WINPEC Working Paper Series No.E1421
February 2015

Epistemic Infinite-Regress Logic

Tai-Wei Hu and Mamoru Kaneko

Waseda INstitute of Political EConomy
Waseda University
Tokyo,Japan

Epistemic Infinite-Regress Logic*

Tai-Wei Hu† and Mamoru Kaneko‡

08 September 2015 (original 16 February 2015)

Abstract

We present a logic called the *epistemic infinite-regress logic* EIR^n for n players. It extends the epistemic logic KD^n, by incorporating the operators expressing infinite regresses. Here, an infinite regress arises from the interdependent thinking of a player about the other's thinking and *vice versa*. If we add Axiom T (truthfulness) for the belief operators to the logic EIR^n, the concept of an infinite regress collapses into the common knowledge. However, we keep the subjective nature of the concept, avoiding Axiom T as well as Axiom 4. We prove the completeness theorem with respect to Kripke semantics and a certain restricted class of Kripke models. These lead to various meta-theorems useful for considerations of prediction/decision making by a player in a game. Using the meta-theorems, we show that a statement about n players can be decomposed into n independent statements, *vice versa,* and even that we can go back and forth from a statement for each player into its epistemic content.

Key words: Infinite regress, Common knowledge, Epistemic logic, Nash equilibrium, Epistemic separation

1 Introduction

We develop the *epistemic infinite-regress logic* EIR^n, which is a fixed-point extension of (propositional) epistemic logic KD^n with n players. The logic is motivated by studies of both *ex ante* prediction/decision making and *ex post* observations in a game; Hu-Kaneko [5] applied the logic EIR^2 to study game-theoretic decidability/undecidability. We introduce new operators to express the concept of an (epistemic) infinite regress that naturally arises in prediction/decision making in a game. Because of interdependence of payoffs, for a player i to make his decision, i needs to make prediction about the decision of the other player, say, j. This prediction, however, also requires i's belief about j's predictions about i, and then the process continues *ad infinitum.* In the 2-person case, this infinite regress of beliefs can be represented as an infinite set of the following form:

$$\{\mathbf{B}_i(A_i), \mathbf{B}_i\mathbf{B}_j(A_j), \mathbf{B}_i\mathbf{B}_j\mathbf{B}_i(A_i), \mathbf{B}_i\mathbf{B}_j\mathbf{B}_i\mathbf{B}_j(A_j), ...\}, \tag{1}$$

*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, Kellogg School of Management, Northwestern University, Evanston, IL, (t-hu@kellogg.northwestern.edu)

‡Faculty of Political Science and Economics, Waseda University, Tokyo, Japan (mkanekoepi@waseda.jp)

where $\mathbf{B}_i(\cdot)$ and $\mathbf{B}_j(\cdot)$ and the belief operators of players i and j ($i \neq j$). In the logic EIR^2, this regress is expressed as the fixed-point operator $\mathbf{Ir}_i(A_i; A_j) := \mathbf{Ir}_i(A_1, A_2)$; for the n-person case, it is expressed as $\mathbf{Ir}_i(A_1, ..., A_n)$. Thus, the logic EIR^n is a fixed-point extension of the epistemic logic KD^n by adding one axiom schema and one inference rule for $\mathbf{Ir}_i(A_1, ..., A_n), i \in N = \{1, ..., N\}$.

The concept of an infinite regress $\mathbf{Ir}_i(A_1, ..., A_n)$ is closely related to that common knowledge. Indeed, it is shown in Section 2.3 that if we add Axiom T(truthfulness): $\mathbf{B}_i(A) \supset A$, then the infinite regress $\mathbf{Ir}_i(A_1, ..., A_n)$ collapses to the common knowledge of $A_1 \wedge ... \wedge A_n$ (Theorem 2.3), and the resulting logic is a common knowledge logic (cf., Fagin *et al.* [3] and Meyer-van der Hoek [14]). We also consider its status when we add Axiom 4 (positive introspection): $\mathbf{B}_i(A) \supset \mathbf{B}_i\mathbf{B}_i(A)$. Nevertheless, we take the KD-type EIR^n as our main logical system for various reasons, which are explained now.

First, we should mention a significant difference in the formulations of the logic EIR^n between the 2-person case and n-person case with $n \geq 3$. For the 2-person case, we can give essentially two different formulations for $\mathbf{Ir}_i(A_1, A_2), i = 1, 2$. One formulation is adopted in Hu-Kaneko [5], but is not available if $n \geq 3$. In this paper, we give another formulation, which is available for either $n = 2$ or $n \geq 3$; and it is equivalent to the formulation given in [5] for $n = 2$. We evaluate these formulations, while developing our theory.

We show that the logic EIR^n is complete with respect to the Kripke semantics (Theorem 3.1). For application purposes, we also present a variant of the completeness result (Theorem 3.2) such that we can restrict the class of models to that of rooted models. These completeness results enable us to develop various meta-theorems for our game-theoretic applications. Those meta-theorems reflect some fundamental principles relevant for a logical system describing players' subjective thinking in a game, and based on them, we evaluate various epistemic axioms based on these meta-theorems. Our choice of KD^n as our base logic is crucial to maintain those principles.

The completeness results show that the operator, $\mathbf{Ir}_i(A_1, ..., A_n)$, fully captures the set in (1). By this fact, it is expected that $\mathbf{Ir}_i(A_1, ..., A_n)$ is within the scope of the belief operator $\mathbf{B}_i(\cdot)$, since every formula in (1) has the outermost $\mathbf{B}_i(\cdot)$. Actually, $\mathbf{Ir}_i(A_1, ..., A_n)$ can be expressed by another equivalent formula, $\mathbf{B}_i[A_i \wedge (\wedge_{j \neq i}\mathbf{Ir}_j(\mathbf{A}))]$, in EIR^n (Theorem 2.2). This allows us to regard $\mathbf{Ir}_i(A_1, ..., A_n)$ as a belief formula for player i, and also to consider the derivability of the epistemic content of the infinite regress in his logical inferences, even though $\mathbf{Ir}_i(A_1, ..., A_n)$ is syntactically indecomposable.

One meta-theorem, called the *Scope Theorem* (Theorems 4.1), shows that $\mathbf{B}_i(A) \vdash \mathbf{B}_i(C)$ is equivalent to $A \vdash \dot{C}$. This equivalence is also applicable to our infinite regress operators (Theorem 4.3). This theorem allows us to change the epistemic scope from player i's subjective perspective to the analyst's, and *vice versa*. This change of scopes is critical in studies of interdependent subjective inferences for prediction/decision making. The Scope Theorem is also needed to evaluate the two different formulations of EIR^n.

Another meta-theorem, called the *Separation Theorem* (Theorem 4.2), shows that $\mathbf{B}_1(A_1) \wedge ... \wedge \mathbf{B}_n(A_n) \vdash \mathbf{B}_1(C_1) \wedge ... \wedge \mathbf{B}_n(C_n)$ is equivalent to $\mathbf{B}_i(A_i) \vdash \mathbf{B}_i(C_i)$ for all i. Thus, a statement about prediction/decision making as a whole can be decomposed into individualistic statements, and *vice versa*. This theorem separates individual subjectivities, in the sense that players' subjective beliefs are the only sources for their ultimate decisions. Thus, in our logic, we can explicitly distinguish the source of belief-changes. In particular, our formulation allows for a meaningful interaction between subjective beliefs and objective observations.

In Section 5, we show applications of our theory on EIR^n to *ex ante* prediction/decision making and their interactions with *ex post* observations. The decidability/undecidability results given in Hu-Kaneko [5] are mentioned, which concern *ex ante* prediction/decision making. While the analysis in Hu-Kaneko [5] is purely subjective and individualistic, our meta-theorems here allow us to decompose the entire situation that includes all players' perspectives and the objective situation into subjective reasonings. Moreover, given this decomposition, we also consider the interaction between *ex ante* subjective reasoning and *ex post* observations, and show that Nash equilibrium describes the situation where players' subjective inferences are consistent with observed behavior.

The scope theorem is newly given in this paper, while it is used in [5]. A primitive form of the separation theorem was proved in Kaneko-Nagashima [8] in an infinitary (predicate) epistemic logic (including Axiom 4) in a proof-theoretic manner, and a more sophisticated version was shown in a model theoretic manner in Kaneko-Suzuki [11] in an epistemic logic of shallow depths. The scope theorem crucially depends upon the choice KD^n as the base logic for EIR^n; we provide counterexamples to show that theorem fails if we add either Axiom T or Axiom 4. The separation theorem fails with Axiom T, but is compatible with Axiom 4.

The paper is organized as follows: Section 2 gives a Hilbert-style formulation of EIR^n and some basic lemmas, and some other variants. Then, Section 3 gives the Kripke semantics and the basic completeness (/soundness) theorem, the ep-rooted completeness theorem, and as an application, we prove scope lemma. Section 4 gives the epistemic separation theorem. Section 5 is a game theoretic application. Section 6 gives the proof of the basic completeness theorem.

2 Epistemic Infinite-Regress Logic EIR^n

Here we formulate the *infinite-regress logic* EIR^n as a fixed-point extension of epistemic logic KD^n with n players. In Section 2.1, we also mention an alternative formulation of EIR^n and that for 2 players given in Hu-Kaneko [5]. In Section 2.2, we give basic lemmas which we utilize for subsequent arguments; Theorem 2.2 (Epistemic content) is specific to the concept of an infinite regress, and plays crucial roles in subsequent arguments. In Section 2.3, we compare EIR^n with the common knowledge logic CKL^n with the presence of Axiom T or Axiom 4.

2.1 Formal system

The language for the infinite-regress logic EIR^n is as follows:

propositional variables: $\mathbf{p}_0, \mathbf{p}_1, ...$; *logical connectives*: $\neg$ (not), $\supset$ (imply), $\wedge$ (and), $\vee$ (or);

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

n-ary infinite regress operators: $\mathbf{Ir}_1(\cdot, ..., \cdot), ..., \mathbf{Ir}_n(\cdot, ..., \cdot)$; *parentheses*: (,).

We denote $PV := \{\mathbf{p}_0, \mathbf{p}_1, ...\}$. The set of players, whose generic element i appears as the subscript of belief operator $\mathbf{B}_i(\cdot)$ and infinite regress operators $\mathbf{Ir}_i(\cdot, ..., \cdot)$, is denoted by $N = \{1, ..., n\}$.

We define the sets of *formulae*, denoted $\mathcal{F}$, by the following induction: (o): each $p \in PV$

is a formula; (i): if A, B are formulae, so are $(A \supset B)$, $(\neg A)$, $\mathbf{B}_i(A)$ for $i \in N$; (ii): if $\mathbf{A} = (A_1, ..., A_n)$ is a vector of formulae, then $\mathbf{Ir}_i(\mathbf{A})$ is also a formula for $i \in N$; (iii): if Φ is a finite (nonempty) set of formulae, then $(\wedge\Phi)$ and $(\vee\Phi)$ are formulae. The set of all formulae is denoted by $\mathcal{F}$, and the set of all nonepistemic formulae, i.e., the ones with no occurrences of $\mathbf{B}_i(\cdot)$ and $\mathbf{Ir}_i(\cdots)$ for $i \in N$, is denoted by $\mathcal{F}^0$.

We abbreviate parentheses when no confusions are expected, and use different parentheses such as [,] for convenience. Also, $\wedge\{A_1, ..., A_n\}$ may be expressed as $A_1 \wedge ... \wedge A_n$, etc. We also abbreviate $(A \supset B) \wedge (B \supset A)$ as $A \equiv B$. We stipulate that when we talk about player i, the other players are denoted by $-i$. The formula $\mathbf{Ir}_i(\mathbf{A}) = \mathbf{Ir}_i(A_1, ..., A_n)$ is also denoted as $\mathbf{Ir}_i(A_i; A_{-i})$. When we refer to a contradictory formula, we will use $(\neg p) \wedge p$, where p is some propositional variable.

The base logic of EIR^n is classical logic, formulated by five axiom (schemata) and three inference rules: for all formulae A, B, C, and finite nonempty sets Φ of formulae,

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

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

Now, we add two epistemic axioms and one inference rule for the belief operators $\mathbf{B}_i(\cdot)$: for all formulae A, C, and for $i = 1, ..., n$,

K: $\mathbf{B}_i(A \supset C) \supset (\mathbf{B}_i(A) \supset \mathbf{B}_i(C))$; **D:** $\neg\mathbf{B}_i(\neg A \wedge A)$;

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

Those axioms and inference rules constitute epistemic logic KD^n.

The *epistemic infinite-regress logic* EIR^n is defined as the system by adding the following axiom schemata and rules to KD^n : for $i \in N$, and any vectors of formulae, $\mathbf{A} = (A_1, ..., A_n)$ and $\mathbf{D} = (D_1, ..., D_n)$,

$\mathbf{IRA}_i : \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\left[\wedge_{j \neq i}\mathbf{Ir}_j(\mathbf{A})\right]$;

$$\mathbf{IRI}_i : \frac{\{D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(\wedge_{k \neq j} D_k) : j \in N\}}{D_i \supset \mathbf{Ir}_i(\mathbf{A})}.$$

The names, IRA$_i$ and IRI$_i$, stand for "infinite regress axiom" and "infinite-regress inference". Note that these are assumed for all $i \in N$. Axiom IRA$_i$ requires a fixed-point property for $\mathbf{Ir}_i(\mathbf{A})$ in the interactive fashion that it contains $\wedge_{j \neq i}\mathbf{Ir}_j(\mathbf{A})$ within the scope of $\mathbf{B}_i(\cdot)$. Rule IRI$_i$ states that if $D_1, ..., D_n$ share the property described by Axioms IRA$_1$, ...,IRA$_n$, then D_i implies $\mathbf{Ir}_i(\mathbf{A})$.

In fact, $\mathbf{Ir}_i(\mathbf{A})$ induces, with the help of the other $\mathbf{Ir}_j(\mathbf{A})$, $j \neq i$, the following infinite set:

$$\begin{aligned} \{\mathbf{B}_{i_0}\mathbf{B}_{i_1}...\mathbf{B}_{i_k}(A_{i_k}) \quad &: \quad (i_0, i_1, ..., i_k) \text{ is a sequence of any finite length in } N \\ \text{with } i_t \quad &\neq \quad i_{t+1} \text{ for } t = 0, ..., k-1 \text{ and } i_0 = i\}. \end{aligned} \tag{2}$$

This is the n-person version of (1). To see that each formula in (2) is logically implied by $\mathbf{Ir}_i(\mathbf{A})$, first note that Axiom IRA$_i$ itself implies $\mathbf{B}_i(A_i)$ and $\mathbf{B}_i(\mathbf{Ir}_j(\mathbf{A}))$. For player i to explicate his belief $\mathbf{B}_i(\mathbf{Ir}_j(\mathbf{A}))$, he needs $\mathbf{B}_i(\text{IRA}_j)$, which he has by Nec. Thus, we obtain $\mathbf{B}_i\mathbf{B}_j(A_j)$ from $\mathbf{B}_i(\mathbf{Ir}_j(\mathbf{A}))$ together with $\mathbf{B}_i(\text{IRA}_j)$. Repeating this argument, we obtain the whole set in (2) as logical implications of $\mathbf{Ir}_i(\mathbf{A})$.

