\Phi \supset A$, where $A \in \Phi$;

L5: $A \supset \vee\Phi$, where $A \in \Phi$;

$$\frac{A \supset C \quad A}{C}\ (\mathrm{MP}) \qquad \frac{\{A \supset B : B \in \Phi\}}{A \supset \wedge\Phi}\ (\wedge\text{-rule}) \qquad \frac{\{B \supset A : A \in \Phi\}}{\vee\Phi \supset A}\ (\vee\text{-rule}).$$

The axiom schemata L1-L5 and the three inference rules together with $\mathcal{P}_\omega$ constitute a classical infinitary logic.[5] Here, formulae are taken from $\mathcal{P}_\omega$ and they may include belief operators $\mathbf{B}_i$, $i \in N$. In $\wedge$-rule and $\vee$-rule, Φ is required to be a permissible set of formulae, but the sets of upper sets of those inferences may not. In this classical logic, each formula with the outermost $\mathbf{B}_i(\cdot)$ behaves just as a proposition variable, independent of its content.

We have two epistemic axiom schemata and one inference rule: for all formulae A, C, permissible Φ and $i \in N$,

K: $\mathbf{B}_i(A \supset C) \supset (\mathbf{B}_i(A) \supset \mathbf{B}_i(C))$;

D: $\neg\mathbf{B}_i(\neg A \wedge A)$;

[4] The *common belief* is defined in the same manner with the replacement of the start $\wedge_{i\in N}\mathbf{B}_i(A)$ in (2.2). Nevertheless, the common belief does not play a role in this paper.

[5] The fragment determined by $\supset$ and $\neg$ with L1-L3 and MP is the same as the classical propositional logic in many textbook such as Mendelson [16]. Basic provable formulae are found in those books.

$\wedge$-Barcan: $\wedge \mathbf{B}_i(\Phi) \supset \mathbf{B}_i(\wedge\Phi)$;

$$\frac{A}{\mathbf{B}_i(A)} \text{ (Necessitation).}$$

By GL_ω, we denote the logical system determined by all the above axioms and inference rules. This is an infinitary version of epistemic logic KD^n with Axiom $\wedge$-Barcan. Axiom K and Necessitation rule imply that all (real and imaginary) players have logical abilities expressed by (infinitary) classical logic. Axiom D means that a belief of an inconsistent statement is not allowed: With Axiom K and Necessitation rule, this is equivalent to $\mathbf{B}_i(\neg A) \supset \neg\mathbf{B}_i(A)$. Axiom D is essential for meaningful discussions of practical problems. The last additional axiom is called the $\wedge$-*Barcan*: If Φ is a finite set, $\wedge\mathbf{B}_i(\Phi) \supset \mathbf{B}_i(\wedge\Phi)$ is provable, but if Φ is an infinite set, $\mathbf{B}_i(\wedge\Phi)$ is logically stronger than $\wedge\mathbf{B}_i(\Phi)$, since the former includes the totality of Φ in the scope of $\mathbf{B}_i$. The axiom $\wedge$-Barcan makes them equivalent. This equivalence is essential for discussions of infinite regress.

A *proof* $P = \langle X, <; \psi\rangle$ in logic GL_ω consists of a countable tree $\langle X, <\rangle$ and a function $\psi : X \to \mathcal{P}_\omega$ with the following requirements:

(o): $\langle X, <\rangle$ has no infinite path from its root;

(i): for each node x in $\langle X, <\rangle$, $\psi(x)$ is a formula attached to x;

(ii): for each leaf x in $\langle X, <\rangle$, $\psi(x)$ is an instance of the axiom schemata;

(iii): for each non-leaf x in $\langle X, <\rangle$,

$$\frac{\{\psi(y) :\ y \text{ is an immediate predecessor of } x\}}{\psi(x)}$$

is a one of the above four inference rules[6]. When $A = \psi(x_0)$ is attached to the root x_0 of $\langle X, <\rangle$, we call $\langle X, <; \psi\rangle$ a *proof of* A.

We say that A is *provable*, denoted by $\vdash A$, iff there is a proof of A. For a (possibly infinite) set of formulae Γ, we write $\Gamma \vdash A$ iff $\vdash A$ or $\vdash \wedge\Gamma' \supset A$ for some nonempty finite subset Γ' of Γ.

We consider the other two logical systems by adding the following axiom schemata: for $i \in N$ and $A \in \mathcal{P}_\omega$,

Axiom T (Truthfulness): $\mathbf{B}_i(A) \supset A$;

Axiom 4 (Positive Introspection): $\mathbf{B}_i(A) \supset \mathbf{B}_i\mathbf{B}_i(A)$.

The logical systems obtained from GL_ω by the addition of Axiom T or 4 is denoted by $\mathrm{GL}_\omega(\mathrm{T})$ or $\mathrm{GL}_\omega(4)$, respectively. The finitary versions of GL_ω, $\mathrm{GL}_\omega(\mathrm{T})$ and $\mathrm{GL}_\omega(4)$ are KD^n, KT^n, and $\mathrm{K4}^n$. Axiom T is a strengthening of Axiom D, which will be important

[6] If x has a unique immediate predecessor y, the set bracket is unnecessary.

in relating the infinite regress to the common knowledge. Axiom 4 also makes a similar relationship them, while keeping subjectivity, which will be discussed in Section 3.2. Nevertheless, the main part of this paper will be discussed in GL_ω.

We will use the facts stated in the following lemma often without referring to them, and we only prove claim (8). Claim (3) is very useful for the subsequent sections[7]. As stated, (6) is equivalent to Axiom D. Claims (7) and (8) are dual, and proved in the dual manner. The converse of (8) is $\wedge$-Barcan.

Lemma 2.1. Let $A, B, C \in \mathcal{P}_\omega$ and let Φ be a permissible set.

(1): $\vdash A \supset B$ and $\vdash B \supset C$ imply $\vdash A \supset C$;

(2): $\vdash (A \supset B) \equiv \neg A \vee B$;

(3): $\vdash [A \supset (B \supset C)] \equiv [A \wedge B \supset C)]$;

(4): $\vdash \neg(\wedge\Phi) \equiv \vee(\neg\Phi)$ and $\vdash \neg(\vee\Phi) \equiv \wedge(\neg\Phi)$;

(5): $\Gamma \vdash \wedge\Phi$ if and only if $\Gamma \vdash A$ for all $A \in \Phi$;

(6): $\vdash \mathbf{B}_i(\neg A) \supset \neg\mathbf{B}_i(\neg A)$;

(7): $\vdash \vee\mathbf{B}_i(\Phi) \supset \mathbf{B}_i(\vee\Phi)$;

(8): $\vdash \mathbf{B}_i(\wedge\Phi) \supset \wedge\mathbf{B}_i(\Phi)$.

Proof. (8): Let A be an arbitrary formula in Φ. Since $\vdash \wedge\Phi \supset A$ by L4, we have $\vdash \mathrm{B}_i(\wedge\Phi \supset A)$ by Necessitation (to be abbreviated as Nec). By K and MP (we will omit references to MP hereafter), we have $\vdash \mathrm{B}_i(\wedge\Phi) \supset \mathrm{B}_i(A)$. Since this holds for all $A \in \Phi$, we have, $\wedge$-rule, $\vdash \mathrm{B}_i(\wedge\Phi) \supset \wedge\mathrm{B}_i(\Phi)$. ■

Logic GL_ω is complete with respect to Kripke semantics, which can be proved, using the method developed in Tanaka-Ono [20]. In this paper, we use proof theory GL_ω almost exclusively for our discussions of provable statements, but use the Kripke semantics only for negative evaluations, i.e., the soundness part.

The logical system GL_ω is a "small" extension of the finitary KD^n, relative to the literature of infinitary logic (cf., Karp [12] and Heifetz [7]. In fact, we can restrict our attention even to a smaller fragment: By restricting the set of formulae $\mathcal{P}_\omega$ to $\mathcal{P}_k$ $(k < \omega)$ in GL_ω, we have the system, denoted by GL_k, which is also Kripke-complete. For practical discussions on infinite regresses such as the game theoretical problem discussed in Section 5, GL_k is enough for some small $k \geq 1$. We will give a comment on this problem in Section 6.

[7] In fact, it takes many steps to prove (3) in our axiomatic system. A derivation is found in the Appendix of Kaneko [10].

3. Basic Properties of Infinite Regresses

In this section, we provide various basic properties related to the concept of infinite regress. We summarize them in Lemmas 3.1 and 3.2. We will use them in the subsequent studies of the Nash solution theory. Lemma 3.1.IRA implies $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset A_i$, but this is a subjective statement in the mind of player i. Then, we give another lemma to guarantee this treatment, which requires our epistemic logic to be the KD-type, rather than the KT or K4-types.

Lemma 3.1. Let $\mathbf{A} = (A_1, ..., A_n)$ be a reference list.

IRA: For all $i \in N$, $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$.

IRI: Let $\mathbf{D} = (D_1, ..., D_n)$ be a reference list. If $\vdash D_i \supset A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(D_j))$ for all $i \in N$, then $\vdash D_i \supset \mathbf{Ir}_i(\mathbf{A})$ for all $i \in N$.

Proof. **IRA**: By L4, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset A_i$. Let $j \neq i$. For each $\nu \geq 0$, $\vdash \mathbf{Ir}_i^{\nu+1}(\mathbf{A}) \supset \mathbf{B}_j(\mathbf{Ir}_j^{\nu}(\mathbf{A}))$ by L4. Hence, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_j(\mathbf{Ir}_j^{\nu}(\mathbf{A}))$ for all $\nu \geq 0$ by L4. Since $\{\mathbf{B}_j(\mathbf{Ir}_j^{\nu}(\mathbf{A})) : \nu \geq 0\}$ is a permissible sequence, we have, by $\wedge$-rule, $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \wedge\{\mathbf{B}_j(\mathbf{Ir}_j^{\nu}(\mathbf{A})) : \nu \geq 0\}$. Thus, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$.

IRI: Suppose $\vdash D_i \supset A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(D_j))$ for all $i \in N$. First, we show by induction that $\vdash D_i \supset \mathbf{Ir}_i^{\nu}(\mathbf{A})$ for all $i \in N$ and all $\nu \geq 0$. For $\nu = 0$, this follows the supposition. We make the induction hypothesis that $\vdash D_i \supset \mathbf{Ir}_i^{\nu}(\mathbf{A})$ for all $i \in N$. Let $j \neq i$. Then, $\vdash \mathbf{B}_j(D_j) \supset \mathbf{B}_j(\mathbf{Ir}_j^{\nu}(\mathbf{A}))$ by Nec and K, which together with $\vdash D_i \supset \mathbf{B}_j(D_j)$ implies $\vdash D_i \supset \mathbf{B}_j(\mathbf{Ir}_j^{\nu}(\mathbf{A}))$. Since $j \neq i$ is arbitrary, we have $\vdash D_i \supset \mathbf{Ir}_i^{\nu+1}(\mathbf{A})$. Thus, we have $\vdash D_i \supset \mathbf{Ir}_i^{\nu}(\mathbf{A})$ for all $\nu \geq 0$. Since $\{\mathbf{Ir}_i^{\nu}(\mathbf{A}) : \nu \geq 0\}$ is permissible, we have $\vdash D_i \supset \wedge\{\mathbf{Ir}_i^{\nu}(\mathbf{A}) : \nu \geq 0\}$. ■

Claim IRA suggests that the infinite regress operators $\mathbf{Ir}_i(\cdot), i \in N$ may be regarded as fixed points having the property IRA. Claim IRI implies that $\mathbf{Ir}_i(\cdot), i \in N$ are the deductively weakest formulae having this property. Indeed, we can construct a fixed-point logic in a finitary manner with the operator symbols $\mathbb{I}r_i(\cdot), i \in N$ and requiring IRA and IRI for them. We can prove that the resulting system is Kripke complete.

Although IRA and IRI give a lot of information about $\mathbf{Ir}_i(\cdot), i \in N$, it would still be useful to list some basic properties.

Lemma 3.2. Let $\mathbf{A} = (A_1, ..., A_n)$ and $\mathbf{B} = (B_1, ..., B_n)$ be reference lists. Then,

(0): $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset (\wedge_{j \neq i} \wedge_{k \neq j} \mathbf{B}_j \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A})))$;

(1): $\vdash \mathbf{Ir}_i(\mathbf{A}) \equiv A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$;

(2): $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{Ir}_1(\mathbf{A}), ..., \mathbf{Ir}_n(\mathbf{A}))$;

(3): $\vdash \mathbf{Ir}_i(A_1 \supset B_1, ..., A_n \supset B_n) \supset (\mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{B}))$;

(4): $\vdash \mathbf{Ir}_i(A_1 \wedge B_1, ..., A_n \wedge B_n) \equiv (\mathbf{Ir}_i(\mathbf{A}) \wedge \mathbf{Ir}_i(\mathbf{B}))$;

(5): If $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset B_i$ for each $i \in N$, then $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{B})$;

(6): $\vdash \mathbf{Ir}_i(\mathbf{A}) \wedge \mathbf{Ir}_i(\mathbf{B}) \supset \mathbf{Ir}_i(A_i; \mathbf{B}_{-i})$, where $(A_i; \mathbf{B}_{-i})$ is the reference list obtained from $\mathbf{B}$ by substituting A_i for B_i in $\mathbf{B}$.

Proof. We prove (0)-(3). The others can be proved in similar manners.

(0): Let $j \neq i$. Then, by IRA, $\vdash \mathbf{Ir}_j(\mathbf{A}) \supset (\wedge_{k \neq j} \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A})))$. By K, and Nec, we have $\vdash \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})) \supset \mathbf{B}_j(\wedge_{k \neq j} \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A})))$. Since $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_j(\wedge_{k \neq j} \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A})))$. Using $\wedge$-rule, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \wedge_{j \neq i} \mathbf{B}_j(\wedge_{k \neq j} \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A})))$. Since the order of $\wedge$ and $\mathbf{B}_j$ can be changed for a finite set of formulae, we have (0).

(1): The direction $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$ follows from IRA. For the other direction, let $B_i = A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$. We claim $\vdash B_i \supset A_i \wedge (\wedge_{j \neq i} \mathbf{B}_j(B_j))$ for $i \in N$. Once this is proved, we have $\vdash B_i \supset \mathbf{Ir}_i(\mathbf{A})$. Let us prove the claim. By L4, we have $\vdash B_i \supset A_i$. Then, for any $j \neq i$, we have, by IRI, $\vdash \mathbf{Ir}_j(\mathbf{A}) \supset A_j$, and hence, by Nec and K, $\vdash \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})) \supset \mathbf{B}_j(A_j)$. Since $\vdash B_i \supset \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A}))$, we have $\vdash B_i \supset \mathbf{B}_j(A_j)$. Similarly, $\vdash \mathbf{Ir}_j(\mathbf{A}) \supset \wedge_{k \neq j} \mathbf{Ir}_k(\mathbf{A})$, and hence, by Nec and K, $\vdash \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})) \supset \mathbf{B}_j(\wedge_{k \neq j} \mathbf{Ir}_k(\mathbf{A}))$. Since $\vdash B_i \supset \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A}))$, we have $\vdash B_i \supset \mathbf{B}_j(\wedge_{k \neq j} \mathbf{Ir}_k(\mathbf{A}))$. Thus, by $\wedge$-rule, $\vdash B_i \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(\wedge_{k \neq j} \mathbf{Ir}_k(\mathbf{A}))$, which can be changed by Lemma 1.1.(8), $\vdash B_i \supset \mathbf{B}_j(A_j \wedge (\wedge_{k \neq j} \mathbf{Ir}_k(\mathbf{A})))$, i.e., $\vdash B_i \supset \mathbf{B}_j(B_j)$. The claim follows from this.

(2): Let $B_i = \mathbf{Ir}_i(\mathbf{A})$ for $i \in N$. Then, for $i \in N$, we have $\vdash B_i \supset \mathbf{Ir}_i(\mathbf{A})$ and $\vdash B_i \supset \wedge_{j \neq i} \mathbf{B}_j(B_j)$ by IRA. Then, by IRI, we have $\vdash B_i \supset \mathbf{Ir}_i(\mathbf{B})$, which is the assertion.

(3): Let $C_i = \mathbf{Ir}_i(A_1 \supset B_1, ..., A_n \supset B_n) \wedge \mathbf{Ir}_i(\mathbf{A})$ for $i \in N$. By IRA, $\vdash C_i \supset (A_i \supset B_i) \wedge A_i$. Thus, $\vdash C_i \supset B_i$. Let $j \neq i$. By IRA, $\vdash C_i \supset \mathbf{B}_j(\mathbf{Ir}_j(A_1 \supset B_1, ..., A_n \supset B_n)) \wedge \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A}))$. By Lemma 3.1.(8), we have $\vdash C_i \supset \mathbf{B}_j(C_j)$. Thus, by IRI, we have $\vdash C_i \supset \mathbf{Ir}_i(B_i)$. By Lemma 3.1.(3), we have the assertion. ■

These have interesting implications: Property (1) implies that $\mathbf{Ir}_i(\mathbf{A}), i \in N$, form a fixed-point, (2) resembles Axiom 4, and (3) is a general form of Axiom K. These properties will be used in the game theoretical applications in Section 5.

We have defined the infinite regress $\mathbf{Ir}_i(\mathbf{A})$ in the mind of player i. If we look at this fact from the outside observer's viewpoint, we should consider $\mathbf{B}_i(\mathbf{Ir}_i(\mathbf{A}))$. Our practical concern is to have implication of the forms such as $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{C})$, and if this occurs in the mind of player i, it can be stated as $\vdash \mathbf{B}_i[\mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{C})]$ or $\vdash \mathbf{B}_i(\mathbf{Ir}_i(\mathbf{A})) \supset \mathbf{B}_i(\mathbf{Ir}_i(\mathbf{C}))$. The following lemma states that those are equivalent in GL_ω.

Lemma 3.3 (Moving out and in from the Scope of $\mathbf{B}_i(A)$): The following three statements are equivalent:

(1): $\vdash A \supset C$; **(2)**: $\vdash \mathbf{B}_i[A \supset C]$; **(3)**: $\vdash \mathbf{B}_i(A) \supset \mathbf{B}_i(C)$.

This result is known for K, KD, and some others (cf., Chellas [3], p.147). The equivalences do not hold when we add Axiom 4. Let p be an atomic formula. Then, $\mathbf{B}_i(p) \supset \mathbf{B}_i\mathbf{B}_i(p)$ is an instance of Axiom 4, but $p \supset \mathbf{B}_i(p)$ is not provable in GL_ω. A semantic proof is possible based on the Kripke-completness of GL_ω, but for self-containedness, we give a proof-theoretic proof.

Proof of Lemma 3.3. It is straightforward to have (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3). We show (3) $\Longrightarrow$ (1). Suppose $\vdash \mathbf{B}_i(A) \supset \mathbf{B}_i(C)$. Then, there is a proof $P = ((X, <); \psi)$ of $\mathbf{B}_i(A) \supset \mathbf{B}_i(C)$.

Let ε_i be the *erasing operator* which erases $\mathbf{B}_i(\cdot)$ only once. Once it erases $\mathrm{B}_i(\cdot)$, it becomes the null symbol. The exact definition is defined inductively as follows:

(0): for any propositional variable p, $\varepsilon_i(p) = p$;

(1): $\varepsilon_i(\neg D) = \neg\varepsilon_i(D)$; (2): $\varepsilon_i(D \supset E) = \varepsilon_i(D) \supset \varepsilon_i(E)$;

(3): $\varepsilon_i(\wedge\Phi) = \wedge\{\varepsilon_i(D) : D \in \Phi\}$; and $\varepsilon_i(\vee\Phi) = \vee\{\varepsilon_i(D) : D \in \Phi\}$;

(4): $\varepsilon_i(\mathbf{B}_i(D)) = D$, but for any $j \neq i$, $\varepsilon_i(\mathbf{B}_j(D)) = \mathbf{B}_j(\varepsilon_i(D))$.

We apply this operator ε_i to the proof $P = (X, <; \psi)$ but not universally.