The upper formulae of Rule IRI_i states that $D_1, ..., D_n$ have the properties described by $\text{IRA}_1, ..., \text{IRA}_n$. Rule IRI_i requires $\mathbf{Ir}_i(\mathbf{A})$ to be deductively weakest among such D_i's. These rules, as shown later with the help of the completeness/soundness theorem, implies that $\mathbf{Ir}_i(\mathbf{A})$ is semantically determined by the set in (2).

A *proof* $P = \langle X, <; \psi\rangle$ in EIR^n is constituted of a finite tree $\langle X, <\rangle$ and a function $\psi : X \rightarrow \mathcal{P}$ with the following three requirements: (i) for each node $x \in X$, $\psi(x)$ 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 $\langle X, <\rangle$, the following is an instance of the above five rules:

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

We call P *a proof of* A iff A is attached to the root of a proof P. We say that A is *provable*, denoted by $\vdash A$, iff there is a proof of A. As stated in Section 1, we need nonlogical axioms for game theoretical applications. We introduce nonlogical assumptions in the following manner: For a set of formulae Γ, we write $\Gamma \vdash A$ iff $\vdash A$ or there is a finite nonempty subset Φ of Γ such that $\vdash \wedge\Phi \supset A$[1]. We say that a set of formulae Γ is *inconsistent* iff $\Gamma \vdash (\neg p) \wedge p$; and Γ is *consistent* otherwise.

As it stands, it may be difficult to interpret rule IRI_i as an inference rule from player i's subjective perspective, because $\mathbf{B}_j[\wedge_{k\neq j}D_k], j \neq i$, appear in the upper formulae of IRI_i. However, we can modify IRI_i as follows:

$$\mathbf{IRI}_i^{\text{Ind}} : \frac{D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i(\wedge_{j\neq i}D_k) \qquad \{\mathbf{B}_i[D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(\wedge_{k\neq j}D_k)] : j \neq i\}}{D_i \supset \mathbf{Ir}_i(\mathbf{A})}.$$

That is, it asserts that when $D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i(\wedge_{j\neq i}D_k)$ and $\mathbf{B}_i[D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(\wedge_{k\neq j}D_k)]$ are are already proved, the lower formula $D_i \supset \mathbf{Ir}_i(\mathbf{A})$ is provable, too. This inference rule is purely subjective from player i's perspective.

Note that Rule IRI_i is permissible in the system with the rule $\text{IRI}_i^{\text{Ind}}$, and hence the later system is deductively stronger than EIR^n. Actually, we have the converse, which is stated in Theorem 2.1.(i) below. We prove this converse using the soundness/completeness theorem for EIR^n in Section 4. Although the system with rule $\text{IRI}_i^{\text{Ind}}$ is better interpreted than EIR^n, they are equivalent and EIR^n is simpler. For this reason, we discuss mainly the logic EIR^n.

Both Axiom IRA_i and Rule IRI_i (and $\text{IRI}_i^{\text{Ind}}$) are interactive in the sense that each includes $\mathbf{Ir}_j(\mathbf{A})$ or D_j for j different from i. For $n = 2$, however, intrapersonal versions of the axiom and rule are given by Hu-Kaneko [5]: for $i = 1, 2$,

$$\mathbf{IRA}_i^{\text{HK}} : \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j\mathbf{Ir}_i(\mathbf{A});$$

$$\mathbf{IRI}_i^{\text{HK}} : \frac{D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j(D_i)}{D_i \supset \mathbf{Ir}_i(\mathbf{A})}.$$

Although the system with IRA_i^{HK} and IRI_i^{HK} is more appealing from the viewpoint of individual inference than EIR^2, only the 2-person version of such system is deductively equivalent to EIR^n,

[1]Given this definition, take Γ as the set in (2), we do not have $\Gamma \vdash \mathbf{Ir}_i(\mathbf{A})$. Hu-Kaneko-Suzuki (2014) permits infinite conjunction of the set in (2) in an infinitary logic.

as stated Theorem 2.1.(ii). In the extension of $\mathrm{IRA}_i^{\mathrm{HK}}$ and $\mathrm{IRI}_i^{\mathrm{HK}}$ to the three player case, for example, $\mathbf{B}_1\mathbf{B}_2\mathbf{B}_1\mathbf{B}_2(A_2)$ is not derived from $\mathbf{Ir}_1(A_1, A_2, A_3)$. The proof will be given in Section 3.3.

Theorem 2.1. *(Equivalent formulations of EIR^n) (i): For any formula $A \in \mathcal{F}$, $\vdash A$ in EIR^n if and only if $\vdash A$ in the system with rule IRI$_i^{Ind}$ instead of IRI$_i$.*

(ii): Let $n = 2$. For any formula $A \in \mathcal{F}$, $\vdash A$ in EIR^n if and only if $\vdash A$ in the system with IRA$_i^{HK}$ and IRI$_i^{HK}$ instead of IRA$_i$ and IRI$_i$.

2.2 Basic Properties of EIR^n

Here, we list some known facts on KD^n, which will be used without referring. The first two are provable formulae in classical logic and the other three are provable in KD^n. In the following, A, B, C are arbitrary formulae, Φ an arbitrary finite nonempty set, and $i = 1, ..., n$.

Lemma 2.1. ***(i)*** $\vdash [A \supset (B \supset C)] \equiv [A \wedge B \supset C)]$; ***(ii)*** $\vdash A \supset B$ *and* $\vdash B \supset C$ *imply* $\vdash A \supset C$; ***(iii)*** $\vdash \mathbf{B}_i(\neg A) \supset \neg\mathbf{B}_i(A)$; ***(iv)*** $\vdash \vee\mathbf{B}_i(\Phi) \supset \mathbf{B}_i(\vee\Phi)$; ***(v)*** $\vdash \mathbf{B}_i(\wedge\Phi) \equiv \wedge\mathbf{B}_i(\Phi)$.

As mentioned previously, $\mathbf{Ir}_i(\mathbf{A}) = \mathbf{Ir}_i(A_1, ..., A_n)$ intends to capture the set in (2), all formulae in which are within the scope of $\mathbf{B}_i(\cdot)$. We may then regard $\mathbf{Ir}_i(\mathbf{A})$ as having the outer $\mathbf{B}_i(\cdot)$, even though $\mathbf{Ir}_i(\cdot, ..., \cdot)$ is a primitive symbol. The following theorem formally justifies this interpretation. It states that $\mathbf{Ir}_i(\mathbf{A})$ is logically equivalent to another formula that begins with $\mathbf{B}_i(\cdot)$, and it gives the epistemic content of $\mathbf{Ir}_i(\mathbf{A})$. This theorem also justifies our claim that the operator is $\mathbf{Ir}_i(\cdot, ..., \cdot)$ captures player i's subjective thinking.

Theorem 2.2. ***(Epistemic Content)****:* $\vdash \mathbf{Ir}_i(\mathbf{A}) \equiv \mathbf{B}_i[A_i \wedge (\wedge_{j\neq i}\mathbf{Ir}_j(\mathbf{A}))]$.

Proof. The one direction $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_i[A_i \wedge (\wedge_{j\neq i}\mathbf{Ir}_j(\mathbf{A}))]$ follows from IRA$_i$. Consider the converse. Let $D_k = \mathbf{B}_k(A_k) \wedge \mathbf{B}_k(\wedge_{j\neq k}\mathbf{Ir}_j(\mathbf{A}))$ for $k \in N$. We show $\vdash D_k \supset \mathbf{B}_k(A_k) \wedge \mathbf{B}_k(\wedge_{j\neq k}D_j)$. It suffices to show $\vdash D_k \supset \mathbf{B}_k(\wedge_{j\neq k}D_j)$. Let $l \neq k$. By IRA$_k$ (Nec and K), $\vdash \mathbf{B}_k(\mathbf{Ir}_l(\mathbf{A})) \supset \mathbf{B}_k\mathbf{B}_l(A_l) \wedge \mathbf{B}_k\mathbf{B}_l(\wedge_{t\neq l}\mathbf{Ir}_t(\mathbf{A}))$, i.e., $\vdash \mathbf{B}_k(\mathbf{Ir}_l(\mathbf{A})) \supset \mathbf{B}_k(\mathbf{B}_l(A_l) \wedge \mathbf{B}_l(\wedge_{t\neq l}\mathbf{Ir}_t(\mathbf{A}))$, which is written as $\vdash \mathbf{B}_k(\mathbf{Ir}_l(\mathbf{A})) \supset \mathbf{B}_k(D_l)$. Since this holds fro all $l \neq k$, we have $\vdash \mathbf{B}_k(\wedge_{l\neq k}\mathbf{Ir}_l(\mathbf{A})) \supset \mathbf{B}_k(\wedge_{l\neq k}D_l)$. Thus, $\vdash D_k \supset \mathbf{B}_k(\wedge_{l\neq k}D_l)$; so $\vdash D_k \supset \mathbf{B}_k(A_k) \wedge \mathbf{B}_k(\wedge_{l\neq k}D_l)$. Since this holds for all $k \in N$, we can apply IRI$_i$, and thus we have $\vdash \mathbf{B}_i(A_i) \wedge \mathbf{B}_i[\wedge_{j\neq i}\mathbf{Ir}_j(\mathbf{A})] \supset \mathbf{Ir}_i(\mathbf{A})$.■

Here, we state various basic properties of $\mathbf{Ir}_i(\cdot, ..., \cdot)$. We write $\mathbf{A} \supset \mathbf{B}$ for $(A_1 \supset B_1, ..., A_n \supset B_n)$. We also write $\mathbf{Ir}_i(\Phi; A_{-i}) := \{\mathbf{Ir}_i(C_i; A_{-i}) : C_i \in \Phi\}$. The properties (i)-(iv) are inherited from the belief operator $\mathbf{B}_i(\cdot)$ satisfying Axioms K and D and Rule Nec. In particular, (i) corresponds to Axiom K, (ii) corresponds to Lemma 2.1.(iii), (iii) corresponds to Lemma2.1.(iv) and (v), and (iv) corresponds to the Rule Nec. Finally, (v) follows from the negation of the Axiom IRA$_i$.

Lemma 2.2. ***(Basic properties of*** $\mathbf{Ir}_i(\cdot, ..., \cdot)$***)****: Let $\mathbf{A}, \mathbf{C}$ be any n-tuples of formulae, and Φ a finite nonempty set of formulae. Then,*

(i)*:* $\vdash \mathbf{Ir}_i(\mathbf{A} \supset \mathbf{B}) \supset (\mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{C}))$.

(ii)*: Let $\mathbf{C}^* = (C_1^*, ..., C_n^*)$, where C_j^* is either C_j or $\neg C_j$ for all $j \in N$. Then, if $C_i^* = \neg C_i$, then* $\vdash \mathbf{Ir}_i(\neg C_i; \mathbf{C}^*_{-i}) \supset \neg\mathbf{Ir}_i(C_i; \mathbf{C}^*_{-i})$*, and if $C_j^* = \neg C_j$ for some $j \in N$, then* $\vdash \mathbf{Ir}_i(\mathbf{C}^*) \supset \neg\mathbf{Ir}_i(\mathbf{C})$.

(iii)*:* $\vdash \mathbf{Ir}_i(\wedge\Phi; \mathbf{A}_{-i}) \equiv \wedge\mathbf{Ir}_i(\Phi; \mathbf{A}_{-i})$*; and* $\vdash \vee\mathbf{Ir}_i(\Phi; \mathbf{A}_{-i}) \supset \mathbf{Ir}_i(\vee\Phi; \mathbf{A}_{-i})$.

(iv): *if* $\mathbf{Ir}_i(\mathbf{A}) \vdash \mathbf{B}_i(C_i) \wedge \mathbf{B}_i[\wedge_{j\neq i}\mathbf{B}_j(C_j)]$, *then* $\mathbf{Ir}_i(\mathbf{A}) \vdash \mathbf{Ir}_i(\mathbf{C})$; *particularly, if* $\vdash C_j$ *for all* $j \in N$, *then* $\vdash \mathbf{Ir}_i(\mathbf{C})$;

(v): *if* $\mathbf{Ir}_i(\mathbf{A}) \vdash \neg[\mathbf{B}_i(C_i) \wedge \mathbf{B}_i(\wedge_{j\neq i}\mathbf{B}_j(C_j))]$, *then* $\mathbf{Ir}_i(\mathbf{A}) \vdash \neg\mathbf{Ir}_i(\mathbf{C})$.

Proof. Those can be proved by choosing appropriate formulae for $\{D_k\}_{k\in N}$ in IRI$_i$. We prove only (i) and (v).

(i): For each $j = 1, ..., n$, let $D_j = \mathbf{Ir}_j(\mathbf{A} \supset \mathbf{C}) \wedge \mathbf{Ir}_j(\mathbf{A})$. Then, by IRA$_j$, $\vdash D_j \supset \mathbf{B}_j((A_j \supset C_j) \wedge A_j)$, which implies $\vdash D_j \supset \mathbf{B}_j(C_j)$. Again by IRA$_j$, we have $\vdash D_j \supset \mathbf{B}_j(\wedge_{k\neq j}D_k)$. Thus, $\vdash D_j \supset \mathbf{B}_j(C_j) \wedge \mathbf{B}_j(\wedge_{k\neq j}D_k)$; i.e., we get the upper formulae of IRI$_i$. Hence, $\vdash D_i \supset \mathbf{Ir}_i(\mathbf{C})$. This implies (i).

(v): Suppose $\mathbf{Ir}_i(\mathbf{A}) \vdash \neg[\mathbf{B}_i(C_i) \wedge \mathbf{B}_i(\wedge_{j\neq i}\mathbf{B}_j(C_j))]$. This is equivalent to $\mathbf{Ir}_i(\mathbf{A}), \mathbf{B}_i(C_i) \wedge \mathbf{B}_i(\wedge_{j\neq i}\mathbf{B}_j(C_j)) \vdash \neg p \wedge p$. By IRA$_i$, we have $\mathbf{Ir}_i(\mathbf{A}), \mathbf{Ir}_i(\mathbf{C}_i) \vdash \neg p \wedge p$. Thus, $\mathbf{Ir}_i(\mathbf{A}) \vdash \neg\mathbf{Ir}_i(\mathbf{C})$.■

2.3 Common knowledge, and Axioms T and 4

The concept of infinite regress is closely related to common knowledge due to Lewis [13] and Aumann [1]. To discuss common knowledge in our logic, we add one unary operator symbol $\mathbf{C}(\cdot)$ to the list of primitive symbols in Section 2 and extend the set of formulae by allowing $\mathbf{C}(A)$. Then, we add Axiom CKA and Inference CKI to IR^n : for any formulae A and D,

CKA: $\mathbf{C}(A) \supset A \wedge (\mathbf{B}_1(\mathbf{C}(A)) \wedge ... \wedge \mathbf{B}_n(\mathbf{C}(A)))$;

CKI: $\dfrac{D \supset A \wedge (\mathbf{B}_1(D) \wedge ... \wedge \mathbf{B}_n(D))}{D \supset \mathbf{C}(A)}$.