We introduce the boundary for this application: We say that $y \in X$ is the *concluding node of Nec*$_i$ (necessitation with B_i) iff the attached formula $\psi(y)$ is the lower formula of Nec_i. We consider a lowest concluding nod of Nec_i, i.e., no other Nec_i occur below y. We denote the set of such lowest concluding nodes of Nec_i by $\{y_1, ..., y_m\}$. These are the *boundary* for the application of ε_i to the attached formula $\psi(x)$: We define $\psi' : X \to \mathcal{P}_\omega$ as follows:

$$\begin{aligned} \psi'(x) &= \psi(x) && \text{if } x > y \text{ for some } y \in \{y_1, ..., y_m\} \\ &= \varepsilon_i\psi(x) && \text{otherwise.} \end{aligned}$$

If the boundary is empty, then all the nodes are regarded as below the boundary. If x is above the boundary, $\psi'(x)$ is the same as $\psi(x)$. We show by the induction from the leaves that $\vdash \psi'(x)$ for any $x \in X$. Consider the subtree determined by the immediate upper node x of some node y of $\{y_1, ..., y_m\}$. Since we did not change the attached formulae, $\vdash \psi'(x)$. Now, we start with the inferences at the boundary. Once this is proved, we have $\vdash \varepsilon_i[\mathbf{B}_i(A) \supset \mathbf{B}_i(C)] = A \supset C$.

Let $y \in \{y_1, ..., y_m\}$, and x be the immediate predecessor of y. Then, $\psi(y)$ is expressed as $\mathbf{B}_i(D)$. Then, since $\psi'(y) = \varepsilon_i\psi(y) = \varepsilon_i\mathbf{B}_i(D) = D = \psi(x) = \psi'(x)$, we have $\vdash \psi'(x)$.

Consider a leaf x below the boundary. The attached formula $\psi(x)$ is an instance of L1-L5, K, D or $\wedge$-Barcan. We should prove $\vdash \varepsilon_i\psi(x)$. Here, we verify only L5, D and $\wedge$-Barcan.

An instance of L5 is expressed as $D \supset \vee\Phi$, where $D \in \Phi$. Then, $\varepsilon_i D \supset \vee\varepsilon_i\Phi$ is also an instance of L5. An instance of Axiom D is expressed as $\neg\mathbf{B}_j(\neg D \wedge D)$. If $j \neq i$, then $\varepsilon_i[\neg\mathbf{B}_j(\neg D \wedge D)] = \neg\mathbf{B}_j[(\neg\varepsilon_i C) \wedge \varepsilon_i C]$, which is another instance of D. If $j = i$, then $\varepsilon_i[\neg \mathbf{B}_i(\neg D \wedge D)] = \neg(\neg D \wedge D)$, which is provable. An instance of $\wedge$-Barcan is expressed as $\wedge \mathbf{B}_j(\Phi) \supset \mathbf{B}_j(\wedge\Phi)$. If $j \neq i$, then $\varepsilon_i[\wedge\mathbf{B}_j(\Phi) \supset \mathbf{B}_j(\wedge\Phi)] = \wedge\mathbf{B}_j(\varepsilon_i\Phi) \supset \mathbf{B}_j(\wedge\varepsilon_i\Phi)$ is also an instance of $\wedge$-Barcan. If $j = i$, then $\varepsilon_i[\wedge\mathbf{B}_i(\Phi) \supset \mathbf{B}_i(\wedge\Phi)] = \wedge\Phi \supset \wedge\Phi$ is provable.

The remaining is to show that the three inferences rules hold below the boundary. Consider only MP. It becomes the form that if $\varepsilon_i D$ and $\varepsilon_i(D \supset E)$ are provable, $\varepsilon_i E$ is provable, i.e., an instance of MP. ■

4. Infinite Regress and Common Knowledge

We have formulated the concept of an infinite regress as a subjective concept. Although $\mathbf{Ir}_i(\mathbf{A})$ and the other $\mathbf{Ir}_j(\mathbf{A})$'s interact with one another, the single $\mathbf{Ir}_i(\mathbf{A})$ is our target and the other $\mathbf{Ir}_j(\mathbf{A})$'s are auxiliary in the reference to the mind of player i. This is well reflected in Lemma 3.2.(1): $\vdash \mathbf{Ir}_i(\mathbf{A}) \equiv A_i \wedge (\wedge_{j\neq i}\mathbf{B}_j(\mathbf{Ir}_j(\mathbf{A})))$. If, however, we assume Axiom T, which blurs the distinction between subjectivity and objectivity, all the $\mathbf{Ir}_i(\mathbf{A})$'s become equivalent, and the infinite regress $\mathbf{Ir}_i(\mathbf{A})$ itself becomes the common knowledge of $\wedge_{j\in N}A_j$. Here, the infinite regress collapses to the common knowledge. This will be shown in Theorem 4.1. With the addition of Axiom 4, we can keep the distinction between those two concepts, but the infinite regress becomes the individual belief of common knowledge. This will be shown in Theorem 4.2. As noted around in Lemma 3.3, here we cannot abbreviate the outer $\mathbf{B}_i[...]$ for the consideration of infinite regress $\mathbf{Ir}_i(\mathbf{A})$.

First, we recall that $\mathrm{GL}_\omega(\mathrm{T})$ is the logic obtained from GL_ω by adding Axiom T.

Lemma 4.1.(1) In GL_ω, $\vdash \mathbf{C}(D) \supset \mathbf{B}_i(\mathbf{C}(D))$ for $i \in N$;

(2) In $\mathrm{GL}_\omega(\mathrm{T})$, $\vdash \mathbf{Ir}_i(\mathbf{A}) \equiv \mathbf{Ir}_j(\mathbf{A})$ for all $i, j \in N$.

Proof.(1): We have $\vdash \mathbf{C}^0(D) \supset D$ and $\vdash \mathbf{C}^{\nu+1}(D) \supset \mathbf{B}_i(\mathbf{C}^\nu(D))$ for all $\nu \geq 0$. Hence, $\vdash \mathbf{C}(D) \supset \mathbf{B}_i(\mathbf{C}^\nu(D))$ for all $\nu \geq 0$. This together with $\wedge$-rule implies $\vdash \mathbf{C}(D) \supset \wedge\{\mathbf{B}_i(\mathbf{C}^\nu(D)) : \nu \geq 0\}$. By $\wedge$-Barcan, we have $\vdash \mathbf{C}(D) \supset \mathbf{B}_i(\wedge\{\mathbf{C}^\nu(D) : \nu \geq 0\})$, which is $\vdash \mathbf{C}(D) \supset \mathbf{B}_i(\mathbf{C}(D))$.

(2): By Lemma 3.1.IRA, we have $\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_j(\mathbf{Ir}_j(A))$ for all $j \neq i$. Thus, by Axiom T, $\vdash \mathbf{Ir}_i(A) \supset \mathbf{Ir}_j(A)$. ■

Now, we have the collapse theorem in $\mathrm{GL}_\omega(\mathrm{T})$.

Theorem 4.1 (Collapse to Common Knowledge). Let $\mathbf{A} = (A_1, ..., A_n)$ and $i \in N$. In $\mathrm{GL}_\omega(\mathrm{T})$, $\vdash \mathbf{Ir}_i(\mathbf{A}) \equiv \mathbf{C}(\wedge_{j\in N}A_j)$.

Proof: First, we show $\vdash \mathbf{C}(\wedge_{j\in N}A_j) \supset \mathbf{Ir}_i(\mathbf{A})$. Let $B = \mathbf{C}(\wedge_{j\in N}A_j)$. By Lemma 4.1.(1), we have $\vdash B \supset \wedge_{j\neq i}\mathbf{B}_j(B)$. By L4 and the definition of $\mathbf{C}$, $\vdash B \supset A_i$. Thus, by IRI, we have $\vdash B \supset \mathbf{Ir}_i(\mathbf{A})$.

Next, we show $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{C}(\wedge_{j\in N}A_j)$. By IRA, $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset A_i$, and $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_j(A_j)$ for each $j \neq i$. Thus, by Axiom T and $\wedge$-rule, $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \wedge_{j\in N}A_j$, i.e., $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{C}^0(\wedge_{j\in N}A_j)$. Suppose the induction hypothesis that $\vdash \mathbf{Ir}_k(\mathbf{A}) \supset \mathbf{C}^\nu(\wedge_{j\in N}A_j)$ for all $k \in N$. Then,

$$\text{for each } k \in N,\ \ \vdash \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A})) \supset \mathbf{B}_k(\mathbf{C}^\nu(\wedge_{j\in N}A_j)) \tag{4.1}$$

by Nec and K. Since $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_k(\mathbf{Ir}_k(\mathbf{A}))$ by IRA, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_k(\mathbf{C}^\nu(\wedge_{j\in N}A_j))$ for all $k \neq i$.

Now, for $k \neq i$, $\vdash \mathbf{Ir}_k(\mathbf{A}) \supset \mathbf{B}_i(\mathbf{Ir}_i(\mathbf{A}))$ by IRA; thus $\vdash \mathbf{Ir}_k(\mathbf{A}) \supset \mathbf{B}_i(\mathbf{C}^\nu(\wedge_{j\in N}A_j))$ by (4.1). But, by Lemma 4.1.(2), we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_k(\mathbf{A})$. Thus, we have $\vdash \mathbf{Ir}_i(\mathbf{A})) \supset \mathbf{B}_i(\mathbf{C}^\nu(\wedge_{j\in N}A_j))$. This and the conclusion of the previous paragraph imply $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \wedge_{k\in N}\mathbf{B}_k(\mathbf{C}^\nu(\wedge_{j\in N}A_j))$, i.e., $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{C}^{\nu+1}(\wedge_{j\in N}A_j)$.

Finally, we have $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \wedge_{\nu\geq 0}\mathbf{C}^\nu(\wedge_{j\in N}A_j)$ by $\wedge$-rule. ■

In Theorem 4.1, one direction does not need Axiom T. Since this statement will be used in the next theorem, we state this result explicitly:

$$\vdash \mathbf{C}(\wedge_{j\in N}A_j) \supset \mathbf{Ir}_i(\mathbf{A}) \text{ in } \mathrm{GL}_\omega. \tag{4.2}$$

On the other hand, Axiom T is used to show $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{C}(\wedge_{j\in N}A_j)$, while $\nvdash \mathbf{Ir}_i(A) \supset \mathbf{C}(\wedge_{j\in N}A_j)$ in GL_ω. This unprovability can be proved by using the soundness part of GL_ω with respect to the Kripke semantics.

Example 4.1. Consider $n = 2$. Let $A_1(p) = p$ and $A_2(p) = \neg p$, where p is an propositional variable. We show that $\mathbf{Ir}_1(A_1, A_2)$ is not contradictory in GL_ω. We construct the following Kripke model: A frame is given as follows:

Diagram 4.1

$$\begin{array}{ccccccccc} w_0 & \underset{1}{\rightarrow} & w_1 & \underset{2}{\rightarrow} & w_2 & \underset{1}{\rightarrow} & w_3 & \underset{2}{\rightarrow} & \cdots \\ \circlearrowleft_2 & & \circlearrowleft_1 & & \circlearrowleft_2 & & \circlearrowleft_1 & & \end{array}$$

That is, $W = \{w_0, w_1, ...\}$ and the accessibility relations R_1, R_2 are given exactly the indexed arrows, which are serial. Now, let τ be an assignment satisfying: $\tau(w_t, p) = \top$ iff t is even (the values for other propositional variables are arbitrary). Then, $(\boldsymbol{F}, \tau, w_0) \models p\ (= \mathbf{Ir}_1^0(A_1, A_2))$, $(\boldsymbol{F}, \tau, w_0) \models \mathbf{B}_1(\neg p)(= \mathbf{Ir}_1^1(A_1, A_2))$, $(\boldsymbol{F}, \tau, w_0) \models \mathbf{B}_1\mathbf{B}_2(p)(= \mathbf{Ir}_1^2(A_1, A_2))$, and so on. Thus, $(\boldsymbol{F}, \tau, w_0) \models \mathbf{Ir}_1(A_1, A_2)$. However, $\vdash \mathbf{C}(A_1 \wedge A_2) \supset p \wedge (\neg p)$, and hence $\vdash \neg\mathbf{C}(A_1 \wedge A_2)$. This implies $(\boldsymbol{F}, \tau, w_0) \nvDash \mathbf{C}(A_1 \wedge A_2)$. Thus, $(\boldsymbol{F}, \tau, w_0) \nvDash \mathbf{Ir}_i(A_1, A_2) \supset \mathbf{C}(A_1 \wedge A_2)$. By soundness, we have $\nvdash \mathbf{Ir}_1(A_1, A_2) \supset \mathbf{C}(A_1 \wedge A_2)$ in GL_ω.

In the 2-person case, Fig.1.1 depicts the one line hierarchy of beliefs in the infinite regress formula $\mathbf{Ir}_1(A_1, A_2)$. However, it may be the case that one player thinks about

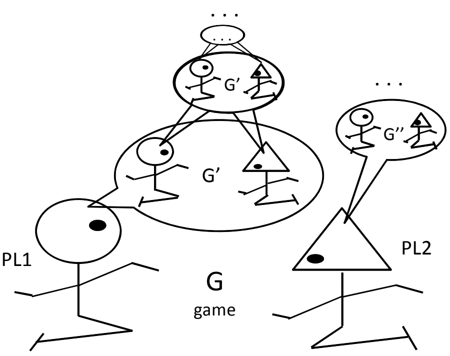

Figure 4.1: Mutual Misunderstanding

the situation with both players including himself: Fig.1.1 is changed to Fig.4.1.When we add Axiom 4 to GL_ω, our concept of infinite regress captures this situation. In this case, however, Lemma 3.3 is no longer available because of Axiom 4. Hence, we should specify the viewpoint for the statement: Theorem 4.2 is stated from the outside observer's viewpoint.

Theorem 4.2 (Individual Belief of Common Knowledge): Let A be any formula in $\mathcal{P}_\omega$. Then, in $\mathrm{GL}_\omega(4)$, $\vdash \mathbf{B}_i\mathbf{C}(A) \equiv \mathbf{B}_i\mathbf{Ir}_i(A, ..., A)$.

Although Theorem 4.2 is not parallel to Theorem 4.1 in that the second has the restriction $\mathbf{A} = (A, ..., A)$, the infinite regress is connected to the common knowledge in the mind of player i. Here, it is allowed for each player to have a belief of different common knowledge. In Fig.4.1, player 1 thinks that the game situation G' is common knowledge between 1 and 2, while player 2 thinks that different G'' is common knowledge. Either differs from the objective G. This mutual misunderstanding is exactly what the comic story called *Konnyaku Mondo* describes (see Kaneko [11]). We note that this differs from the concept of *common beliefs* discussed in the literature of epistemic logic, which, in GL_ω, is defined to be $\mathbf{B}_1\mathbf{C}(A) \wedge \mathbf{B}_2\mathbf{C}(A)$.

In the literatures of epistemic logic as well as game theory, the S5-type logics are typically adopted to have discussions on common knowledge (cf., Fagin, *et al.* [4] and Myer-van der Hoek [15]). One question is the evaluations of epistemic axioms such as Axiom T, 4, and 5 (Negative Introspection). In Theorem 4.1, Axiom T makes the concept of infinite regress to collapse to common knowledge[8], and hence under Axiom T, we have no meaningful distinction between them. In contrast, Theorem 4.2 shows that Axiom 4 makes a subjective connection between infinite regress and common knowledge.

[8] Kaneko [10], Section 5, showed that Axiom T can be described inside $\mathrm{KD4}^n$. The same holds in GL_ω.

In our treatment Axiom 5 seems less relevant and it seems difficult to have a precise evaluation of its role.

Proof of Theorem 4.2. The direction $\vdash \mathbf{B}_i\mathbf{C}(A) \supset \mathbf{B}_i\mathbf{Ir}_i(A, ..., A)$ follows (4.2) by Nec and K.

Now, we show that $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i\mathbf{C}(A)$. For any $j \in N$, by IRA, $\vdash \mathbf{Ir}_j(A, ..., A) \supset A$, and hence, $\vdash \mathbf{B}_j\mathbf{Ir}_j(A, ..., A) \supset \mathbf{B}_j\mathbf{C}^0(A)$. Suppose the induction hypothesis that $\vdash \mathbf{B}_j\mathbf{Ir}_j(A, ..., A) \supset \mathbf{B}_j\mathbf{C}^\nu(A)$ for any $j \in N$. By Axiom 4, we have

$$\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i\mathbf{B}_i\mathbf{C}^\nu(A). \tag{4.3}$$

By IRA, $\vdash \mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_j(\mathbf{Ir}_j(A, ..., A))$ for all $j \neq i$. Hence, by the induction hypothesis, we have $\vdash \mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_j\mathbf{C}^\nu(A)$ for all $j \neq i$, and by $\wedge$-rule, $\vdash \mathbf{Ir}_i(A, ..., A) \supset \wedge_{j\neq i}\mathbf{B}_j\mathbf{C}^\nu(A)$. Thus, by Nec and K, we have $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i(\wedge_{j\neq i}\mathbf{B}_j\mathbf{C}^\nu(A))$. This, together with (4.3), implies $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i(\wedge_{j\in N}\mathbf{B}_j\mathbf{C}^\nu(A))$, that is, $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i(\mathbf{C}^{\nu+1}(A))$.

By the mathematical induction principle, and $\wedge$-rule, we have $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \wedge\{\mathbf{B}_i(\mathbf{C}^\nu(A)) : \nu \geq 0\}$. By $\wedge$-Barcan, we have $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i(\wedge\{\mathbf{C}^\nu(A) : \nu \geq 0\})$, i.e., $\vdash \mathbf{B}_i\mathbf{Ir}_i(A, ..., A) \supset \mathbf{B}_i\mathbf{C}(A)$. $\blacksquare$

5. An Application to Game Theory

Here, we apply the concept of infinite regress to the prediction/decision making discussed in Section 1. Distinguishing between predictions and decisions, the criterion and the resulting outcome are naturally expressed in terms of infinite regresses. In this paper, we consider only the 2-person case and a certain class of games. More discussions will be given in separate papers. The main theorems (Theorems 5.2 and 5.3) will be proved in Section 7.

5.1. Basic Concepts for 2-Person Games

Let $G = (\{1, 2\}, \{S_1, S_2\}, \{h_1, h_2\})$ be a finite 2-person game, where $1, 2$ are the players, S_i is the finite set of *strategies* and $h_i : S := S_1 \times S_2 \to \mathbb{R}$ is the *payoff function* for player $i = 1, 2$. An example is given in Table 1.1. When we focus on player i, the other player is denoted by j. We also write $(s_i; s_j)$ for $s = (s_1, s_2) \in S$. A strategy s_i for player i is a *best-response* against s_j iff $h_i(s_i; s_j) \geq h_i(t_i; s_j)$ for all $t_i \in S_i$. A strategy pair $s \in S$ is a *Nash equilibrium* in G iff s_i is a best response against s_j for $i = 1, 2$. We denote $E(G)$ the set of all Nash equilibria in G, where $E(G)$ may be empty. We say that s_i is a *Nash strategy* iff $(s_i; s_j) \in E(G)$ for some $s_j \in S_j$, and denote the set of all Nash strategies for player i by $E_i(G)$. In the game of Table 1.1, $(\mathbf{s}_{11}, \mathbf{s}_{21})$ is the unique Nash equilibrium, and hence $\mathbf{s}_{i1}$ is the Nash strategy for player $i = 1, 2$. (See Luce-Raiffa [14], Osborne-Rubinstein [18] for basic game theoretic concepts.)

We say that game G is *solvable* iff $E(G) \neq \emptyset$ and for $i = 1, 2$,

$$\text{if } s, s' \in E(G), \text{ then } (s_i; s'_j) \in E(G). \tag{5.1}$$

The game of Table 1.1 is solvable, but there are many unsolvable games (cf., [14]). It can be observed that (5.1) is equivalent to $E(G) = E_1(G) \times E_2(G)$. In this case, $E(G)$ is called the *Nash solution*.

When G is unsolvable and $E(G) \neq \emptyset$, Nash [17] provided the concept of a subsolution. In this paper, however, we focus only on solvable games.

The above language is incapable in distinguishing explicitly between decisions and predictions. However, it would be useful to give a description of prediction/decision making in this language. The prediction/decision criterion $\mathrm{N1}^o$-$\mathrm{N2}^o$ in Section 1 is now formulated as follows: Let E_i be a subset of S_i for $i = 1, 2$ satisfying

$$E_1 \neq \emptyset \Longleftrightarrow E_2 \neq \emptyset. \tag{5.2}$$

Then, $\mathrm{N1}^o$-$\mathrm{N2}^o$ are:

N1: for any $s_1 \in E_1$, s_1 is a best response against s_2 for all $s_2 \in E_2$;

N2: for any $s_2 \in E_2$, s_2 is a best response against s_1 for all $s_1 \in E_1$.

We have the following theorem.