These also form a fixed-point definition of $\mathbf{C}(A)$. The logical implications of $\mathbf{C}(A)$ include all formulae in the following set:

$$\{A\} \cup \{\mathbf{B}_{i_0}\mathbf{B}_{i_1}...\mathbf{B}_{i_k}(A) : (i_0, i_1, ..., i_k) \text{ is a sequence of any finite length in } N\}. \tag{3}$$

This is obtained by repeated uses of CKA. By the completeness/soundness theorem for CKL, in terms of semantics, this set is exactly captured by $\mathbf{C}(A)$.

The set (3) makes it clear that $\mathbf{C}(A)$ is formulated from the outside analyst's perspective, as it includes all players' subjective thinking as well as the objective situation. It therefore differs from the infinite regress, $\mathbf{Ir}_i(A, ..., A)$, which is formulated purely from player i's subjective perspective. However, this subjectivity would disappear if we impose the truth axiom that is typically assumed in CKL. To make this point formally, we suppose that the logic EIR^n includes $\mathbf{C}(\cdot)$ for Axiom CKA and Rule CKI as stated above. We denote by EIR^n(T) by assuming **Axiom T**: $\mathbf{B}_i(A) \supset A$ $(A \in \mathcal{F})$. The following theorem shows that the infinite regress $\mathbf{Ir}_i(A_1, ..., A_n)$ collapses to the common knowledge $\mathbf{C}(A_1 \wedge ... \wedge A_n)$ in the logic EIR^n(T), and hence the subjectivity is destroyed.

Theorem 2.3. ***(Collapse under Axiom T):*** *For any* $i \in N$ *and* n*-tuple* $(A_1, ..., A_n)$ *of formulae,*

$$EIR^n(T) \ \vdash \mathbf{Ir}_i(A_1, ..., A_n) \equiv \mathbf{C}(A_1 \wedge ... \wedge A_n). \tag{4}$$

Proof*. We prove* $\vdash \mathbf{Ir}_i(A_1, ..., A_n) \supset \mathbf{C}(A_1 \wedge ... \wedge A_n)$. *We have, from Axiom IRA*$_i$*, using Axiom T several times, that* $\vdash \mathbf{Ir}_i(A_1, ..., A_n) \supset (A_1 \wedge ... \wedge A_n) \wedge [\wedge_{j\in N}\mathbf{B}_j(\mathbf{Ir}_j(A_1, ..., A_n))]$. *This and CKI imply* $\vdash \mathbf{Ir}_i(A_1, ..., A_n) \supset \mathbf{C}(A_1 \wedge ... \wedge A_n)$. *The converse is proved similarly.*■

Without imposing Axiom T, a closely related formula to $\mathbf{Ir}_i(A, ..., A)$ by using the operator $\mathbf{C}(\cdot)$ is $\mathbf{B}_i(\mathbf{C}(A))$, which also describes player i's subjective perspective. The formula $\mathbf{B}_i(\mathbf{C}(A))$ then corresponds to the set of formulae in (3) but adding the operator $\mathbf{B}_i(\cdot)$ before each of them. The key difference between the two, however, lies in the fact that $\mathbf{B}_i(\mathbf{C}(A))$ includes all formulae of the form $\mathbf{B}_{i_0}\mathbf{B}_{i_1}...\mathbf{B}_{i_k}(A)$ for any finite sequence $(i_0, i_1, ..., i_k)$ with $i_0 = i$, while $\mathbf{Ir}_i(A, ..., A)$ only includes those with an alternating epistemic structure. In fact, our next theorem shows that the two formulae are equivalent if we add **Axiom 4 (Positive Introspection)**, $\mathbf{B}_i(A) \supset \mathbf{B}_i\mathbf{B}_i(A)$ ($A \in \mathcal{F}$ and $i \in N$), to EIR^n. We denote the logic obtained from EIR^n by EIR^n(4) by adding Axiom 4.

Theorem 2.4. ***(Belief of common knowledge under Axiom 4)****: for any formula A and $i \in N$.*

$$EIR^n(4) \ \vdash \mathbf{Ir}_i(A, ..., A) \equiv \mathbf{B}_i\mathbf{C}(A) \tag{5}$$

Proof. We abbreviate $\mathbf{Ir}_i(A, ..., A)$ as $\mathbf{Ir}_i(A)$. Let us prove $\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_i\mathbf{C}(A)$. Consider the formula $D = A \wedge \mathbf{Ir}_1(A) \wedge ... \wedge \mathbf{Ir}_n(A)$. Since $\vdash \mathbf{Ir}_i(A) \equiv \mathbf{B}_i(A \wedge (\wedge_{j\neq i}\mathbf{Ir}_i(A)))$ by Theorem 2.2, we have $\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_i(A)$, $\vdash \mathbf{Ir}_i(A) \equiv \mathbf{B}_i(\wedge_{j\neq i}\mathbf{Ir}_i(A)))$, and finally $\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_i(\mathbf{Ir}_i(A)))$ by Axiom 4. By combining these three, we have $\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_i(A \wedge (\wedge_{j\in N}\mathbf{Ir}_j(A)))$, i.e.,

$$\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_i(D). \tag{6}$$

Since this holds for any $i \in N$, we have $\vdash D \supset A \wedge (\wedge_{j\in N}\mathbf{B}_j(D))$, which is an upper formula of CKA. Hence, $\vdash D \supset \mathbf{C}(A)$. This implies $\vdash \mathbf{B}_i(D) \supset \mathbf{B}_i\mathbf{C}(A)$ by Nec and K. This together with (6) implies $\vdash \mathbf{Ir}_i(A) \supset \mathbf{B}_i\mathbf{C}(A)$.

The converse is simple: Let $D = \mathbf{B}_i\mathbf{C}(A)$. Then, $\vdash D \supset \mathbf{B}_i(A) \wedge \mathbf{B}_i(\wedge_{j\neq i}\mathbf{B}_j(A))$ by CKI. By IRI$_i$, we have $\vdash D \supset \mathbf{Ir}_i(A)$, i.e., $\vdash \mathbf{B}_i\mathbf{C}(A) \supset \mathbf{Ir}_i(A)$.■

Without Axiom 4, however, the two formulae $\mathbf{B}_i\mathbf{C}(A)$ and $\mathbf{Ir}_i(A, ..., A)$ behave differently. For many applications, including our game theoretical ones given in Section 5, the infinite regress operator with its straightforward interpretation arises naturally in such a context and has a convenient mathematical structure. In contrast, it is difficult to interpret the arbitrary rounds of self-introspection involved in $\mathbf{B}_i\mathbf{C}(A)$ in such contexts, and, without Axiom 4, they also bring in cumbersome epistemic structures that obscure the epistemic analysis in games.

3 Kripke Semantics and Completeness

We give the Kripke semantics for EIR^n, which is the same as the semantics for KD^n with the additional valuation for $\mathbf{Ir}_i(\cdot, ..., \cdot)$ for $i \in N$. In Section 3.1, we give the basic (soundness-) completeness theorem, and in Section 3.2, we give a restriction on the semantics to facilitate further discussions.

3.1 Basic completeness theorem

A Kripke frame $F = \langle W; R_1, ..., R_n\rangle$ consists of a nonempty set W of possible worlds and an accessibility relation $R_i \subseteq W \times W$ for player $i \in N$. We say that a frame $F = \langle W; R_1, ..., R_n\rangle$ is *serial* iff for $i = 1, ..., n$ and for all $w \in W$, wR_iu for some $u \in W$. A *truth assignment* τ is a function from $W \times PV$ to $\{\top, \bot\}$. A pair $M = (F, \tau)$ is called a *model*. When F is serial, we say that M is a *serial model*.

To define the semantic valuation, we need the concept of an alternating chain. We call a sequence $[(v_0, i_0), (v_1, i_1), ..., (v_m, i_m), v_{m+1}]$ a *chain* iff $v_0, v_1, ..., v_{m+1} \in W$, $i_0, i_1, ..., i_m \in N$ and $(v_t, v_{t+1}) \in R_{i_t}$ for $t = 0, ..., m$. We say that a chain $[(v_0, i_0), (v_1, i_1), ..., (v_m, i_m), v_{m+1}]$ is *alternating* iff $i_t \neq i_{t+1}$ for $t = 0, ..., m-1$.

The valuation in (M, w), denoted by $(M, w) \models$, is defined over $\mathcal{F}$ by induction on the length of a formula as follows:

V0: for any $A \in PV$, $(M, w) \models A \Longleftrightarrow \tau(w, A) = \top$;

V1: $(M, w) \models \neg A \Longleftrightarrow (M, w) \not\models A$;

V2: $(M, w) \models A \supset B \Longleftrightarrow (M, w) \not\models A$ or $(M, w) \models B$;

V3: $(M, w) \models \wedge\Phi \Longleftrightarrow (M, w) \models A$ for all $A \in \Phi$;

V4: $(M, w) \models \vee\Phi \Longleftrightarrow (M, w) \models A$ for some $A \in \Phi$;

V5: $(M, w) \models \mathbf{B}_i(A) \Longleftrightarrow (M, v) \models A$ for all v with wR_iv;

V6: $(M, w) \models \mathbf{Ir}_i(A_1, ..., A_n) \Longleftrightarrow (M, v_{m+1}) \models A_{i_m}$ for all alternating chains $[(v_0, i_0), (v_1, i_1), ..., (v_m, i_m), v_{m+1}]$ with $(v_0, i_0) = (w, i)$.

Definitions V0-V5 are standard, and V6 corresponds exactly to (2). The semantic valuation for the common knowledge operator in the common knowledge logic CKL is much simpler: it requires $A_1 = ... = A_n = A$ and $(M, v_m) \models A$ for all chains $[v_0, v_1, ..., v_m, v_{m+1}]$ (cf., Fagin *et al.* [3]).

The basic completeness theorem is as follows:

Theorem 3.1. ***(Basic completeness)***. *Let A be any formula. Then, $\vdash A$ if and only if $(M, w) \models A$ for any (finite) serial models $M = ((W; R_1, ..., R_n), \tau)$ and any $w \in W$.*

This theorem will be proved in Section 6. The proof guarantees the *finite model property*, i.e., we can restrict the latter part of the theorem to the finite models. Another restriction is to require each model to be *connected*, i.e., for any $w, u \in W$, there is a chain from w to u. In Section 3.2, we give further restrictions, which will be used in Section 4.

It follows from Theorem 3.1 that a set Γ of formulae is consistent if and only if for any finite nonempty subset $\Gamma' \subseteq \Gamma$, there is a serial model M and a possible world w in M such that $(M, w) \models A$ for all $A \in \Gamma'$.

We cannot extend Theorem 3.1 to strong completeness: $\Gamma \vdash A$ if and only if $(M, w) \models A$ for any serial models M of Γ and $w \in W$. One counterexample for the "if" direction is to take the set in (2) as Γ and to take $\mathbf{Ir}_i(A_1, ..., A_n)$ as A with $n = 2$. By V6, any model of Γ would satisfy $(M, w) \models A$, but, in general, for any finite subset Γ' of Γ, $\wedge\Gamma' \vdash A$ is not provable in EIR^n.

The infinite-regress logic EIR^n is a *conservative extension* of KD^n, i.e., for any formula A with no occurrences of $\mathbf{Ir}_i(\cdot, ..., \cdot)$ for any $i \in N$, $\vdash A$ in $\mathrm{EIR}^n \Longleftrightarrow \vdash A$ in KD^n. As a result, we can convert meta-theorems in KD^n (e.g., epistemic depth theorem in Kaneko-Suzuki [11]) into EIR^n.

The proof of Theorem 3.1 given in Section 6 can be modified without difficulties to the logic

EIR^n with Axiom T, Axiom 4, and/or Axiom 5 : $\neg\mathbf{B}_i(A) \supset \mathbf{B}_i(\neg\mathbf{B}_i(A))$. In the corresponding cases, we need to add reflexivity, transitivity, and/or euclideanness on the accessibility relations R_i. Necessary modifications on the proof will be mentioned in Section 6.3. However, certain meta-theorems given in Section 4 fail in the logic with Axioms T, 4, or 5.

3.2 Completeness with respect to ep-rooted models

Theorem 3.1 can be modified to a more convenient form that will become very useful in Section 4. Let $(W; R_1, ..., R_n)$ be a Kripke frame. We use $W_i(w)$ to denote the set of all possible worlds u's that are *accessible from* w with the initial reference by player $i_0 = i$, that is, u's for which there is a chain $[(v_0, i_0), (v_1, i_1), ..., (v_m, i_m), v_{m+1}]$ $(m_0 \geq 0)$ with $v_0 = w, i_0 = i$ and $u = v_{m+1}$. Let $w_0 \in W$. We say that a frame $(W; R_1, ..., R_n)$ is *ep-rooted at* w_0 iff it satisfies

$$\{w_0\}, W_1(w_0), ..., W_n(w_0) \text{ are mutually disjoint.} \tag{7}$$

Note that if $(W; R_1, ..., R_n)$ is *ep-rooted at* w_0, then $w_0 \notin W_i(w_0)$ for all i. The union $\{w_0\} \cup W_1(w_0) \cup ... \cup W_n(w_0)$ consists of w_0 and the worlds accessible from w_0. It may be possible for w in this set to be accessible from some world $w \in W - \{w_0\} \cup W_1(w_0) \cup ... \cup W_n(w_0)$. For semantic evaluations at w_0, however, it is more important that w_0 is not accessible from any $w \in W_1(w_0) \cup ... \cup W_n(w_0)$ and each $W_i(w_0)$ is separated from $W_j(w_0)$ for $i \neq j$. This enables to evaluate the truthfulness of any formula at w_0 by referring only to $W_1(w_0) \cup ... \cup W_n(w_0)$.

An *ep-rooted frame* F *at* w_0 is denoted as $F = (W; w_0; R_1, ..., R_n)$. The set of ep-rooted serial frames is a proper subset of the set of serial frames. Nevertheless, we have the following theorem, which will be proved in the end of this section.

Theorem 3.2. ***(Completeness with respect to ep-rooted models)****. Let A be any formula. Then, $\vdash A$ if and only if $(M, w_0) \models A$ for any ep-rooted serial models $M = ((W, w_0; R_1, ..., R_n), \tau)$.*

Although the sematic requirement here is weaker than in Theorem 3.1, this theorem asserts that validity remains equivalent. Here, the semantic valuation is stated only at the root w_0 for each rooted model $M = ((W, w_0; R_1, ..., R_n), \tau)$. The other parts $W_1(w_0), ..., W_n(w_0)$ are needed to evaluate $\mathbf{B}_1(\cdot), ..., \mathbf{B}_n(\cdot)$ at w_0, and the remaining part $W - \{w_0\} \cup W_1(w_0) \cup ... \cup W_n(w_0)$ is not used at all. For simplicity, we still allow this remaining part to be nonempty.

Theorem 3.2 fails with Axiom T, because reflexivity required for R_i is violated by (7). With Axiom 4, Theorem 3.2 remains if transitivity required for each R_i in M is compatible with(7).

The *only-if* part of Theorem 3.2 directly follows from Theorem 3.2. We prove the *if* part. First, let us prove the following lemma.