Theorem 5.1. Let G be a game satisfying (5.1), and let E_i be a subset of S_i for $i = 1, 2$ with (5.2). Then $E = E_1 \times E_2$ is the Nash solution $E(G)$ of G if and only if (E_1, E_2) is the greatest pair satisfying (5.2) and N1-N2.[9]

The nonemptiness of $E(G)$ is not assumed for Theorem 5.1, but is necessary for decision making itself.

Theorem 5.1 is proved in a separate paper. It is interpreted as meaning that a Nash strategy is a possible final decision for player i, and the set of Nash strategies for player j is i's predictions. However, this is purely interpretational in that the mathematical formalism has no distinction between a decision and a prediction. To have a faithful description of prediction/decision making, we adopt epistemic logic GL_ω.

To discuss the above game theoretical problems, we adopt the following symbolic expressions as the list of propositional variables in GL_ω: for $i = 1, 2$,

atomic preference formulae: $\mathrm{Pr}_i(s; t)$ for $s, t \in S$;

atomic decision/prediction formulae: $I_i(s_i)$ for $s_i \in S_i$.

The expression $\mathrm{Pr}_i(s; t)$ is intended to mean that player i *weakly prefers* strategy pair s to strategy pair t. The expression $I_i(s_i)$ means that s_i is a *possible final decision* for

[9] The "greatest" is relative to the componentwise set-inclusions.

player i in the scope of $\mathbf{B}_i(\cdots)$. Player i's prediction about j's decision is expressed as $\mathbf{B}_j(I_j(s_j))$ occurring in $\mathbf{B}_i(\cdots)$, which will be clarified below.

We describe the payoff functions h_1, h_2 in G as follows: for each $i = 1, 2$,

$$g_i = \wedge \left[\{\mathrm{Pr}_i(s;t) : h_i(s) \geq h_i(t)\} \cup \{\neg \mathrm{Pr}_i(s;t) : h_i(s) < h_i(t)\} \right]. \tag{5.3}$$

The concepts of best response and Nash equilibrium are now formulated as follows: The statement "s_i is a best response to s_j" is given as $\mathrm{best}_i(s_i; s_j) := \wedge_{t_i \in S_i} \mathrm{Pr}_i(s_i, s_j; t_i, s_j)$, "$s = (s_1, s_2) \in S$ is a Nash equilibrium" is $\mathrm{nash}(s) := \mathrm{best}_1(s_1; s_2) \wedge \mathrm{best}_2(s_2; s_1)$, and "$s_i$ is a Nash strategy for player i" is $\vee_{t_j \in S_j} \mathrm{nash}(s_i; t_j)$.

5.2. Epistemic Axioms for Prediction/Decision Making

Here, we analyze the prediction/decision making by player i in a game in epistemic logic GL_ω. First, recall that Lemma 3.3 allows us to abbreviate the outer $\mathbf{B}_i[\cdots]$; for example, we can consider Diagram 1.2 forgetting the outer $\mathbf{B}_1[\cdots]$. Thus, we focus on the inside of the scope of $\mathbf{B}_i[\cdots]$. Nevertheless, we should keep this outer $\mathbf{B}_i[\cdots]$ in our mind, since we adopt the description of our axioms particularly from the perspective of player i.

This is reflected in the description of player i's decision:

(i): $I_i(s_i)-$ s_i is a possible decision for player i.

However, when player i predicts about j's decision, i takes the perspective of player j, i.e.,

(ii): $\mathbf{B}_j(I_j(s_j))-$ it is predicted by player i that s_j is a possible decision for j.

Player i's decision $I_i(s_i)$ itself occurs in the scope of the outer $\mathbf{B}_i[\cdots]$ in (i), and we do not need to take the form $\mathbf{B}_i(I_i(s_i))$.

Another step is that "player j thinks about i's decision" in the scope of the outer $\mathbf{B}_i[\cdots]$ is expressed as $\mathbf{B}_j\mathbf{B}_i(I_i(s_i))$; since the outer $\mathbf{B}_i[\cdots]$ does not sneak into the scope of $\mathbf{B}_j(\cdot)$, we need to put $\mathbf{B}_i(I_i(s_i))$. This will be used in the second axiom below. Since this point is subtle, we will return after Theorem 5.2.

We consider the following three axioms for player i's prediction/decision making:

$\mathbf{D0}_i$ (Best Choice toward Predictions): $\wedge_{s \in S}[I_i(s_i) \supset \langle \mathbf{B}_j(I_j(s_j)) \supset \mathrm{best}_i(s_i; s_j)\rangle]$;

$\mathbf{D1}_i$ (Predictability): $\wedge_{s_i \in S_i}[I_i(s_i) \supset \mathbf{B}_j\mathbf{B}_i(I_i(s_i))]$;

$\mathbf{D2}_i$ (Necessary Prediction): $\wedge_{s_i \in S_i}[I_i(s_i) \supset \vee_{s_j \in S_j} \mathbf{B}_j(I_j(s_j))]$.

Axioms $\mathrm{D0}_1$ and $\mathrm{D0}_2$ formalize the prediction/decision criterion N1-N2 taking the distinction between (i) and (ii) into account. Axioms $\mathrm{D1}_1$ and $\mathrm{D1}_2$ state that each player's decisions are correctly predictable by the other player. Since each $\mathrm{D1}_i$ occurs in $\mathbf{B}_i[\cdots]$, player i himself believes this correctness. Axioms $\mathrm{D2}_1$ and $\mathrm{D2}_2$, corresponding to (5.2), mean that one's decision requires his prediction about the other.

Let $\mathrm{D}_i = \mathrm{D0}_i \wedge \mathrm{D1}_i \wedge \mathrm{D2}_i$ for $i = 1, 2$, and $\mathbf{D} = (\mathrm{D}_1, \mathrm{D}_2)$. As discussed in Section 1, the first layer D_i is not enough and requires $\mathbf{B}_j(\mathrm{D}_j)$ to determine the meaning of player i's predictions $\mathbf{B}_j(I_j(s_j))$, which further requires $\mathbf{B}_j\mathbf{B}_i(\mathrm{D}_i)$ to determine $\mathbf{B}_j\mathbf{B}_i(I_i(s_j))$, and so on. This sequence is described by the infinite regress formula $\mathbf{Ir}_i(\mathbf{D}) = \mathbf{Ir}_i(\mathrm{D}_1,\mathrm{D}_2)$.

To study $\mathbf{Ir}_i(\mathbf{D}) = \mathbf{Ir}_i(\mathrm{D}_1,\mathrm{D}_2)$ as a requirement for prediction/decision making, we make an analogy of regarding $\mathbf{Ir}_i(\mathbf{D})$ as a system of simultaneous equations having unknown symbols $I_i(s_i)$'s and $I_j(s_j)$'s. More precisely, the target unknowns are player i's decisions $I_i(s_i)$ and predictions $\mathbf{B}_j(I_j(s_j))$. Then, we look for solutions for these unknowns in $\mathcal{P}_\omega$. Now, we have the following candidates for them: for each $s_i \in S_i, i = 1, 2$,

$$A_i^*(s_i) := \vee_{s_j \in S_j} \mathbf{Ir}_i[\mathrm{best}_i(s_i; s_j); \mathrm{best}_j(s_j; s_i)]. \tag{5.4}$$

This corresponds to the Nash strategy $\vee_{s_j \in S_j} \mathrm{nash}(s_i; s_j)$, but the infinite regress is needed in accordance with $\mathbf{Ir}_i(\mathbf{D}) = \mathbf{Ir}_i(\mathrm{D}_1,\mathrm{D}_2)$.

We show that the formulae given in (5.4) are solutions for $\mathbf{Ir}_i(\mathbf{D})$. For this, we take two steps (a): explication; and (b) choice of the weakest formulae for $\mathbf{Ir}_i(\mathbf{D})$. When we obtain a solution for a system of equations, first we derive a candidate for the system, which is (a); and then we show that it is the deductively weakest formulae satisfying $\mathbf{Ir}_i(\mathbf{D})$. In this subsection, we give a theorem for step (a), and in Section 5.3, we take step (b).

Theorem 5.2.(Explication of $\mathbf{Ir}_i(\mathbf{D})$):

$$\mathbf{Ir}_i(\mathbf{D}) \vdash \wedge_{s \in S} \langle \mathbf{Ir}_i[I_1(s_1) \supset A_1^*(s_1), I_2(s_2) \supset A_2^*(s_2)] \rangle. \tag{5.5}$$

It follows from Theorem 5.2 and IRA of Lemma 3.1 that $\mathbf{Ir}_i(\mathbf{D}) \vdash I_i(s_i) \supset A_i^*(s_i)$ and $\mathbf{Ir}_i(\mathbf{D}) \vdash \mathbf{B}_j(I_j(s_j)) \supset \mathbf{B}_j(A_j^*(s_j))$. That is, $\mathbf{Ir}_i(\mathbf{D})$ determines player i's possible decision $I_i(s_i)$ and his prediction $\mathbf{B}_j(I_j(s_j))$ to necessarily be $A_i^*(s_i)$ and $\mathbf{B}_j(A_j^*(s_j))$.

Based on Lemma 3.3, (5.5) can be written as two equivalent forms:

$$\vdash \mathbf{B}_i[\mathbf{Ir}_i(\mathbf{D}) \supset \wedge_{s \in S} \langle \mathbf{Ir}_i[I_1(s_1) \supset A_1^*(s_1), I_2(s_2) \supset A_2^*(s_2)] \rangle]; \tag{5.6}$$

$$\vdash \mathbf{B}_i[\mathbf{Ir}_i(\mathbf{D})] \supset \mathbf{B}_i[\wedge_{s \in S} \langle \mathbf{Ir}_i[I_1(s_1) \supset A_1^*(s_1), I_2(s_2) \supset A_2^*(s_2)] \rangle]. \tag{5.7}$$

The statement (5.6) is to cover $\mathbf{B}_i[\cdot]$ to (5.5), which sees the statement from the outside observer's perspective. Then, (5.7) is obtained by distributing $\mathbf{B}_i[\cdot]$, which is an statement from the scope of the outsider, and some inferences are made by the outsider. Mathematically speaking, by Lemma 3.3, we can substitute $\mathbf{B}_j[\cdot]$ for the outer $\mathbf{B}_i[\cdot]$ in (5.6) and (5.7). However, this substitution violates our intended interpretation because of the distinction (*i*) and (*ii*) in the beginning of this section.