Lemma 3.1. . *Suppose that $M = (F, \tau) = ((W; R_1, ..., R_n), \tau)$ is a serial model, and choose a fixed $w \in W$. Then, there is an ep-rooted serial model $M^* = ((W^*, w_0^*; R_1^*, ..., R_n^*), \tau^*)$ such that for any formula A,*

$$(M, w) \models A \text{ if and only if } (M^*, w_0^*) \models A. \tag{8}$$

Proof. Let w_0^* be a new symbol. We define an ep-rooted model

$$M^* = (F^*, \tau^*) = ((W^*, w_0^*; R_1^*, ..., R_n^*), \tau^*)$$

as follows: for $i \in N$,

$$W_i^* = \{(u,i) : u \in W\}, \quad (9)$$
$$R_i^* = \{(w_0^*, (u,i)) : (w,u) \in R_i\} \cup \{[(v,k),(u,k)] : (v,u) \in R_i, k \in N\},$$

and

$$W^* = \{w_0^*\} \cup W_1^* \cup ... \cup W_n^*, \quad (10)$$
$$\tau^*(w_0^*, p) = \tau(w,p),\ \tau^*((u,k),p) = \tau(u,p) \text{ for all } k \in N,\ p \in PV.$$

Each W_i^* is a copy of W in that the restriction of M^* to W_i^* is essentially the same as $((W; R_1, ..., R_n), \tau)$. That is, we have a natural bijection ψ_i between W and W_i^*. This bijection ψ_i preserves the structure of the accessibility relations $R_j, j \in N$ and also the truth assignment τ. This preservation includes the fact that for $i \neq j$, W_i^* and W_j^* are entirely separated and also the fact that w_0^* is not referred at all from W_i^*. This implies that for each $i \in N$ and each $u \in W$,

$$(M^*, (u,k)) \models A \text{ if and only if } (M,u) \models A. \quad (11)$$

On the other hand, at w_0^*, (u,i) is referred by R_i^* as u is referred at w by R_i. Thus, the new frame F^* is ep-rooted and serial.

Let us see (8). Consider each $i \in N$. The accessibility relation R_i^* at w_0 keeps the same structure of R_i at w, while (w,i) is used when w is referred. Also, τ^* in W_i^* preserves the same values as τ in W_i^*, and τ^* at w_0 coincides with τ at w. The equivalence (8) follows from these facts and (11).■

Proof of Theorem 3.2. We show the contrapositive of the *if* part. Suppose $\nvdash A$. Then, by Theorem 3.1, there is a KD^n-model $M = (F, \tau)$ such that $(M,w) \nvDash A$ for some $w \in W$. By Lemma 3.1, we have an ep-rooted model $M^* = ((W^*, w_0^*; R_1^*, ..., R_n^*), \tau^*)$ so that $(M^*, w_0^*) \nvDash A$. ■

4 Metatheorems for subjectivity

In this section we give few metatheorems that demonstrate the subjectivity of agents' thinking in our EIR^n.

4.1 A Scope Theorem

One critical feature of EIR^n is that each player enjoys the logical ability described by the classical logic, and their reasonings are purely subjective in the sense that they are independent of the objective situation. These features are captured by our next theorem, called the *scope theorem*, as it states that we may change the epistemic scope from the inside player to the outside analyst when making inferences in EIR^n. As shown below, such independence does not exit if we impose Axiom 4 or Axiom T, and hence the choice of KD^n is crucial to maintain these features. Our proof utilizes the convenience of ep-rooted models. Scope Theorem is also crucial to our proof of Theorem 2.1, which is given right after it.

Theorem 4.1. ***(Scope Theorem 1)**: Let $i \in N$.*

***(i)**: Let A, C be formulae. Then, the following statements are equivalent:*

(a) $\vdash A \supset C$; (b) $\vdash \mathbf{B}_i(A \supset C)$; and (c) $\vdash \mathbf{B}_i(A) \supset \mathbf{B}_i(C)$.

***(ii)**: Let A, C be formulae. Then, the following statements are equivalent:*

(a) $\vdash A \supset \neg C$; (b) $\vdash \mathbf{B}_i(A \supset \neg C)$; and (c) $\vdash \mathbf{B}_i(A) \supset \neg\mathbf{B}_i(C)$.

Proof. (i): It is easy to see $(a) \Longrightarrow (b) \Longrightarrow (c)$. Here, we show $(c) \Longrightarrow (a)$, and prove its contrapositive. Suppose $\nvdash A \supset C$. Then, by the completeness part of Theorem 3.2, there is an ep-rooted model $M = (F, \tau) = ((W, w_0; R_1, ..., R_2), \tau)$ such that $(M, w_0) \nvDash A \supset C$, i.e., $(M, w_0) \models A$ and $(M, w_0) \nvDash C$. We add a new element w_0^* to F as follows:

$$W^* = W \cup \{w_0^*\}; \text{ and } R_j^* = R_j \cup \{(w_0^*, w_0)\} \text{ for } j \in N.$$

This is a KD-frame (but not ep-rooted). Let τ^* be the assignment such that $\tau^*(w, \cdot) = \tau(w, \cdot)$ for all $w \in W$ and $\tau^*(w_0^*, \cdot) = \tau(w_0, \cdot)$. Let $M^* = ((W^*, R_1^*, ..., R_n^*), \tau^*)$. Then, $(M^*, w_0) \models A$ and $(M^*, w_0) \nvDash C$, which imply $(M^*, w_0^*) \models \mathbf{B}_i(A)$ and $(M^*, w_0^*) \nvDash \mathbf{B}_i(C)$. Hence, $(M^*, w_0^*) \nvDash \mathbf{B}_i(A) \supset \mathbf{B}_i(C)$. Thus, $\nvdash \mathbf{B}_i(A) \supset \mathbf{B}_i(C)$ by the soundness part of Theorem 3.1.

(ii): We show $(c) \Longrightarrow (a)$. Let $\vdash \mathbf{B}_i(A) \supset \neg\mathbf{B}_i(C)$. Then $\{\mathbf{B}_i(A), \mathbf{B}_i(C)\}$ is inconsistent, so $\vdash \mathbf{B}_i(A \wedge C) \supset \mathbf{B}_i(\neg p \wedge p)$. By (1) above, we have $\vdash A \wedge C \supset \neg p \wedge p$; so $\vdash A \supset \neg C$.■

Theorem 4.1 fails with Axiom 4 and/or with Axiom T. A counter example in $\mathrm{EIR}^2(4)$ is an instance of Axiom 4 itself: $\vdash \mathbf{B}_i(p) \supset \mathbf{B}_i\mathbf{B}_i(p)$ but $\nvdash p \supset \mathbf{B}_i(p)$. A counter example in $\mathrm{EIR}^2(\mathrm{T})$ is: $\vdash \mathbf{B}_i(\neg\mathbf{B}_i(p) \wedge p)) \supset \mathbf{B}_i(\neg p \wedge p)$ in $\mathrm{EIR}^2(\mathrm{T})$. Indeed, since $\vdash \mathbf{B}_i(\neg\mathbf{B}_i(p) \wedge p)) \supset \mathbf{B}_i(\neg\mathbf{B}_i(p)) \wedge \mathbf{B}_i(p)$ and then $\vdash \mathbf{B}_i(\neg\mathbf{B}_i(p) \wedge p)) \supset \neg\mathbf{B}_i(p) \wedge \mathbf{B}_i(p)$. Thus, $\mathbf{B}_i(\neg\mathbf{B}_i(p) \wedge p)$ is contradictory; and so $\vdash \mathbf{B}_i(\neg\mathbf{B}_i(p) \wedge p)) \supset \mathbf{B}_i(\neg p \wedge p)$. On the other hand, $\nvdash \neg\mathbf{B}_i(p) \wedge p \supset \neg p \wedge p$, which is obtained by constructing a counter model.

Proof of Theorem 2.1.(i): First, we see that Rule IRI_i is admissible in the logic $\mathrm{EIR}^{\mathrm{In},n}$. Suppose the upper formulae of IRI_i are proved, i.e., $\vdash D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(\wedge_{k \neq j} D_k)$ for all $j \in N$. Then, by Nec, we have $\vdash \mathbf{B}_i(D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(\wedge_{k \neq j} D_k))$ for all $j \neq i$. Hence, All the upper formulae of $\mathrm{IRI}_i^{\mathrm{In}}$ are provable. Hence, by $\mathrm{IRI}_i^{\mathrm{In}}$, we have $\vdash D_i \supset \mathbf{Ir}_i(A_1, ..., A_n)$. It is similar that Rule $\mathrm{IRI}_i^{\mathrm{In}}$ is admissible in the logic EIR^n: We use Theorem 4.1.(1) in the start.

(ii): Again, we show that IRI_i is admissible in $\mathrm{EIR}^{\mathrm{HK},2}$, and that $\mathrm{IRI}_i^{\mathrm{HK},2}$ is admissible in EIL^2. Suppose $\vdash D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i(D_j)$ and $\vdash D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(D_i)$. The latter implies $\vdash \mathbf{B}_i(D_j) \supset \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j(D_i)$.These imply $\vdash D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j(D_i)$. Hence, $\vdash D_i \supset \mathbf{Ir}_i(A_1, A_2)$ by $\mathrm{IRI}_i^{\mathrm{HK},2}$. Conversely, suppose $\vdash D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j(D_i)$. Then, let $D_j = \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(D_i)$. Then, $\vdash D_j \supset \mathbf{B}_j(A_j) \wedge \mathbf{B}_j(D_i)$ and also $\vdash D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i(D_j)$. Thus, by IRI_i, $\vdash D_j \supset \mathbf{Ir}_i(A_1, A_2)$.■

Theorem 4.1 can be extended to the $\mathbf{Ir}_i$ operator. However, it requires the contents of the operator have proper scopes, and we need an epistemic separation theorem for that purpose, which is introduced in the next subsection.

4.2 Epistemic Separations and Disjunction Properties

This section presents several results, which illustrate that each player's subjective inferences are independent from those of other players in EIR^n. This principle is formalized by our Separation

Theorem (Theorem 4.2). It also enables us to obtain the corresponding Scope Theorem for the infinite regress operators (Theorem 4.1). We also obtain a theorem on disjunction properties for the belief operator (Theorem 4.4). The first two theorems will be used in a game theoretic context in Section 5. Weaker forms of the first and the third theorems were given for the epistemic logic GL_{EF} of shallow depths in Kaneko-Suzuki [11].

4.2.1 Epistemic separation theorem

First, let $PV_0, PV_1, ..., PV_n$ be a partition of PV, where some PV_i's are allowed to be empty. For $i \in N$, we say that a formula A is *i-indecomposable* iff $A \in PV_i$ or the outermost symbol of A is $\mathbf{B}_i(\cdot)$ or $\mathbf{Ir}_i(\cdot, ..., \cdot)$, and that A is an *i-formula* iff it is generated only from i-indecomposable formulae with the four logical connectives. A nonepistemic formula including only propositional variables in PV_0 is called a *0-formula*. For example, $\mathbf{B}_i(\mathbf{B}_j(A) \supset p)$ is an i-formula, and so is $\mathbf{B}_i(\mathbf{B}_j(A) \supset p) \wedge p$ if $p \in PV_i$. Also, $\mathbf{Ir}_i(A_1, ..., A_n)$ itself is an i-formula. We note that these i-formulae for $i \in N$ and 0-formulae are relative to the given partition $(PV_0, PV_1, ..., PV_n)$ of PV. For simplicity, we let $N^* = \{0\} \cup N$; a 0-formula is also expressed as an i-formula for $i \in N^*$.

The Separation Theorem below shows that we can discuss each player's logical inferences independently of others' beliefs and inferences in EIR^n. In our game-theoretical applications, this theorem is crucial to allow for a consistent framework of subjective beliefs and inferences where players may hold completely different views about the game situation. We take into account the partition of the propositional variables into $(PV_0, PV_1, ..., PV_n)$, which allow more freedom for game theoretical applications than in [11].

Theorem 4.2. ***(Epistemic Separation for EIR^n):*** *Let Γ_i be a set of i-formulae and A_i an i-formula for $i \in N^*$. We let $\Gamma = \cup_{i \in N^*}\Gamma_i$. Then,*

(o): Γ is consistent if and only if Γ_i is consistent for all $i \in N^*$.

(i): Suppose that Γ is consistent. Then, $\Gamma \vdash \wedge_{i \in N^*} A_i$ if and only only if $\Gamma_i \vdash A_i$ for $i \in N^*$.

(ii): $\Gamma \vdash \vee_{i \in N^*} A_i$ if and only only if $\Gamma_i \vdash A_i$ for some $i \in N^*$.

Proof. In the following, for any set Λ of formulae, we abbreviate $(M, w) \models A$ for all $A \in \Lambda$ as $(M, w) \models \Lambda$. We stipulate $(M, w) \models \emptyset$.

(o): The *if* part is essential. Suppose the consistency of Γ_i for all $i \in N^*$. Let Γ' be a finite subset of Γ. We let $\Gamma'_i = \Gamma' \cap \Gamma_i$ for $i \in N^*$. For each $i \in N^*$, there is an ep-rooted model $M^i = (F^i, \tau^i) = ((W^i, w_0^i; R_1^i, ..., R_n^i), \tau^i)$ such that $(M^i, w_0^i) \models \Gamma'_i$. Recall that each formula in Γ'_i is an i-formula. For $i = 0$, only the valuations at the root w_0^0 is relevant; and for $i \in N$, only the valuations at the set $\{w_0^i\} \cup W_i^i(w_0^i)$ (recall $W_i^i(w_0^i) := \{w \in W^i : w_0^i R_i^i w\}$) are relevant. These facts enable us to combine the sets $W_i^i(w_0^i), i \in N$ with w_0^0 and construct the truth valuations appropriately.

We define $M^* = (F^*, \tau^*) = ((W^*, w_0^*; R_1^*, ..., R_n^*), \tau^*)$ with $w_0^* = w_0^0$ as follows:

$$W^* = \{w_0^*\} \cup W_1^1(w_0^1) \cup ... \cup W_n^n(w_0^n), \tag{12}$$

$$\begin{aligned} \text{for } i &\in N, \ R_i^* = \{(w_0^*, u) : (w_0^i, u) \in R_i^i\} \cup \\ \{(u, v) &: u \in W^*, u \neq w_0^i, \text{ and } (u, v) \in R_i^1 \cup ... \cup R_i^n\}; \end{aligned} \tag{13}$$

$$\begin{aligned} \tau^*(w_0^*, p) &= \tau^i(w_0^i, p) \text{ if } p \in PV_i, i \in N^*; \\ \tau^*(u, p) &= \tau^i(u, p) \text{ if } p \in PV,\ u \in W_i^i(w_0^i), i \in N^*. \end{aligned} \quad (14)$$

Thus, the new root, w_0^*, and the partial worlds, $W_1^1(w_1^1), ..., W_n^n(w_n^1)$, are connected keeping the original accessibility relations by (13). Condition (14) makes τ^* coincide with $\tau^i(w_0^i, p)$ if $p \in PV^i$ for $i \in N^*$ and also $\tau^*(w, \cdot)$ is the same $\tau^i(w, \cdot)$ in $W_i^i(w_0^i), i \in N$.

Now we show that $(M^*, w_0^*) \models \Gamma'$. First, by (14), for any $i \in N^*$ and $p \in PV^i$, $(M^*, w_0^*) \models p$ if and only if $(M^i, w_0^i) \models p$. Because $(M^0, w_0^0) \models \Gamma_0'$, it holds that $(M^*, w_0^*) \models \Gamma_0'$. Let $i \in N$ be fixed. Then, by an induction argument, it holds for any $w \in W_i^i(w_0^i)$ and any C, $(M^*, w) \models C \iff (M^i, w) \models C$. This implies $(M^*, w_0^*) \models \mathbf{B}_i(C) \iff (M^i, w_0^i) \models \mathbf{B}_i(C)$, and also $(M^*, w_0^*) \models \mathbf{Ir}_i(\mathbf{A}) \iff (M^i, w_0^i) \models \mathbf{Ir}_i(\mathbf{A})$. These facts imply that for $i \in N^*$, if $A \in \Gamma_i'$, then $(M^*, w_0^*) \models A$. That is, M^* is a model for Γ'. Because Γ' is an arbitrary finite subset of Γ, this shows that Γ is consistent.

(i): It suffices to show the *only-if* part. We show its contrapositive: $\Gamma_i \nvdash A_i$ for some $i \in N^*$ implies $\Gamma \nvdash A_0 \wedge A_1 \wedge ... \wedge A_n$, that is, for any finite subset $\Gamma' \subseteq \Gamma$, $\Gamma' \nvdash A_0 \wedge A_1 \wedge ... \wedge A_n$.

Suppose $\Gamma_i \nvdash A_i$ for some $i \in N^*$. Let $\Gamma' \subseteq \Gamma$ be an arbitrary finite subset. Since Γ is consistent, we have an ep-rooted model $M^* = (F^*, \tau^*) = ((W^*, w_0^*; R_1^*, ..., R_n^*), \tau^*)$ such that

$$(M^*, w_0^*) \models \Gamma'. \quad (15)$$

First, suppose that $\Gamma_0 \nvdash A_0$; hence $\Gamma_0' \nvdash A_0$, where $\Gamma_i' = \Gamma' \cap \Gamma_i$ for $i \in N^*$. By Theorem 3.2, we have another ep-rooted model $M^0 = (F^0, \tau^0) = ((W^0, w_0^0; R_1^0, ..., R_n^0), \tau^0)$ with root w_0 such that $(M^0, w_0^0) \models \Gamma_0'$ but $(M^0, w_0^0) \nvDash A_0$. We modify τ^* as follows: for all $w \in W^*$,

$$\tau^{*o}(w, p) = \begin{cases} \tau^0(w_0^0, p) & \text{if } w = w_0^* \text{ and } p \in PV_0 \\ \tau^*(w, p) & \text{otherwise.} \end{cases}$$

Then, τ^{*o} differs from τ^* only in w_0^* and in PV_0. Any formula in $\Gamma_0' \cup \{A_0\}$ has only propositional variables in PV_0 and is non-epistemic. Thus, $((F^*, \tau^{*o}), w_0^*) \models \Gamma_0'$ but $((F^*, \tau^{*o}), w_0^*) \nvDash A_0$. Since the change in w_0^* does not affect the other parts in M^*, we have $((F^*, \tau^{*o}), w_0^*) \models \Gamma_1' \cup ... \cup \Gamma_n'$. Hence, $\Gamma' \nvdash A_0 \wedge A_1 \wedge ... \wedge A_n$ by Theorem 3.2 (completeness part).

Now suppose that $\Gamma_i \nvdash A_i$ for some $i \in N$; hence $\Gamma_i' \nvdash A_i$. Then, we have an ep-rooted model $M^i = (F^i, \tau^i) = ((W^i, w_0^i; R_1^i, ..., R_n^i), \tau^i)$ such that $(M^i, w_0^i) \models \Gamma_i'$ and $(M^i, w_0^i) \nvDash A_i$. Here, we can assume $w_0^i = w_0^*$. Here we modify combine $M^* = (F^*, \tau^*)$ into $M^{*i} = (F^{*i}, \tau^{*i})$, so that only the part of $M^i = (F^i, \tau^i)$ relevant to player i is taken into $M^* = (F^*, \tau^*)$, as follows:

$$W^{i*} = \{w_0^*\} \cup \left[\cup_{j \in N^*, j \neq i} W_j^*(w_0^*)\right] \cup W_i^i(w_0^i);$$

$$\begin{aligned} R_k^{i*} &= [R_k^* \cap \left[\cup_{j \in N^*, j \neq i} \langle (W_j^*(w_0^*) \cup \{w_0^*\}) \times W_j^*(w_0^*) \rangle\right]] \\ &\quad \cup [R_k^i \cap [(W_i^i(w_0^i) \cup \{w_0^i\}) \times W_i^i(w_0^i)]] \text{ for all } k \text{ in } N^*; \end{aligned}$$

$$\tau^{i*}(w, p) = \begin{cases} \tau^i(w, p) & \text{if } w \in W_i^i(w_0^i) \cup \{w_0^*\} \text{ and } p \in PV_i \\ \tau^*(w, p) & \text{otherwise.} \end{cases}$$

Since $(M^i, w_0^i) \models \Gamma_i'$ by the choice of M^i and $(M^*, w_0^*) \models \Gamma_i'$ by (15), we have $(M^{*i}, w_0^*) \models \Gamma_i'$ and $(M^{*i}, w_0^*) \models \Gamma_j'$ for any $j \in N^*$ with $j \neq i$. Hence, $(M^{*i}, w_0^*) \models \Gamma'$. By $(M^i, w_0^i) \nvDash A_i$, we have $(M^{*i}, w_0^*) \nvDash A_i$. By this and Theorem 3.2, we have $\Gamma' \nvdash A_0 \wedge A_1 \wedge ... \wedge A_n$.

(ii): This can be proved in a similar manner to (i), without assuming the consistency of Γ.■

Theorem 4.2 states that the provability of a statement on the entire situation can be decomposed into each player's subjective perspective as well as the objective situation, and the inference in each component is independent of the other. Given this decomposition, we can focus on each player's provability and the objective situation separately. The choice of KD^n is crucial to obtain this separation result. Indeed, Theorem 4.2 fails with Axiom T, although it remains valid with Axiom 4. A counter example against (o) is as follows: Let $\Gamma_0 = \{p\}, \Gamma_1 = \{\mathbf{B}_1(p)\}, \Gamma_2 = \{\mathbf{B}_2(\neg p)\}$. Then $\Gamma_0 \cup \Gamma_1 \cup \Gamma_2$ is inconsistent in the presence of Axiom T, but each formula is consistent. This gives also a counterexample for (ii): $\Gamma_0 \cup \Gamma_1 \cup \Gamma_2 \vdash q \vee \mathbf{B}_1(q) \vee \mathbf{B}_2(\neg q)$ by Axiom T, but $p \nvdash q$, $\mathbf{B}_1(p) \nvdash \mathbf{B}_1(q)$ and $\mathbf{B}_2(\neg p) \nvdash \mathbf{B}_2(\neg q)$. An example against (i) is as follows: $\{p \supset q\} \cup \{\mathbf{B}_1(p)\} \cup \{\mathbf{B}_2(p)\} \vdash q$ holds under Axiom T but $p \supset q \nvdash q$.

We end this subsection with the Scope Theorem for the formula of the form $\mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{C})$, which also serves as an application of Theorem 4.2 because of the interactive nature of the infinite regress operator.

Theorem 4.3. ***(Scope Theorem 2)****: Let $i \in N$. Let A_k, C_k be a k-formula or a 0-formula for all $k \in N$. Suppose that $\mathbf{Ir}_i(\mathbf{A})$ is consistent. Then, the following statements are equivalent:*

(a) $\vdash (A_k \supset C_k)$ for each $k = 1, ..., n$; (b) $\vdash \mathbf{Ir}_i(\mathbf{A} \supset \mathbf{C})$; (c) $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{C})$.

Proof. It follows from Lemma 2.2.(4) that $(a) \Longrightarrow (b)$. By Lemma 2.2.(1), we have $(b) \Longrightarrow (c)$. Consider $(c) \Longrightarrow (a)$. Suppose $\vdash \mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{Ir}_i(\mathbf{C})$. By Theorem 2.2, this is equivalent to $\vdash \mathbf{B}_i[A_i \wedge [\wedge_{j \neq i} \mathbf{Ir}_j(\mathbf{A})]] \supset \mathbf{B}_i[C_i \wedge [\wedge_{j \neq i} \mathbf{Ir}_j(\mathbf{C})]]$. By (1) of Theorem 4.1.(1), we have $\vdash A_i \wedge [\wedge_{j \neq i} \mathbf{Ir}_j(\mathbf{A})] \supset C_i \wedge [\wedge_{j \neq i} \mathbf{Ir}_j(\mathbf{C})]$, i.e., $\{A_i\} \cup \{\mathbf{Ir}_j(\mathbf{A}) : j \neq i\} \vdash C_i \wedge [\wedge_{j \neq i} \mathbf{Ir}_j(\mathbf{C})]$. Since $\mathbf{Ir}_i(\mathbf{A})$ is consistent, $A_i \wedge [\wedge_{j \neq i} \mathbf{Ir}_j(\mathbf{A})]$ is also consistent by Axiom D. Then, since both A_i, C_i are either i-formulae or 0-formula, and since $\mathbf{Ir}_j(\mathbf{A}), \mathbf{Ir}_j(\mathbf{C})$ are j-formulae, we have, by Theorem 4.2.(1), $\vdash A_i \supset C_i$ and $\vdash \mathbf{Ir}_j(\mathbf{A}) \supset \mathbf{Ir}_j(\mathbf{C})$ for all $j \neq i$. Now we repeat the same argument, we have $\vdash A_j \supset C_j$ for all $j \in N$.■

4.2.2 Epistemic disjunction properties

In EIR^n, we have various disjunction properties. Here, we mention the following theorem presenting two results. The first is an extension of the theorem given for the epistemic logic GL_{EF} in Kaneko-Suzuki [11] to the logic IRL^n. The second is an extension to various players, which will be applied to game theoretic prediction/decision statement in Section 5. Here, let S be a nonempty subset of N, and we denote $\wedge_{j \in S} \mathbf{B}_j(C_j)$ by $\mathbf{B}_S(\mathbf{C})$.

Theorem 4.4. ***(Epistemic Disjunction Properties)****: (i): Let $i \in N$. Suppose that Λ^i and Θ^i are both finite nonempty sets of formulae. Then,*

$$\mathbf{B}_i(\Lambda^i) \vdash \vee \mathbf{B}_i(\Theta^i) \text{ if and only if } \mathbf{B}_i(\Lambda^i) \vdash \mathbf{B}_i(C) \text{ for some } C \in \Theta^i.$$

(ii): Let $\mathbf{\Phi}^S = \Pi_{j \in S} \mathbf{\Phi}^j$, where $\mathbf{\Phi}^j$ a finite nonempty set of formulae for $j \in S$. Then,

$$\mathbf{B}_S(\mathbf{A}) \vdash \vee_{\mathbf{C} \in \mathbf{\Phi}^S} \mathbf{B}_S(\mathbf{C}) \text{ if and only if } \mathbf{B}_S(\mathbf{A}) \vdash \mathbf{B}_S(\mathbf{C}) \text{ for some } \mathbf{C} \in \mathbf{\Phi}^S.$$

Proof. (i): The *only-if* part is essential, and we prove its contrapositive. Let $\Theta_i = \{C_1, ..., C_m\}$. Suppose $\mathbf{B}_i(\Lambda_i) \nvdash \mathbf{B}_i(C_t)$ for any $t = 1, ..., m$. By Theorem 3.2, for each $t = 1, ..., m$, we have an ep-rooted KD-models $M^t = ((W^t; w_0^t; R_1^t, ..., R_n^t), \tau^t)$ of $\mathbf{B}_i(\Lambda_i)$ with $(M^t, w_0^t) \nvDash \mathbf{B}_i(C_t)$. We can assume without loss of generality that $W^t \cap W^{t'} = \emptyset$ for distinct t, t'. Let w_0^* be a new symbol. Then, we replace each w_0^t by the common w_0^* : That is, we let $W^* = (\cup_{t=1}^m (W^t - \{w_0^t\})) \cup \{w_0^*\}$, and define, for $k \in N$,

$$R_k^* = \cup_{t=1}^m \left[\{(w_0^*, u) : (w_0^t, u) \in R_k^t\} \cup (R_k^t - \{(w_0^t, u) : (w_0^t, u) \in R_k^t\}) \right]. \tag{16}$$

Finally, for any $p \in PV$, we let $\tau^*(w, p) = \tau^t(w, p)$ if $w \in W^t - \{w_0^t\}$; and $\tau^*(w_0^*, p) = \tau^1(w_0^1, p)$. This $M^* = ((W^*, w^*; R_1^*, ..., R_n^*), \tau^*)$ is a KD-model. We show that $(M^*, w_0^*) \models \mathbf{B}_i(A)$ for any $A \in \Lambda_i$. Note that any $w \in W^* - \{w_0^*\}$ belongs to $W^t - \{w_0^t\}$ for some unique t. If $w_0^t R_i w$, then $w_0^* R_i^* w$, which implies that $(M^*, w) \models A$. Since this holds for any $w \in W^* - \{w_0^*\}$, we have $(M^*, w_0^*) \models \mathbf{B}_i(A)$. In sum, M^* is a model of Λ.

First, $(M^t, w_0^t) \nvDash \mathbf{B}_i(C_t)$ implies $(M^t, w) \nvDash C_t$ for some w with $w_0^t R_i w$. Since $(M^t, w) \nvDash C_t$ implies $(M^*, w) \nvDash C_t$ by (16), we have $(M^*, w) \nvDash C_t$ for some w with $w_0^* R_i^* w$. Hence, $(M^*, w_0^*) \nvDash \mathbf{B}_i(C_t)$. Since this holds for $t = 1, ..., m$, we have $(M^*, w_0^*) \nvDash \vee \mathbf{B}_i(\Phi)$. By soundness, we have $\mathbf{B}_i(\Gamma) \nvdash \vee \mathbf{B}_i(\Phi)$.