5.3. Choice of the Weakest Formulae

Step (b) is to show that the candidates $\mathcal{A}_i^* = \{A_i^*(s_i) : s_i \in S_i\}, i = 1, 2$, satisfy $\mathbf{Ir}_i(\mathbf{D})$. We formulate this idea in logic GL_ω as follows: Let $\mathcal{A}_i = \{A_i(s_i) : s_i \in S_i\}$ be a class of formulae indexed with $s_i \in S_i$ for $i = 1, 2$. We substitute those formulae for the occurrences of $I_i(s_i), s_i \in S_i$ and $i = 1, 2$ in $\mathbf{Ir}_i(\mathrm{D}_1,\mathrm{D}_2)$. To be precise, we define $\mathrm{D0}_i(\mathcal{A})$ as the formula obtained from $\mathrm{D0}_i$ by substituting each $A_i(s_i)$ for the occurrences of $I_i(s_i)$ in $\mathrm{D0}_i$ for $s_i \in S_i$ and $i = 1, 2$. We define $\mathrm{D1}_i(\mathcal{A})$ and $\mathrm{D2}_i(\mathcal{A})$ in a parallel fashion. Let $\mathrm{D}_i(\mathcal{A}) = \mathrm{D0}_i(\mathcal{A}) \wedge \mathrm{D1}_i(\mathcal{A}) \wedge \mathrm{D2}_i(\mathcal{A})$[10].

We adopt the following axiom schema:

$\mathbf{WF}_i(\mathcal{A})$ (Choice of the Weakest Formulae) :

$$\mathrm{D}_i(\mathcal{A}) \wedge \mathbf{B}_j(\mathrm{D}_j(\mathcal{A})) \supset \wedge_{t \in S}\langle [A_i(t_i) \supset I_i(t_i)] \wedge \mathbf{B}_j[A_j(t_j) \supset I_j(t_j)]\rangle.$$

Thus, $\mathrm{WF}_i(\mathcal{A})$ states that if $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ satisfies D_i and $\mathbf{B}_j(\mathrm{D}_j)$, i.e., $\mathrm{D}_i(\mathcal{A})$ and $\mathbf{B}_j(\mathrm{D}_j(\mathcal{A})))$ hold, then each of $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ implies the corresponding formula of $\mathcal{I} = (\mathcal{I}_1, \mathcal{I}_2)$. In other words, $\mathcal{I} = (\mathcal{I}_1, \mathcal{I}_2)$ are the deductively weakest among the formulae satisfying D_i and $\mathbf{B}_j(\mathrm{D}_j)$. Then, we take the infinite regress $\mathbf{Ir}_i(\mathbf{WF}(\mathcal{A})) = \mathbf{Ir}_i(\mathrm{WF}_1(\mathcal{A}), \mathrm{WF}_2(\mathcal{A}))$. We denote $\{\mathbf{Ir}_i(\mathbf{WF}(\mathcal{A})) : \mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ is the pair of formulae indexed by $s_i \in S_i, i = 1, 2\}$ by $\mathbf{Ir}_i(\mathbf{WF})$.

As needed in Theorem 5.1, this step needs some additional condition corresponding to interchangeability (5.1). We first consider the following formula "subjective Nash": for $i = 1, 2$ and $s \in S$,

$$\mathrm{sash}_i(s) := \mathrm{best}_i(s_i; s_j) \wedge \mathbf{B}_j(\mathrm{best}_j(s_j; s_i)). \tag{5.8}$$

That is, from the perspective of player i, strategy s_i is a best against his prediction s_j and player j believes that s_j is a best against s_i. Using these formulae, we formulate "subjective interchangeability" in the perspective of player i :

$$\mathrm{Snt}_i := \wedge_{s \in S}\langle (\vee_{t_j \in S_j}\mathrm{sash}_i(s_i; t_j)) \wedge \mathbf{B}_j(\vee_{t_i \in S_i}\mathrm{sash}_j(s_j; t_i)) \supset \mathrm{best}_i(s_i; s_j)\rangle. \tag{5.9}$$

That is, player i's decision is a best response to any of his predictions about j, where "decision" and "prediction" are described in terms of subjective Nash. Let $\mathbf{Snt} = (\mathrm{Snt}_1, \mathrm{Snt}_2)$ and hence the infinite regress of these formulas is $\mathbf{Ir}_i(\mathbf{Snt})$. We have the following theorem, which has the converse conclusion of Theorem 5.2.

Theorem 5.3 (Choice of the Weakest Formulae):

$$\mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF}) \vdash \wedge_{s \in S}\langle \mathbf{Ir}_i[A_1^*(s_1) \supset I_1(s_1), A_2^*(s_2) \supset I_2(s_2)]\rangle. \tag{5.10}$$

[10] We may require each formula in $\mathcal{A}_i = \{A_i(s_i) : s_i \in S_i\}, i = 1, 2$ not to include $I_1(\cdot)$ and $I_2(\cdot)$ at all. This restriction is natural but we do not use it in the present paper.

The assumption $\mathbf{Ir}_i(\mathbf{WF})$ is a set of formulae, while $\mathbf{Ir}_i(\mathbf{Snt})$ is one formula. Mathematically speaking, in fact, only the single formula, $\mathbf{Ir}_i(\mathbf{WF}(\mathcal{A}^*))$, is used to have the conclusion of the theorem, which will be seen in the proof of the theorem in Section 7. Nevertheless, we should adopt $\mathbf{Ir}_i(\mathbf{WF})$ to allow any candidates satisfying D_i and $\mathbf{B}_j(\mathrm{D}_j)$.

Now, we combine Theorems 5.2 and 5.3.

Theorem 5.4 (Full Characterization 1):

$$\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF}) \vdash \wedge_{s \in S} \langle \mathbf{Ir}_i[I_1(s_1) \equiv A_1^*(s_1), I_2(s_2) \equiv A_2^*(s_2)] \rangle. \tag{5.11}$$

Theorem 5.4 gives a full epistemic characterization of prediction/decision making in terms of the infinite regresses of D0-D2, subjective interchangeability and the choice of the weakest formulae satisfying $\mathrm{D}_i \wedge \mathbf{B}_j(\mathrm{D}_j)$. We can regard this derivation occurs in the mind of player i; that is, player i explicate those infinite regresses $\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF})$, and his logical inferences lead to $A_i^*(s_i)$.

In Theorem 5.4, we do not use a particular game formula, but we use only $\mathbf{Ir}_i(\mathbf{Snt})$. Theorem 5.1 requires specific game structure with (5.1). In fact, we can change Theorem 5.4 in a parallel form. Let G be a game with (5.1), and $\mathbf{g} = (g_1, g_2)$ the pair of formulae defined by (5.3). Then, we have the following:

$$\vdash \mathbf{Ir}_i(\mathbf{g}) \supset \mathbf{Ir}_i(\mathbf{Snt}). \tag{5.12}$$

That is, the infinite regress of payoff functions for game G implies the infinite regress of subjective version of interchangeability, (5.9). Using this, Theorem 5.4 implies the following theorem:

Theorem 5.5 (Full Characterization 2):

$$\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{g}), \mathbf{Ir}_i(\mathbf{WF}) \vdash \wedge_{s \in S} \langle \mathbf{Ir}_i[I_1(s_1) \equiv A_1^*(s_1), I_2(s_2) \equiv A_2^*(s_2)] \rangle. \tag{5.13}$$

This theorem directly corresponds to Theorem 5.1. We can formalize Theorem 5.1 by ignoring all the occurrences of belief operators $\mathbf{B}_1(\cdot)$ and $\mathbf{B}_2(\cdot)$ in the above argument. Here, we make a connection from Theorem 5.5 to the formalized version of Theorem 5.1. This connection enables us to see the consistency of $\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{g}), \mathbf{Ir}_i(\mathbf{WF})$, *a fortiori,* $\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF})$, in GL_ω.

First, we introduce the (universal) eraser ε_0 to eliminate all $\mathbf{B}_i(\cdot), i = 1, 2$ in any given formula, which is defined in a similar manner to ε_i in the proof of Lemma 3.3. Then, Kaneko-Nagashima [8] proved: for any formula A,

$$\vdash A \text{ implies } \vdash_0 \varepsilon_0 A,$$

where $\vdash_0$ is the provability relation of classical logic. Then, the statement obtained from (5.13) by applying ε_0 is exactly the non-epistemic formalization of Theorem 5.1. Note that Axiom $\mathrm{D1}_i$ becomes a redundant tautology with the application of ε_0.

We can show the consistency (contadiction-freeness) of the premises $\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{g})$, $\mathbf{Ir}_i(\mathbf{WF})$, *a fortiori*, $\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF})$, in GL_ω in the same manner of the reduction by the eraser ε_0.

Kaneko [9] considered a certain epistemic characterization of the Nash solution in logic $\mathrm{GL}_\omega(\mathrm{T4})$; the resulting formula was given as $\vee_{t_j \in S_j} \mathbf{C}(\mathrm{nash}(s_i; t_j))$ under the common knowledge of interchangeability $\mathbf{C}(\mathrm{int})$. It follows from Theorem 4.1 that in $\mathrm{GL}_\omega(\mathrm{T})$, $\vdash A_i^*(s_i) \equiv \vee_{t_j \in S_j} \mathbf{C}(\mathrm{nash}(s_i; t_j))$ for $s_i \in S_i$. Theorem 5.4 becomes:

$$\mathbf{Ir}_i(\mathbf{D}), \mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF}) \vdash \mathbf{C}(\wedge_{i=1,2}[I_i(s_1) \equiv \vee_{t_j \in S_j} \mathbf{C}(\mathrm{nash}(s_i; t_j))]).$$

However, Kaneko [9] did not make the clear-cut distinction between decisions and predictions; he provided certain axioms similar to $\mathrm{D0}_i$-$\mathrm{D2}_i$ to characterize $\vee_{t_j \in S_j} \mathbf{C}(\mathrm{nash}(s_i; t_j))$. The axiomatization was possible without the clear-distinction because of Axiom T. In our approach in GL_ω, we need the clear-cut distinction between decisions and predictions; otherwise, there would be multiplicity in the axiomatization and in the characterization.

5.4. Necessity of Infinite Regresses

Here we give a few remarks on the necessity of various components in our approach, in particular, of the infinite regress of the axioms D0-D2 to obtain the full characterization of prediction/decision making. Those are all described as unprovability statements, which are proved by making counter models in Kripke semantics. We omit the proofs.