(ii): The *if* part is straightforward. We prove the *only-if* part. Let $i \in S$. Suppose $\mathbf{B}_S(\mathbf{A}) \vdash \vee_{\mathbf{C} \in \mathbf{\Phi}^S} [\wedge_{j \in S} \mathbf{B}_j(C_j)]$. Since $\vdash \vee_{\mathbf{C} \in \mathbf{\Phi}^S} [\wedge_{j \in S} \mathbf{B}_j(C_j)] \supset \vee_{C_i \in \mathbf{\Phi}^i} \mathbf{B}_i(C_i)$, we have $\mathbf{B}_S(\mathbf{A}) \vdash \vee_{C_i \in \mathbf{\Phi}^i} \mathbf{B}_i(C_i)$. Let $\top_j = \mathbf{B}_j(p_0) \vee (\neg \mathbf{B}_j(p_0))$ and $\top_j = \mathbf{B}_j(p_0) \wedge (\neg \mathbf{B}_j(p_0))$ for $j \in N^* - S$, where $\top_0 = p_0 \vee (\neg p_0)$ and $\top_0 = p_0 \wedge (\neg p_0)$. Then, by $\mathbf{B}_S(\mathbf{A}) \vdash \vee_{C_i \in \mathbf{\Phi}^i} \mathbf{B}_i(C_i)$, we have $\mathbf{B}_S(\mathbf{A}) \wedge (\wedge_{j \in N^* - S} \top_j) \vdash [\vee_{C_i \in \mathbf{\Phi}^i} \mathbf{B}_i(C_i)] \vee (\vee_{j \in N^* - S} \top_j)$. Applying Theorem 4.2.(2) to this, we have $\mathbf{B}_i(A_i) \vdash \vee_{C_i \in \mathbf{\Phi}^i} \mathbf{B}_i(C_i)$. By (1) of this theorem, we have $\mathbf{B}_i(A_i) \vdash \mathbf{B}_i(C_i)$ for some $C_i \in \Phi^i$. Since i is arbitrary in S, we have $\mathbf{B}_i(A_i) \vdash \mathbf{B}_i(C_i)$ for all $i \in S$. Hence, $\wedge_{j \in S} \mathbf{B}_j(A_j) \vdash \wedge_{j \in S} \mathbf{B}_j(C_j)$, i.e., $\mathbf{B}_S(\mathbf{A}) \vdash \mathbf{B}_S(\mathbf{C})$ and $\mathbf{C} \in \mathbf{\Phi}^S$.■

The above disjunction properties differ from those in intuitionistic logic, which require some sufficient conditions on the referred formulae to satisfy the properties. Those formulae are typically called Harrop formulae (cf., Troelstra-Schwichtenberg [19]). In contrast, Theorem 4.4 only requires the formulae to be within the scope of $\mathbf{B}_i(\cdot)$.

Theorem 4.4.(i) remains with Axiom 4 but fails with Axiom T or Axiom 5. First we give a counterexample under Axiom T. Let $\Gamma = \{\mathbf{B}_i(p) \vee \mathbf{B}_i(q)\}$. Then, $\mathbf{B}_i(\Gamma) \vdash \mathbf{B}_i(\mathbf{B}_i(p) \vee \mathbf{B}_i(q))$; so $\mathbf{B}_i(\Gamma) \vdash \mathbf{B}_i(p) \vee \mathbf{B}_i(q)$ by Axiom T. If Theorem 4.4.(i) holds, then $\mathbf{B}_i(\Gamma) \vdash \mathbf{B}_i(p)$ or $\mathbf{B}_i(\Gamma) \vdash \mathbf{B}_i(q)$. We can eliminate the outer $\mathbf{B}_i(\cdot)$ from those two provability statements, which preserves provability, and then we have $p \vee q \vdash p$ or $p \vee q \vdash q$ in classical logic, which is impossible. Now we consider Axiom 5. By 5, we have $\vdash \mathbf{B}_i(p) \vee \mathbf{B}_i(\neg \mathbf{B}_i(p))$. If Theorem 4.4.(i) holds, then $\vdash \mathbf{B}_i(p)$ or $\vdash \mathbf{B}_i(\neg \mathbf{B}_i(p))$, but either is impossible.

5 Game Theoretic Applications

As mentioned in Section 1, the logic EIR^n was motivated by prediction/decision making by an individual player facing a game situation. Here, we return to this game theoretic problem and discuss how we analyze the problem in EIR^n. In particular, using the meta-theorems given in the previous sections, we can go back and forth from inferences about the entire situation to the decomposed individual inferences. Although we consider the general n-person case here, we refer to Hu-Kaneko [5] for the basic game theoretic results in the 2-person case whenever those results apply to the more general case in a straightforward manner.

5.1 Basic game theoretic concepts

We consider an n-person strategic game $G = (N, \{S_i\}_{i \in N}, \{h_i\}_{i \in N})$, where S_i is a finite nonempty set of *strategies* (actions) and $h_i : S := S_1 \times ... \times S_n \to \mathbf{R}$ is a (real-valued) *payoff function* for player $i \in N$. We write $h_i(s_i; s_{-i})$ for $h_i(s)$ when we focus on player i. We say that $s_i \in S_i$ is a *best response* to $s_{-i} \in S_{-i}$ iff $h_i(s_i; s_{-i}) \geq h_i(t_i; s_{-i})$ for all $t_i \in S_i$. A profile of strategies $s = (s_1, ..., s_n) \in S$ is a *Nash equilibrium* iff s_i is a best response to s_{-i} for each $i \in N$. For each player i, a strategy s_i is a *Nash strategy* iff $(s_i; s_{-i})$ is a Nash equilibrium for some s_{-i}. Table 1.1 is a 2-person game with 3 strategies for each players, and has a unique Nash equilibrium, designated by the superscript NE, and Table 5.2 has two NE's. We use $E(G)$ to denote the set of Nash equilibria in G.[2]

Table 5.1

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	$(2,2)$	$(2,4)$	$(4,0)$
$\mathbf{s}_{12}$	$(4,2)$	$(3,3)^{NE}$	$(3,0)$
$\mathbf{s}_{13}$	$(5,5)$	$(0,0)$	$(2,6)$

Table 5.2

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	$(2,1)^{NE}$	$(0,0)$
$\mathbf{s}_{12}$	$(0,0)$	$(1,2)^{NE}$

In a game such as Tables 5.1 and 5.2, each player's payoff is interdependent in the sene that they depend not only upon his own choice but also on other player' choice. Thus, he makes a possible decision together with a prediction about the other's possible decisions. This prediction/decision making is described as the two statements:

Na_1: 1 chooses his best strategy against all of his predictions about 2's choice based on Na_2;

Na_2: 2 chooses his best strategy against all of his predictions about 1's choice based on Na_1.

A possible final decision for 1 is determined by Na_1, but because Na_2 is included in Na_1 as his prediction criterion, 1 also needs to assume that 2 uses Na_2 for his decision making. The symmetric form Na_2 determines a decision for player 2 with predictions about 1's decisions.

Player 1's decision making is described by his belief $\mathbf{B}_1(\mathrm{Na}_1)$ in Diagram 5.1, and in his prediction making, 1 thinks about 2's thinking by putting himself to 2' shoe, which is expressed as $\mathbf{B}_1\mathbf{B}_2(\mathrm{Na}_2)$ in Diagram 5.1. In fact, Na_1 occurs again in this $\mathbf{B}_1\mathbf{B}_2(\mathrm{Na}_2)$, which requires the third $\mathbf{B}_1\mathbf{B}_2\mathbf{B}_1(\mathrm{Na}_1)$, and so on. Taking Na_i as A_i, the infinite sequence forms the set (1), which is captured by the formula $\mathbf{Ir}_1(A_1, A_2)$ in the logic EIR^2. If we make no distinction between decisions and predictions, Diagram 5.1 collapses to $\mathrm{Na}_1 \longleftrightarrow \mathrm{Na}_2$: Na_1-Na_2 is a circular definition of decisions and predictions, i.e., as an equilibrium. This is Nash's [16] noncooperative theory. The logic EIR^n facilitates a study of his theory in a manner faithfully to prediction/decision making.

Diagram 5.1

$$\mathbf{B}_1(\mathrm{Na}_1) \quad \to \quad \mathbf{B}_1\mathbf{B}_2(\mathrm{Na}_2) \quad \to \quad \mathbf{B}_1\mathbf{B}_2\mathbf{B}_1(\mathrm{Na}_1) \quad \to \quad \mathbf{B}_1\mathbf{B}_2\mathbf{B}_1\mathbf{B}_2(\mathrm{Na}_2) \quad \to \quad \cdots$$

The concept of Nash equilibrium allows different interpretations. We follow the *ex ante* approach to study decision making, due to Nash [16], in Section 5.2, and we also consider it from the *ex post* perspective in Section 5.3. In the *ex ante* approach, from player i's subjective viewpoint, in a profile $(s_i; s_{-i})$, s_i is interpreted as a possible final decision for i and s_{-i} is

[2] Here we only focus on Nash equilibria in pure strategies, and hence the set $E(G)$ may be empty.

interpreted as his prediction of possible decisions for the others. Following Nash [16], we say that G is *solvable* iff $E(G) = E_1(G) \times ... \times E_n(G)$, where $E_i(G) = \{s_i \in S_i : (s_i; s_{-i}) \in E(G)$ for some $s_j \in S_j\}$ for each $i \in N$. This condition captures independent decision-making by the players, and we shall explain this formally later.

To express the game theoretical concepts in the language of EIR^n, we adopt the following atomic formulae for the propositional variables PV : For $i \in N$,

$2n$-ary preferences; $R_i(s;t)$ for $s, t \in S$; unary decision symbol; $\mathrm{I}_i(s_i)$ for $s_i \in S_i$;

ex post observation symbol; $\pi_i(s)$ for $s \in S$.

We denote the set of those atomic formulae for $i \in N$ by AF. Preference expressions $R_i(s;t)$ are used to express the payoff functions h_i. Decision expressions $\mathrm{I}_i(s_i)$ are intended to mean that s_i is a possible final decision for player i. *Ex post* observations $\pi_i(s)$ expresses the actions that are observed by player i after the actual play of the game.

To define i-formulae, we choose a partition $(AF_0, AF_1, ..., AF_n)$ of AF as follows:

$$\begin{aligned} AF_0 \ &: \ = \text{the set of all atomic preference and decision expressions;} \\ AF_i \ &: \ = \{\pi_i(s) : s \in S\} \text{ for } i \in N. \end{aligned} \tag{17}$$

With this partition, we can talk about the subjective understanding of the game situation for each player as well as his own prediction/decision criterion.

Let the objective game be $G^0 = (N, \{S_j\}_{j \in N}, \{h_j^0\}_{j \in N})$. Each $i \in N$ has a subjective perception about the game being played, and we use $G^i = (N, \{S_j\}_{j \in N}, \{h_j^i\}_{j \in N})$ to denote his perceived game. We allow h_j^i to be totally different from the objective payoff function h_j^0 for $j \in N$. In our formal language, for $i \in \{0\} \cup N$, the payoff function h_j^i for player $j \in N$ in G^i is expressed in terms of preferences:

$$g_j^i := \wedge[\{R_j(s;t) : h_j^i(s) \geq h_j^i(t), s, t \in S\} \cup \{\neg R_j(s;t) : h_j^i(s) < h_j^i(t), s, t \in S\}].$$

Here, N and $\{S_j\}_{j \in N}$ are assumed to be common, the game is identified by the profile of formulae $\boldsymbol{g}^i := (g_1^i, ..., g_n^i)$ and $\boldsymbol{g}^0 := (g_1^0, ..., g_n^0)$.

We express "s_i is a best response to s_{-i}" as the formula $\mathrm{bst}_i(s_i; s_{-i}) := \wedge\{R_i(s_i; s_{-i} : t_i; s_{-i}) : t_i \in S_i\}$. We may write $\mathrm{bst}_i(s) = \mathrm{bst}_i(s_i; s_{-i})$. We also express "$s$ is a Nash equilibrium" by the formula $\mathrm{Nash}(s) := \wedge_{j \in N} \mathrm{bst}_j(s_j; s_{-j})$. Note that those concepts are formulated without referring to a specific game.

5.2 *Ex ante* prediction/decision making

In Section 5, we described the prediction/decision making criterion Na_1 and Na_2 by player 1 in a 2-person game in a non-formalized manner. Here, we give its n-person version in the formalized language. Following Hu-Kaneko [5], we also introduce two auxiliary axioms. These axioms are intended to be the contents of player i's basic beliefs and hence they occur in player i's mind, i.e., in the scope of $\mathbf{B}_i(\cdot)$ (actually, $\mathbf{Ir}_i(\cdots)$);

N0$_i$ (Optimization against all predictions): $\wedge_{s \in S}[\mathrm{I}_i(s_i) \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathrm{I}_j(s_j))) \supset \mathrm{bst}_i(s_i; s_{-i})]$.

N1$_i$ (Predictions by the others): $\wedge_{s \in S}[\mathrm{I}_i(s_i) \wedge (\wedge_{j \neq i} \mathbf{B}_j(\mathrm{I}_j(s_j))) \supset \wedge_{j \neq i} \mathbf{B}_j(\wedge_{l \neq j} \mathbf{B}_l(\mathrm{I}_l(s_l)))]$.

N2$_i$ (Necessity of predictions): $\wedge_{s_i \in S_i}[\mathrm{I}_i(s_i) \supset \vee_{s_{-i} \in S_{-i}}(\wedge_{j \neq i}\mathbf{B}_j(\mathrm{I}_j(s_j)))]$.

For each $i \in N$, let $\mathrm{N}_i = \mathrm{N0}_i \wedge \mathrm{N1}_i \wedge \mathrm{N2}_i$, and let $\mathbf{N} = (\mathrm{N}_1, ..., \mathrm{N}_n)$.

Axiom $\mathrm{N0}_i$ translates Na_i into our formal language, taking the belief operators into account. The premises of Axioms $\mathrm{N0}_i$ and $\mathrm{N1}_i$ are the same; $\mathrm{I}_i(s_i)$ is a possible final decision for i, and $\wedge_{j \neq i}\mathbf{B}_j(\mathrm{I}_j(s_j))$ his predictions. Axiom $\mathrm{N1}_i$ requires, in the mind of player i, each $j \neq i$ makes predictions in the same manner. Axiom $\mathrm{N2}_i$ requires predictions for any possible final decision, and is needed in order to separate $\mathrm{I}_i(s_i)$ from his predictions $\wedge_{j \neq i}\mathbf{B}_j(\mathrm{I}_j(s_j))$.

We assume the infinite regress $\mathbf{Ir}_i(\mathbf{N}) = \mathbf{Ir}_i(\mathrm{N}_1, ..., \mathrm{N}_n)$ of those axioms, which corresponds to Diagram 5.1 (adding $\mathrm{N1}_i \wedge \mathrm{N2}_i, i \in N$). We take the infinite regress $\mathbf{Ir}_i(\mathbf{N})$ as basic beliefs for player i's prediction/decision making.

By Theorem 2.2, the epistemic content of $\mathbf{Ir}_i(\mathbf{A})$ is $A_i \wedge (\wedge_{j \neq i}\mathbf{Ir}_j(\mathbf{A}))$, which is denoted as $\mathbf{Ir}_i^o(\mathbf{A}) := A_i \wedge (\wedge_{j \neq i}\mathbf{Ir}_j(\mathbf{A}))$. Using the expression $\mathbf{Ir}_i^o$, we can write a candidate for $\mathrm{I}_i(s_i)$ as follows: for each $s_i \in S_i$,

$$A_i^*(s_i) := \vee_{s_{-i} \in S_{-i}}\mathbf{Ir}_i^o[\mathrm{bst}_1(s_1; s_{-1}), ..., \mathrm{bst}_n(s_n; s_{-n})]. \tag{18}$$

In the logic $\mathrm{EIR}^n(\mathrm{T})$, Theorem 2.3 implies that $A_i^*(s_i)$ can be written as $\vee_{s_{-i} \in S_{-i}}\mathbf{C}(\mathrm{nash}(s_i; s_{-i}))$. However, we are interested in the case without Axiom T; the reason will be manifested presently.