First, we let $\widehat{\mathbf{Ir}}_i^\nu(A_1, A_2) := \wedge\{\mathbf{Ir}_i^\kappa(A_1, A_2) : 0 \leq \kappa \leq \nu\}$, which is the regress formula up to depth ν. First, we have the following fact:

(a): For any natural number ν, $\widehat{\mathbf{Ir}}_i^\nu(\mathbf{D}) \nvdash \wedge_{s_i \in S_i}[I_i(s_i) \supset A_i^*(s_i)]$.

That is, if we stop the regress for $\mathbf{D} = (\mathrm{D}_1,\mathrm{D}_2)$ at finite ν, we cannot have the statement corresponding to Theorem 5.2, which means that the infinite regress of the premise is inevitable for the theorem. This is proved by using the *depth lemma* for KD^n (see Kaneko [10]) and the fact that GL_ω is an conservative extension of KD^n.

Notice that $A_i^*(s_i)$ includes the infinite regress, which makes us to have $\vdash A_i^*(s_i) \supset \mathbf{B}_j\mathbf{B}_i(A_i^*(s_i))$, that is, $A_i^*(s_i)$ satisfies Axiom $\mathrm{D1}_i$. Again, if we stop the regress at a finite ν in $A_i^*(s_i)$, this would not hold. Let $A_i^\nu(s_i) = \vee_{s_j \in S_j} \widehat{\mathbf{Ir}}_i^\nu(\mathrm{best}_i(s_i; s_j);\mathrm{best}_j(s_j; s_i))$ for $\nu \geq 0$. Then,

(b) For any natural number ν, $\nvdash A_i^\nu(s_i) \supset \mathbf{B}_j\mathbf{B}_i(A_i^\nu(s_i))$.

This is also proved by the depth lemma.

A related fact is hat if we delete Axiom D1 from $\mathbf{Ir}_i(\mathbf{D})$, Theorem 5.2 would fail to hold.

(c) $\mathbf{Ir}_i(\mathrm{D0}_1 \wedge \mathrm{D2}_1, \mathrm{D0}_2 \wedge \mathrm{D2}_2) \nvdash \wedge_{s_i \in S_i}[I_i(s_i) \supset A_i^\nu(s_i)]$ for all $\nu \geq 1$.

That is, even $A_i^{\nu}(s_i)$ with $\nu = 1$ is not derived if D1 is deleted from the premise.

6. Conclusions

First, we summarize the developments in the present paper in terms of steps IR1-IR3 in Section 1. Step IR1 was accomplished in Section 1, which discusses the infinite regresses arising from prediction/decision making in a game situation. Then, Sections 2, 3, and 4 form step IR2: We formulated the infinite regress formula in the infinitary logic GL_{ω}, and showed its basic properties as well as relations to the common knowledge formula, revealing the role of the two additional epistemic axioms, i.e., Axiom T and Axiom 4. Then, in Section 5, we applied the developed theory to the game theoretical prediction/decision making, which is step IR3. Although the theory and application already become a long discourse, many new developments may thrive from the current study both from the viewpoints of logic and game theory. Here, we give a few comments on further developments.

From the viewpoint of logic, GL_{ω} is an infinitary logic. As stated in Section 3, we can formulate the logic as a fixed-point logic, since Lemma 3.1 indicates the corresponding fixed-point approach. Also, since the discourse in this paper is based on Lemma 3,1, all the results after Lemma 3.1 can be reconstructed in the corresponding fixed-point logic. The infinitary logic and the fixed-point logic approaches are mutually complementary: The infinitary approach treats infinitary regresses in a faithful manner, while the fixed-point approach gives different merits as a finitary logic.

Although GL_{ω} is quite small relative to the standard infinitary logic approach such as Karp [12] and Heifetz [7], specific problems such as those discussed in Section 5 need even much smaller fragments of GL_{ω}. In particular, the infinite regresses arising from prediction/decision making needs only a small extension of the finitary KD^n. A precise evaluation of this "smallness" requires a hierarchy of logics up to GL_{ω} : Each logic can be proved to be Kripke complete based on the method developed in Tanaka-Ono [20]. In this approach, we can evaluate our axiomatization of prediction/decision making more precisely. Also, the status of the fixed-point approach will become clearer.

Here, we turn to the game theoretical problems. In this paper, we have not touched unsolvable games. Our approach, however, can be used to make a more substantive contributions in understanding prediction/decision-making in games. Although Nash [17] gave the definition of a subsolution, it gives no particular meaning of "unsolvability" beyond its mathematical definition. Our approach would produce an incompleteness result on prediction/decision making for an unsolvable game. It states that a player can prove neither that a given strategy is a final decision nor that it is not. Thus, unsolvability is expressed by an incompleteness result. This will be discussed in a separate paper.

Finally, our approach may be applied to other theories of *ex ante* prediction/decision-making in games, such as the rationalizability concept proposed by Bernheim and Pearce (cf., Osborne-Rubinstein [18]): the theory of rationalizable strategies has also a certain structure of infinite regress, but differs from our starting point N1^o-N2^o.

7. Proofs of Theorems 5.2 and 5.3

7.1. Proof of Theorem 5.2

We prove the theorem with the following claims:

Lemma 7.1. For any $i = 1, 2$ and $s \in S$,

(1) $\mathbf{Ir}_i(\mathrm{D1}_1, \mathrm{D1}_2) \vdash [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_j))] \supset \mathbf{Ir}_i(I_1(s_1), I_2(s_2))$;

(2) $\mathbf{Ir}_i(\mathrm{D0}_1, \mathrm{D0}_2) \vdash \mathbf{Ir}_i(I_1(s_1), I_2(s_2)) \supset \mathbf{Ir}_i(\mathrm{best}_1(s_1; s_2), \mathrm{best}_2(s_2; s_1))$;

(3): $\mathbf{Ir}_i(\mathrm{D01}_1,\mathrm{D01}_2) \vdash [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_j))] \supset \mathbf{Ir}_i(\mathrm{best}_1(s_1; s_2),\mathrm{best}_2(s_2; s_1))$;

Proof. (3) follows (1) and (2).

(1): Let $\alpha_i = \mathbf{Ir}_i(\mathrm{D1}_1,\mathrm{D1}_2) \wedge [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_j))]$. We show

$$\vdash \alpha_i \supset \mathbf{B}_j(\alpha_j). \tag{7.1}$$

Suppose that (7.1) is proved for $i = 1, 2$. Then, since $\vdash \alpha_i \supset I_i(s_i)$ for $i = 1, 2$, it follows from (7.1) and IRI that $\vdash \alpha_i \supset \mathbf{Ir}_i(I_1(s_1), I_2(s_2))$. This is (1).

Let us prove (7.1). By IRA,

$$\vdash \alpha_i \supset \mathbf{B}_j(\mathbf{Ir}_j(\mathrm{D1}_1, \mathrm{D1}_2)). \tag{7.2}$$

Then, we should show $\vdash \alpha_i \supset \mathbf{B}_j[I_j(s_i)) \wedge \mathbf{B}_i(I_i(s_i))]$, which and (7.2) imply (7.1). Since $\vdash I_i(s_i) \wedge \mathbf{B}_j(I_j(s_i)) \supset \mathbf{B}_j(I_j(s_i))$ and $\vdash \mathrm{D1}_i \supset (I_i(s_i) \supset \mathbf{B}_j\mathbf{B}_i(I_i(s_i))$, we have

$$\vdash \mathrm{D1}_i \supset [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_i) \supset \mathbf{B}_j(I_j(s_i)) \wedge \mathbf{B}_j\mathbf{B}_i(I_i(s_i))].$$

This is equivalent to $\vdash \mathrm{D1}_i \wedge I_i(s_i) \wedge \mathbf{B}_j(I_j(s_i)) \supset \mathbf{B}_j(I_j(s_i)) \wedge \mathbf{B}_j\mathbf{B}_i(I_i(s_i))$. Since $\vdash \alpha_i \supset \mathrm{D1}_i$ and $\vdash \alpha_i \supset I_i(s_i) \wedge \mathbf{B}_j(I_j(s_i))$, we have $\vdash \alpha_i \supset \mathbf{B}_j(I_j(s_i)) \wedge \mathbf{B}_j\mathbf{B}_i(I_i(s_i))$. Then, $\vdash \alpha_i \supset \mathbf{B}_j[I_j(s_i)) \wedge \mathbf{B}_i(I_i(s_i))]$ by Lemma 2.1.(8).

(2): Now, let $\alpha_i = \mathbf{Ir}_i(\mathrm{D0}_1,\mathrm{D0}_2)$ and $\beta_i = \mathbf{Ir}_i(I_1(s_1), I_2(s_2))$. We show that

$$\text{(a) } \vdash \alpha_i \wedge \beta_i \supset \mathrm{best}_i(s_i; s_j) \text{ and (b)} \vdash \alpha_i \wedge \beta_i \supset \mathbf{B}_j(\alpha_j \wedge \beta_j). \tag{7.3}$$

Once this is proved, we have, by IRI, $\vdash \alpha_i \wedge \beta_i \supset \mathbf{Ir}_i(\mathrm{best}_1(s_1; s_2), \mathrm{best}_2(s_2; s_1))$, which is equivalent to (2). Let us prove (a). Since $\vdash \alpha_i \supset \mathrm{D0}_i$, we have the conclusion

$\vdash \alpha_i \supset [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_j)) \supset \mathrm{best}_i(s_i; s_j)]$. Since $\vdash \beta_i \supset I_i(s_i)$ and $\vdash \beta_i \supset \mathbf{B}_j(I_j(s_j))$ by IRA, we have $\vdash \beta_i \supset I_i(s_i) \wedge \mathbf{B}_j(\beta_j)$. This together with the conclusion implies (a). Assertion (b) follows IRA and Lemma 2.1.(4). ■

Now, we can finish the proof of Theorem 5.2: Let s_i be an arbitrary element in S_i. We show that

$$\vdash \mathbf{Ir}_i(\mathbf{D}) \supset [I_i(s_i) \supset A_i^*(s_i)] \text{ for } i = 1, 2, \tag{7.4}$$

where $A_i^*(s_i) = \vee_{t_j \in S_j} \mathbf{Ir}_i(\mathrm{best}_1(s_i; t_j), \mathrm{best}_2(s_i; t_j))$. Since $\vdash \mathbf{Ir}_i(\mathbf{D}) \supset \mathbf{B}_j(\mathbf{Ir}_j(\mathbf{D}))$ by IRA, we obtain, by (7.4) and IRI, $\vdash \mathbf{Ir}_i(\mathbf{D}) \supset \mathbf{Ir}_i(I_1(s_1) \supset A_1^*(s_1), I_2(s_2) \supset A_2^*(s_2))$, which implies the assertion of Theorem 5.2.