First, we have the following result: For $i \in N$,

$$\mathbf{Ir}_i(\mathbf{N}) \vdash \mathbf{B}_i(\mathrm{I}_i(s_i) \supset A_i^*(s_i)) \text{ for all } s_i \in S_i. \tag{19}$$

The 2-person version of (19) is given and proved in Hu-Kaneko [5]. Here, since it is slightly different and more general, a proof is given in the end of this section.

The result (19) gives a necessary condition for a possible final decision $\mathrm{I}_i(s_i)$. To have a full characterization, we add the infinite regress of payoff functions $\mathbf{Ir}_i(\mathbf{g}^i) := \mathbf{Ir}_i(g_1^i, ..., g_n^i)$. In fact, this is not enough: In classical mathematics, it is a standard practice to regard a given property as the largest set satisfying the property. The counterpart of this requirement in logic is to look for the deductively weakest formula. In our case, we consider n families of formulae as candidates for $\{\mathrm{I}_1(s_1)\}_{s_1 \in S_1}, ..., \{\mathrm{I}_n(s_n)\}_{s_n \in S_n}$. That is, we consider $\mathcal{A} = (\mathcal{A}_1, ..., \mathcal{A}_n) = (\{A_1(s_1)\}_{s_1 \in S_1}, ..., \{A_n(s_n)\}_{s_n \in S_n})$ so that we substitute $A_i(s_i)$ for $\mathrm{I}_i(s_i)$ ($s_i \in S_i, i \in N$) in the above axioms N0-N2$_i$. We call $\mathcal{A} = (\mathcal{A}_1, ..., \mathcal{A}_n)$ a profile of *candidate families* iff $\mathcal{A}_i$ is a family of formulae indexed by $s_i \in S_i$. The resulting axioms are written as $\mathrm{N0}_i(\mathcal{A})$,$\mathrm{N1}_i(\mathcal{A})$,$\mathrm{N2}_i(\mathcal{A})$, and we let $\mathrm{N}_i(\mathcal{A}) = \mathrm{N0}_i(\mathcal{A}) \wedge \mathrm{N1}_i(\mathcal{A}) \wedge \mathrm{N2}_i(\mathcal{A})$.

We formalize the choice of the deductive weakest families by the following axiom $\mathrm{WF}_i(\mathcal{A})$:

$$\begin{aligned} &\mathrm{N}_i(\mathcal{A}) \wedge (\wedge_{j \neq i}\mathbf{B}_j(\mathrm{N}_j(\mathcal{A}))) \wedge [\wedge_{s \in S}\langle \mathrm{I}_i(s_i) \wedge (\wedge_{j \neq i}\mathbf{B}_j(\mathrm{I}_i(s_i)) \supset A_i(s_i) \wedge [\wedge_{j \neq i}\mathbf{B}_j(A_i(s_i))\rangle] \\ \supset\ &\wedge_{s_i \in S_i}\langle A_i(s_i) \supset \mathrm{I}_i(s_i)\rangle. \end{aligned} \tag{20}$$

We use $\mathbf{Ir}_i(\mathbf{WF})$ to denote the set of all formulae of the form $\mathbf{Ir}_i(\mathrm{WF}_1(\mathcal{A}), ..., \mathrm{WF}_n(\mathcal{A}))$ for candidate families $\mathcal{A}$.

Then the basic beliefs for player i are given as:

$$\Delta_i(\mathbf{g}^i) := \{\mathbf{Ir}_i(\mathbf{g}^i), \mathbf{Ir}_i(\mathbf{N})\} \cup \mathbf{Ir}_i(\mathbf{WF}).$$

The following theorem summarizes individual inferences involved in this decision making process given in Hu-Kaneko [5] for the 2-person case. Later on we examine these inferences from the perspective of our Scope Theorems and Separation Theorem. The proofs of the claims of Theorem 5.1 can be proved in, more or less, the same manner as in [5] and we skip them.

Theorem 5.1. *Let $i \in N$. Let G^i be a game belied by player i, and $\mathbf{g}^i$ its formalized payoffs.*

(o) *$\Delta_i(\mathbf{g}^i)$ is consistent for each $i = 1, ..., n$, and $\wedge_{j \in N}\mathbf{g}_j^0$ is consistent.*

(i) *Let G^i be a solvable game. Then,*

$$\Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i(I_i(s_i) \equiv A_i^*(s_i)) \text{ for all } s_i \in S_i, \tag{21}$$

$$\Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i\mathbf{B}_j(I_j(s_j) \equiv A_j^*(s_j)) \text{ for all } s_j \in S_j \text{ and } j \neq i, \tag{22}$$

$$\text{for any } s_i \in S_i, \text{ either } \Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i(I_i(s_i)) \text{ or } \Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i(\neg I_i(s_i)), \tag{23}$$

$$\text{for any } s_{-i} \in S_{-i}, \text{ either } \Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i[\wedge_{j\neq i}\mathbf{B}_j(I_j(s_j))] \text{ or } \Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i[\vee_{j\neq i}\mathbf{B}_j(\neg I_j(s_j))]. \tag{24}$$

(ii) *Let G^i be a unsolvable game. Assume that it is generic in the sense that for all strategy profiles entail distinct payoffs for all players. Then, for all Nash strategy $s_i \in S_i$,*

$$\Delta_i(\mathbf{g}^i) \nvdash \mathbf{B}_i(I_i(s_i)) \text{ and } \Delta_i(\mathbf{g}^i) \nvdash \mathbf{B}_i(\neg I_i(s_i)). \tag{25}$$

Claim (o) states that the individual belief set $\Delta_i(\mathbf{g}^i)$ is consistent (contradiction-free) in EIR^n; game theoretical studies with $\Delta_i(\mathbf{g}^i)$ could be meaningless without this consistency. Claim (i) states that when the game G^i believed by i is solvable, the decision $\mathrm{I}_i(s_i)$ is characterized by the formula $A_i^*(s_i)$ given by (18), and also each strategy is decidable as a possible final decision or not in the mind of player i. In fact, $\Delta_i(\mathbf{g}^i) \vdash A_i^*(s_i)$ if s_i is a Nash strategy and $\Delta_i(\mathbf{g}^i) \vdash \neg A_i^*(s_i)$ if s_i otherwise (see Hu-Kaneko [5]). Claim (ii), however, states when G^i is unsolvable and is generic, player i cannot decide for any Nash strategy, whether it is a possible final decision or not. In fact, since (19) implies that $\Delta_i(\mathbf{g}^i)$ entails any non-Nash strategy as a negative decision, he cannot reach a positive decision for any strategy[3].

Hu-Kaneko [5] concentrated on the game theoretic decidability/undecidability from the perspective of an individual player. Our various meta-theorems in Section 4 allows us to the entire situation via epistemic separation and, here, we analyze the relationships between different individualistic views and the objective reality. The entire situation may be described by

$$\Delta(\mathbf{g}) := (\wedge_{j\in N}\mathbf{g}_j^0) \cup \Delta_1(\mathbf{g}^1) \cup .... \cup \Delta_n(\mathbf{g}^n).$$

The set $\Delta(\mathbf{g})$ includes all players' basic beliefs, as well as the objective description of the game. Theorem 5.1.(o) shows that each $\Delta_i(\mathbf{g}^i)$, as well as $\wedge_{j\in N}\mathbf{g}_j^0$, is consistent. Then, applying Theorem 4.2.(o), we have:

Step 0 (Consistency): The union of beliefs $\Delta(\mathbf{g})$ is consistent in the logic EIR^n.

Note that we allow $\mathbf{g}^i \neq \mathbf{g}^j$ for distinct i and j (and, perhaps, 0). Thus, each player may have completely different beliefs. However, as long as the differences are not revealed and only exist in their minds, the entire set $\Delta(\mathbf{g})$ is consistent. This would be inconsistent in the logic $\mathrm{EIR}^n(\mathrm{T})$ with Axiom T, which requires $\mathbf{g}^i = \mathbf{g}^0$ for all players. Instead of imposing Axiom T, we study such (in-)consistency from the *ex post* point of view.

Taking $\Delta(\mathbf{g})$ as given, we may ask what are the decisions made by players from the analyst's perspective. When a player can make a decision on a strategy s_i, the statement $\mathbf{B}_i(\mathrm{I}_i(s_i))$ or $\mathbf{B}_i(\neg\mathrm{I}_i(s_i))$ is provable. Let $\mathrm{I}_i^*(s_i)$ be $\mathrm{I}_i(s_i)$ or $\neg\mathrm{I}_i(s_i)$ for $i \in N$. Then, by Theorem 4.2.(i), we can decompose this collective decision problem into individual ones:

[3] For unsolvable games with nongeneric payoffs, Hu-Kaneko [5] gives a full characterization of strategies satisfying (25).

Step 1 (Decomposition): $\Delta(\mathbf{g}) \vdash \wedge_{i\in N}\mathbf{B}_i(\mathrm{I}_i^*(s_i))$ if and only if $\Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i(\mathrm{I}_i^*(s_i))$ for all $i \in N$.

This guarantees that an analysis of n players' decidabilities together is decomposed to a study of each player's, and *vice versa.* From Theorem 5.1.(i) and (ii), for each player i, whether he can reach a possible final decision or not only depend on whether G^i is solvable (and with a Nash equilibrium) or not, independent of G^j and G^0 for any $j \neq i$. Thus, by Theorem 5.1.(ii), if one of the perceived game G^i is unsolvable, the the whole situation is undecidable. For the remaining two steps we focus on the decidable case and discuss how to relate subjective inference to objective inference in the individualistic decision problem.

Focusing on individual inferences, (21) in Theorem 5.1.(i) shows that possible final decisions are fully characterized by the candidate formulae $A_i^*(s_i)$. To discuss derivability of $\mathbf{B}_i(A_i^*(s_i))$, we only need to refer to $\mathbf{Ir}_i(\mathbf{g}^i)$ in player i's basic beliefs. By Theorem 4.3, we have the following step:

Step 2 (Epistemic Reduction): $\mathbf{Ir}_i(\mathbf{g}^i) \vdash \mathbf{B}_i(A_i^*(s_i))$ if and only if $\mathbf{Ir}_i^o(\mathbf{g}^i) \vdash A_i^*(s_i)$.

Notice that the game $\mathbf{g}^i$ believed by player i appears in $\wedge_{j\neq i}\mathbf{Ir}_j(\mathbf{g}^i)$, that is, player i believes that player j believes $\mathbf{g}^i$ is being played instead of $\mathbf{g}^j$; player j may hold a complete different belief from the outside analyst's perspective.

For i's predictions, (22) shows that his predictions about j's decisions are fully characterized by the candidate formulae $A_j^*(s_j)$. As in the previous step, to discuss derivability of $\mathbf{B}_i\mathbf{B}_j(A_j^*(s_j))$, we only need to refer to $\mathbf{Ir}_i(\mathbf{g}^i)$ in player i's basic beliefs. By applying Theorem 4.3 twice we have the following step:

Step 3 (Epistemic Reduction for Predictions): $\mathbf{Ir}_i(\mathbf{g}^i) \vdash \mathbf{B}_i\mathbf{B}_j(A_j^*(s_j))$ if and only if $\mathbf{Ir}_j^o(\mathbf{g}^i) \vdash A_j^*(s_j)$.

In Step 3, the prediction problem for player i is reduced to his simulated inference for player j. However, note that in his simulation, i assumes j also believes that the game being played is the same as i's perception. In this sense, although inferences *per se* are purely objective and described by classical logic, players differ in their basic beliefs, i.e., the starting points of their inferences.

Proof of (19). Let $i \in N$, and $s \in S$ be fixed. First, we have the first claim, which corresponds to the content version of Rule IRI_i;

(0): for any formulae $D_k, k \in N$ and $A = (A_1, ..., A_n)$, if $\vdash D_k \supset A_k \wedge (\wedge_{j\neq k}\mathbf{B}_j(D_j))$ for all $k \in N$, then $\vdash D_i \supset \mathbf{Ir}_i(\mathbf{A})$.

This can be proved without much difficulty.

Now, the following step (1) is crucial and different from the corresponding step in [5].

(1): $\mathbf{Ir}_i^o[\mathrm{N01}_1, ..., \mathrm{N01}_n] \vdash \mathrm{I}_i(s_i) \wedge (\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(s_j))) \supset \mathbf{Ir}_i^o[\mathrm{bst}_1(s_1; s_{-1}), ..., \mathrm{bst}_n(s_n; s_{-n})]$;

(2): $\mathbf{Ir}_i^o[\mathrm{N}_1, ..., \mathrm{N}_n] \vdash \mathrm{I}_i(s_i) \supset A^*(s_i)$.

To prove (1), we let $\theta_i(s) := \mathbf{Ir}_i^o[\mathrm{N01}_1, ..., \mathrm{N01}_n] \wedge [\mathrm{I}_i(s_i) \wedge (\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(s_j)))]$. We show that

$$\vdash \theta_i(s) \supset \mathrm{bst}_i(s_i; s_{-i}) \wedge (\wedge_{j\neq i}\mathbf{B}_j(\theta_j(s))). \tag{26}$$

Once this is proved, we have, by (0), $\vdash \theta_i(s) \supset \mathbf{Ir}_i^o[\mathrm{bst}_1(s_1; s_{-1}), ..., \mathrm{bst}_n(s_n; s_{-n})]$. The first part of (26), $\vdash \theta_i(s) \supset \mathrm{bst}_i(s_i; s_{-i})$, comes from $\mathrm{N0}_i$ and $\mathrm{I}_i(s_i) \wedge (\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(s_j)))$. Consider the second part. Let $j \neq i$. By Theorem 2.2, we have $\vdash \mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n] \supset \mathrm{N1}_i$. Hence, $\vdash \mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n] \wedge \mathrm{I}_i(s_i) \wedge (\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(s_j))) \supset \wedge_{j\neq i}\mathbf{B}_j(\wedge_{l\neq j}\mathbf{B}_l(\mathrm{I}_l(s_l)))$. Hence,

$$\vdash \theta_i(s) \supset \wedge_{j\neq i}\mathbf{B}_j[\mathbf{B}_i(\mathrm{I}_i(s_i)) \wedge \mathrm{I}_j(s_j))].$$

This together with $\vdash \mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n] \supset \wedge_{j\neq i}\mathbf{B}_j(\mathbf{Ir}_j^o[\mathrm{N01}_1,...,\mathrm{N01}_n])$ implies $\vdash \theta_i(s_i;s_{-i}) \supset \wedge_{j\neq i}\mathbf{B}_j(\theta_j(s_j;s_{-j}))$.