Now we prove (7.4). By Lemma 7.1.(3), we have

$$\mathbf{Ir}_i(\mathbf{D}) \vdash [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_j))] \supset \mathbf{Ir}_i(\mathrm{best}_1(s_1; s_2), \mathrm{best}_2(s_2; s_1)). \tag{7.5}$$

By L5, this implies $\mathbf{Ir}_i(\mathbf{D}) \vdash [I_i(s_i) \wedge \mathbf{B}_j(I_j(s_j))] \supset \vee_{t_j \in S_j}[\mathbf{Ir}_i(\mathrm{best}_i(s_i; t_j); \mathrm{best}_j(t_j; s_i))]$. Since this holds for all $s_j \in S_j$, we have, by $\vee$-rule,

$$\mathbf{Ir}_i(\mathbf{D}) \vdash \vee_{t_j \in S_j}[I_i(s_i) \wedge \mathbf{B}_j(I_j(t_j))] \supset \vee_{t_j \in S_j}[\mathbf{Ir}_i(\mathrm{best}_i(s_i; t_j); \mathrm{best}_j(t_j; s_i))]. \tag{7.6}$$

Since $\mathrm{D2}_i \vdash I_i(s_i) \supset \vee_{t_j \in S_j} \mathbf{B}_j(I_j(t_j))$, we have $\mathrm{D2}_i \vdash I_i(s_i) \supset \vee_{t_j \in S_j}[I_i(s_i) \wedge \mathbf{B}_j(I_j(t_j))]$. Combining this with (7.6), we have (7.4). ■

7.2. Proof of Theorem 5.3

We begin with the following lemma.

Lemma 7.2.(1): $\mathrm{Snt}_i \vdash A_i^*(s_i) \wedge \mathbf{B}_j(A_j^*(s_j)) \supset \mathrm{best}_i(s_i; s_j)$;

(2): $\vdash A_i^*(s_i) \supset \mathbf{B}_j \mathbf{B}_i(A_i^*(s_i))$;

(3): $\vdash A_i^*(s_i) \supset \vee_{t_j \in S_j} \mathbf{B}_j(A_j^*(t_j))$.

Proof. We denote $\mathbf{Ir}_i(\mathrm{best}_1(s_1; s_2), \mathrm{best}_2(s_2; t_1))$ by $A_i(s_1, s_2)$.

(1): First, it holds that $\vdash A_i(s_1, s_2) \supset \mathrm{best}_i(s_i; s_j) \wedge \mathbf{B}_j(\mathrm{best}_i(s_i; s_j))$, i.e., $\vdash A_i(s_1, s_2) \supset \mathrm{sash}_i(s_i; s_j)$, for $i = 1, 2$. By applying L5 and $\vee$-rule successively to this, we have $\vdash \vee_{t_j} A_i(s_1, s_2) \supset \vee_{t_j} \mathrm{sash}_i(s_i; t_j)$, i.e.,

$$\vdash A_i^*(s_i) \supset \vee_{t_j} \mathrm{sash}_i(s_i; t_j) \text{ for } i = 1, 2. \tag{7.7}$$

Applying Nec and K to (7.7) for j, we have $\vdash \mathbf{B}_j(A_j^*(s_j)) \supset \mathbf{B}_j(\vee_{t_i} \mathrm{sash}_j(s_j; t_i))$. It follows from this and (7.7) that $\vdash A_i^*(s_i) \wedge \mathbf{B}_j(A_j^*(s_j)) \supset [\vee_{t_j} \mathrm{sash}_i(s_i; t_j) \wedge \mathbf{B}_j(\vee_{t_i} \mathrm{sash}_j(s_j; t_i))]$. The conclusion of this formula is the premise of the inside of Axiom Snt_i. Hence, using Snt_i, we have $\vdash A_i^*(s_i) \wedge \mathbf{B}_j(A_j^*(s_j)) \supset \mathrm{best}_i(s_i; s_j)$.

(2): Since $\vdash A_i(s_1, s_2) \supset \mathbf{B}_j(A_j(s_1, s_2))$ for $i = 1, 2$ by IRA, we have, by Nec and K, $\vdash \mathbf{B}_j(A_j(s_1, s_2)) \supset \mathbf{B}_j\mathbf{B}_i(A_i(s_1, s_2))$. Thus, $\vdash A_i(s_1, s_2) \supset \mathbf{B}_j\mathbf{B}_i(A_i(s_1, s_2))$. By L5, we have $\vdash A_i(s_1, s_2) \supset \vee_{t_j}\mathbf{B}_j\mathbf{B}_i(A_i(s_i;t_j))$. Applying Lemma 2.1.(7) and (1) twice, we have $\vdash A_i(s_1, s_2) \supset \mathbf{B}_j\mathbf{B}_i(\vee_{t_j}A_i(s_i;t_j))$. Since this holds for all $s_j \in S_j$, we have, using $\vee$-rule, $\vdash \vee_{t_j}A_i(s_1, s_2) \supset \mathbf{B}_j\mathbf{B}_i(\vee_{t_j}A_i(s_i;t_j))$. This is (2).

(3): Since $\vdash A_i(s_1, s_2) \supset \mathbf{B}_j(A_j(s_1, s_2))$ by IRA, we have, by L5, $\vdash A_i(s_1, s_2) \supset \vee_{t_i}\mathbf{B}_j(A_j(s_j;t_i))$. By Lemma Lemma 2.1.(7) and (1), we have $\vdash A_i(s_1, s_2) \supset \mathbf{B}_j(\vee_{t_i}A_j(s_j;t_i))$. By L5, we have $\vdash A_i(s_1, s_2) \supset \vee_{t_j}\mathbf{B}_j(\vee_{t_i}A_j(t_j;t_i))$. This holds for all $s_j \in S_j$. Hence, by $\vee$-rule, $\vdash \vee_{t_j}A_i(s_i;t_j) \supset \vee_{t_j}\mathbf{B}_j(\vee_{t_i}A_j(t_j;t_i))$, which is (3). ■

Now, we can finish the proof of Theorem 5.3. Recall $\mathcal{A}_i^* = \{A_i^*(s_i) : s_i \in S_i\}$ and $\mathcal{A}^* = (\mathcal{A}_1^*, \mathcal{A}_2^*)$. Lemma 7.2 shows that for $i = 1, 2$,

$$\mathrm{Snt}_i \vdash \mathrm{D}_i(\mathcal{A}^*).$$

Thus, by IRA, we have $\mathbf{Ir}_i(\mathbf{Snt}) \vdash \mathrm{D}_i(\mathcal{A}^*) \wedge \mathbf{B}_j(\mathrm{D}_j(\mathcal{A}^*))$, and hence

$$\mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF}(\mathcal{A}^*)) \vdash A_i^*(s_i) \supset I_i(s_i) \text{ for } i = 1, 2.$$

Using Lemma 3.2.(6), we have

$$\mathbf{Ir}_i(\mathbf{Snt}), \mathbf{Ir}_i(\mathbf{WF}(\mathcal{A}^*)) \vdash \mathbf{Ir}_i(A_1^*(s_1) \supset I_1(s_1), A_2^*(s_2) \supset I_2(s_2)).$$

Theorem 5.3 follows this. ■

References

[1] Aumann, R. J. (1976), Agreeing to Disagree, *Annals of Statistics* 4, 1236–1239.

[2] Aumann, R. J., and A. Brandenburger, (1995), Epistemic Conditions for Nash Equilibrium, *Econometrica* 63, 1161-1180.

[3] Chellas, B., (1980), *Modal logic.* Cambridge: Cambridge University Press.

[4] Fagin, R., J. Y. Halpern, Y. Moses and M. Y. Verdi, (1995), *Reasoning about Knowledge*, The MIT Press, Cambridge.

[5] Gratton, C., (2010), Infinite Regress Arguments, Springer, Hiderberg.

[6] Harsanyi, J. C. (1967/68), Games with Incomplete Information Played by Bayesian Players, I-III, *Management Science* 14, 159-182, 320-334, 486-502.

[7] Heifetz, A., (1999), Iterative and Fixed Point Common Belief, *Journal of Philosophical Logic* 28, 61-79.

[8] Kaneko, M., and T. Nagashima, (1996), Game logic and its applications I, *Studia Logica* 57, 325–354.

[9] Kaneko, M., (1999), Epistemic Considerations of Decision Making in Games. *Mathematical Social Sciences* 38, 105–137.

[10] Kaneko, M., (2002a), Epistemic logics and their game theoretical applications: Introduction. *Economic Theory* 19 (2002), 7-62.

[11] Kaneko, M., (2002b), *Game Theory and Mutual Misunderstanding*, Springer, Heidelberg.

[12] Karp, C. R. (1964), *Languages with Expressions of Infinite Length,* North Holland, Amsterdam.

[13] Lewis, D. K., (1969), *Convention:A Philosophical Study*, Harvard University Press.

[14] Luce, R. D. and H. Raiffa, (1957), *Games and Decisions*, John Wiley Sons, New York.

[15] Meyer, J.-J. Ch., van der Hoek, W., (1995), *Epistemic logic for AI and computer science*. Cambridge.

[16] Mendelson, E., (1988), *Introduction to Mathematical Logic*, Wadsworh, Monterey.

[17] Nash, J. F., (1951), Non-cooperative Games, *Annals of Mathematics* 54, 286-295.

[18] Osborne, M., and A. Rubinstein, (1994), *A Course in Game Theory*, MIT Press, Cambridge.

[19] Tan T., and S. D. C. Werlang, (1988), The Bayesian Foundations of Solution Concepts of Games, *Journal of Economic Theory* 45, 370-391.

[20] Tanaka, Y., and H. Ono, (1998), The Rasiowa-Sikorski lemma and Kripke Completeness of Predicate and Infinitary Modal Logics, *Advances in Modal Logics '98.*

[21] Van Benthem, *Logic in Games*, Institute for Logic, Language and Computation.

[22] Van Benthem, J., E. Pacuit, and O. Roy, (2011), Toward a Theory of a Play: A Logical Perspective on Games and Interaction, *Games* 2, 52-86.