Now we consider (2). It follows from (1) that

$$\mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n] \vdash \mathrm{I}_i(t_i) \wedge (\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(s_j))) \supset \vee_{t_{-i}\in S_{-i}}\mathbf{Ir}_i^o[\mathrm{bst}_1(t_1;t_{-1}),...,\mathrm{bst}_n(t_n;t_{-n})].$$

This is equivalent to $\mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n] \vdash (\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(s_j))) \supset (\mathrm{I}_i(t_i) \supset A_i^*(t_i))$. Hence $\mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n] \vdash \vee_{t_{-i}\in S_{-i}}(\wedge_{j\neq i}\mathbf{B}_j(\mathrm{I}_j(t_j))) \supset (\mathrm{I}_i(s_i) \supset A_i^*(s_i))$. By adding $\mathrm{N2}_i$ to $\mathbf{Ir}_i^o[\mathrm{N01}_1,...,\mathrm{N01}_n]$, we delete the first disjunctive formula, i.e., $\mathbf{Ir}_i^o[\mathrm{N}_1,...,\mathrm{N}_n] \vdash \mathrm{I}_i(s_i) \supset A_i^*(s_i)$.■

5.3 Consistency between *ex ante* predictions and *ex post* observations

In our analysis of the *ex ante* decision/prediction making, all interactions occur in each player's mind but not among the players in the physical world. From this *ex ante* viewpoint, we have used Separation Theorem to study each player's subjective decision/prediction making. Here we introduce external interactions through actual plays and *ex post* observations. Once the game is played and the actions taken are observed by the players, we can study the potential conflicts arising from a player's *ex ante* predictions about the others and his *ex post* observations.

To connect *ex ante* predictions to *ex post* observations, we impose the following axiom:

$$\mathrm{EX}_i := \wedge\{\pi_i(s) \supset \mathbf{B}_i(\mathrm{I}_i(s_i)) \wedge [\wedge_{j\neq i}\mathbf{B}_i\mathbf{B}_j(\mathrm{I}_j(s_j))] : s \in S\}. \tag{27}$$

Recall that $\pi_i(s)$ means that player i observes s as the actual play. EX_i states that if a profile $s = (s_i; s_{-i})$ is observed *ex post*, then s_i becomes a possible final decision for i and s_{-i} becomes his predictions of others' possible decisions.

Now, assuming that a particular outcome, $s^o \in S$, is observed (by all players), and, for each player $i \in N$, we consider the set

$$\hat{\Gamma}_i := \Delta_i(\mathbf{g}^i) \cup \mathrm{EX}_i \cup \{\pi_i(s^o)\}. \tag{28}$$

In this case, the consistency of $\hat{\Gamma}_i$ may be problematic. We have the following theorem.

Theorem 5.2. ***(*Ex Post *Consistencies)*** *(i):* $\mathbf{g}^o \cup \hat{\Gamma}_1 \cup ... \cup \hat{\Gamma}_n$ *is consistent if and only if* $\hat{\Gamma}_i$ *is consistent for each* $i \in N$.

(ii): Suppose that G^i *has generic payoffs*[4]. $\hat{\Gamma}_i$ *is consistent if and only if* s^o *is a Nash equilibrium of* G^i.

Proof. (i) is proved by Theorem 4.2.(i), just counting $\pi_i(s)$, $s \in S$ to be i-formulae as in (17).

(ii): The only-if part (the contrapositive) is proved by (19), which states that if s^o is not a Nash equilibrium, then either $\Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i(\neg\mathrm{I}_i(s_i^o))$, or $\Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i\mathbf{B}_j(\neg\mathrm{I}_j(s_j^o))$ for some $j \neq i$. In either case, it is inconsistent with $\mathrm{EX}_i \cup \{\pi_i(s^o)\}$, i.e., $\hat{\Gamma}_i$ is inconsistent.

[4] This result holds for nongeneric payoffs as well, but we assume genericity for simplicity.

To prove the if part, we first claim that if s^o is a Nash equilibrium, then

$$\Delta_i(\mathbf{g}^i) \nvdash \mathbf{B}_i(\neg \mathrm{I}_i(s_i^o)) \text{ and } \Delta_i(\mathbf{g}^i) \nvdash \mathbf{B}_i\mathbf{B}_j(\neg \mathrm{I}_j(s_j^o)).$$

When G^i is solvable, this follows from Theorem 5.1 (i); when it is unsolvable but has generic payoffs, this follows from Theorem 5.1 (ii). It implies that we can construct a model such that $\Delta_i(\mathbf{g}^i)$, $\mathbf{B}_i(\mathrm{I}_i(s_i^o)), \mathbf{B}_i\mathbf{B}_j(\mathrm{I}_j(s_j^o))$ hold at a specific world. Since the truth value of $\pi_i(s^o)$ is independent of those formulae, we may add it to hold in that world. This gives a model for $\hat{\Gamma}_i$. ■

Again, (i) states that the consistency of the whole statement is decomposed into that for each player. Assertion (ii) is more substantive: player i's observations $\pi_i(s^o)$ are coherent with his *ex ante* basic beliefs if and only if the observed strategy profile constitutes a Nash equilibrium in his perceived game G^i. Theorem 5.2 shows that an observed strategy profile s^o is stable in the sense that it does not bring in inconsistency in any player's mind, if and only if s^o is a Nash equilibrium for each player's perceived game.

When the perceived game G^i is solvable, we may restate (ii) as

$$\hat{\Gamma}_i \text{ is consistent } \iff \Delta_i(\mathbf{g}^i) \vdash \mathbf{B}_i(\mathrm{I}_i(s_i^o)) \wedge [\wedge_{j\neq i}\mathbf{B}_i\mathbf{B}_j(\mathrm{I}_j(s_j^o))]. \tag{29}$$

In this case, $\hat{\Gamma}_i$ is consistent if and only if the observations are already positively predicted in in the *ex ante* stage. As long as the right-hand side of (29) holds (and hence s^o is a Nash equilibrium of G^i), player i can learn nothing from the *ex post* experience.

In contrast, if G^i is unsolvable, $\hat{\Gamma}_i$ can still be consistent unless the observed behavior is predicted not to be played in the *ex ante* stage, and hence, it can be consistent without the right-hand side of (29). However, in this case, $\mathbf{B}_i(\mathrm{I}_i(s_i^o)) \wedge [\wedge_{j\neq i}\mathbf{B}_i\mathbf{B}_j(\mathrm{I}_j(s_j^o))]$ are derived from EX_i and $\pi_i(s^o)$. Then, player i may include $\mathbf{B}_i(\mathrm{I}_i(s_i^o)) \wedge [\wedge_{j\neq i}\mathbf{B}_i\mathbf{B}_j(\mathrm{I}_j(s_j^o))]$, which is learned from the previous experience, into his basic beliefs for his *ex ante* decision in the next play. This introduces a two-way between ex ante decision making and ex post observations; a full development is beyond the current paper.

6 Proof of the Soundness-Completeness of EIR^n

The following proof of completeness is a variant of a known proof of common knowledge logic (cf., Fagin *et al.* [3] and Meyer-van der Hoek [14]). Nevertheless, since we need to take several new steps, we give a full proof.

Since the base logic of EIR^n is classical logic, we use classical tautologies. Lemmas 2.1 and 2.2 list a few basic properties on $\mathbf{B}_i(\cdot)$, and $\mathbf{Ir}_i(\cdot)$.

6.1 Soundness of EIR^n

It suffices to show that all logical axioms are valid, and the four inference rules preserve validity $\models$. Here, we consider these for Axiom IRA$_i$ and Inference IRI$_i$.

Axiom IRA$_i$. Suppose $(M, w) \models \mathbf{Ir}_i(A_1, ..., A_n)$. By V6, $(M, u) \models A_i$ for any u with wR_iu. Hence, $(M, w) \models \mathbf{B}_i(A_i)$. We take an arbitrary u with wR_iu. Let $j \neq i$, and $[(w_0, j_0), ..., (w_\nu, j_\nu), w_{\nu+1}]$

an alternating chain with $(w_0, j_0) = (u, j)$. Then, $[(w, i), (w_0, j_0), ..., (w_\nu, j_\nu), w_{\nu+1}]$ is alternating, too. By V6, we have $(M, w_{\nu+1}) \models A_{i_\nu}$ because $(M, w) \models \mathbf{Ir}_i(A_1, ..., A_n)$. Since $[(w_0, j_0), ..., (w_\nu, j_\nu), w_{\nu+1}]$ is arbitrary with $(w_0, j_0) = (u, j)$, we have $(M, u) \models \mathbf{Ir}_j(\mathbf{A})$. Since u is arbitrary with wR_iu, we have $(M, w) \models \mathbf{B}_i\mathbf{Ir}_j(\mathbf{A})$.

Inference $\mathbf{IRI}_i$: Let $\mathbf{D} = (D_1, ..., D_n)$ be an n-tuple of formulae, and suppose that for all $i \in N$,

$$(M, u) \models D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i(\wedge_{j \neq i} D_j) \text{ for all } u \in W. \tag{30}$$

Let w be arbitrary element in W. If $(M, w) \not\models D_i$, we have $(M, w) \models D_i \supset \mathbf{Ir}_i(\mathbf{A})$. Now, let $(M, w) \models D_i$. Suppose that $[(w_0, i_0), ..., (w_\nu, i_\nu), w_{\nu+1}]$ is an alternating chain with $(w_0, i_0) = (w, i)$. Then, we prove by induction that $(M, w_{k+1}) \models A_{i_k} \wedge (\wedge_{j \neq i_k} D_j)$ for all $k = 0, ..., \nu$. The induction base, i.e., $k = 0$ is: Since $(M, w) \models D_i$, we have $(M, w) \models \mathbf{B}_i(A_i) \wedge \mathbf{B}_i(\wedge_{j \neq i} D_j)$ by (30). By wR_iw_1, it holds that $(M, w_1) \models A_i \wedge (\wedge_{j \neq i} D_j)$. Suppose the induction hypothesis that $(M, w_{k+1}) \models A_{i_k} \wedge (\wedge_{j \neq i_k} D_j)$. This implies $(M, w_{k+1}) \models D_{i_{k+1}}$ since $i_{k+1} \neq i_k$. It follows from this and (30) that $(M, w_{k+1}) \models \mathbf{B}_{i_{k+1}}(A_{i_{k+1}}) \wedge \mathbf{B}_{i_{k+1}}(\wedge_{j \neq i_{k+1}} D_j)$. Since $w_{k+1} R_{i_{k+1}} w_{k+2}$, we have $(M, w_{k+2}) \models A_{i_{k+1}} \wedge (\wedge_{j \neq i_{k+1}} D_j)$. This completes the induction argument. Thus, $(M, w) \models D_i \supset \mathbf{Ir}_i(\mathbf{A})$. Since w is arbitrarily chosen, it holds that $M \models D_i \supset \mathbf{Ir}_i(\mathbf{A})$.

6.2 Completeness for $\mathbf{EIR}^n$

As is standard, supposing $\nvdash A$, we construct a (finite) model $M = (F; \tau) = ((W; R_1, ..., R_n); \tau)$ so that $(M, w) \not\models A$ for some $w \in W$. In the following, A is an arbitrarily fixed formula with $\nvdash A$.

We start with the following facts: Let $\mathcal{A}_m^o = \{A_0, ..., A_m\}$ be a finite set of formulae ($m \geq 0$), and $\mathcal{A}_m = \mathcal{A}_m^o \cup \{\neg A : A \in \mathcal{A}_m^o\}$. Let $\mathbb{W}(\mathcal{A}_m)$ be the set of maximally consistent subsets in $\mathcal{A}_m$. We can construct a maximally consistent in the standard manner; thus, $\mathbb{W}(\mathcal{A}_m)$ is nonempty. We write $\varphi_w = \wedge w$ for $w \in \mathbb{W}(\mathcal{A}_m)$. We stipulate $\wedge\emptyset$ to be $(\neg p) \vee p$.

Lemma 6.1. ***(1)*** *if $w \in \mathbb{W}(\mathcal{A}_m)$ and $t \leq m$, then either $A_t \in w$ or $\neg A_t \in w$;*

(2) *if $w \in \mathbb{W}(\mathcal{A}_m)$, then $w \cap \mathcal{A}_{m-1} \in \mathbb{W}(\mathcal{A}_{m-1})$; and if $w \in \mathbb{W}(\mathcal{A}_{m-1})$, then $w \cup \{A_m\} \in \mathbb{W}(\mathcal{A}_m)$ or $w \cup \{\neg A_m\} \in \mathbb{W}(\mathcal{A}_m)$;*

(3) *for any consistent $v \subseteq \mathcal{A}_m$, $\vdash \wedge v \equiv \vee_{v \subseteq w \in \mathbb{W}(\mathcal{A}_m)} \varphi_w$;*

(4) $\vdash \vee_{w \in \mathbb{W}(\mathcal{A}_m)} \varphi_w$.

Proof. (1) is standard. (4) follows from (3) taking $v = \emptyset$.

(2): Consider the former: Since $w \in \mathbb{W}(\mathcal{A}_m)$, $w \cap \mathcal{A}_{m-1}$ is consistent. Also, $w \cap \mathcal{A}_{m-1}$ is maximal in $\mathcal{A}_{m-1}$ by (1). Now consider the latter: Let $w \in \mathbb{W}(\mathcal{A}_{m-1})$. Then $w \cup \{A_m\}$ or $w \cup \{\neg A_m\}$ is consistent; in either case, it is maximally consistent in $\mathcal{A}_m$ by (1).

(3): Let v be a consistent subset of $\mathcal{A}_m$. Let $\mathcal{A}_m^o(v) = \{C \in \mathcal{A}_m^o : C \notin v \text{ and } \neg C \notin v\}$. If $\mathcal{A}_m^o(v) = \emptyset$, then v is maximal; so $\vee_{v \subseteq w \in \mathbb{W}(\mathcal{A}_m)} \varphi_w$ is written as $\vee\{\varphi_v\}$ and is equivalent to $\varphi_v = \wedge v$ itself. Let $\mathcal{A}_m^o(v) \neq \emptyset$. Take any $C \in \mathcal{A}_m^o(v)$. Let $\mathcal{A}_m' = \mathcal{A}_m - \{C, \neg C\}$. We show

$$\vdash [\vee_{v \subseteq u \in \mathbb{W}(\mathcal{A}_m')} \varphi_u] \equiv [\vee_{v \subseteq u \in \mathbb{W}(\mathcal{A}_m)} \varphi_u]. \tag{31}$$

Using this, we can eliminate, by induction, all such formulae $C, \neg C$ from $\mathcal{A}_m$, and then the first formula in (31) becomes equivalent to $\wedge v$. Let us prove (31). Now, we have

$$\vdash \vee_{v \subseteq u \in \mathbb{W}(\mathcal{A}_m')} \varphi_u \equiv \vee_{v \subseteq u \in \mathbb{W}(\mathcal{A}_m')} (\varphi_u \wedge C) \vee (\varphi_u \wedge \neg C).$$

Since u includes v, one of the disjuncts of the latter is contradictory; a contradictory formula can be deleted from the disjunction. The the remaining one $u \cup \{C\}$ or $u \cup \{\neg C\}$ is consistent and also maximal in $\mathcal{A}_m$. Thus, the latter formula is equivalent to $\vee_{v \subseteq w \in \mathbb{W}(\mathcal{A}_m)} \varphi_w$.■

Now, suppose $\nvdash A$ throughout the following. Now, we construct a model $M = (F, \tau) = ((W; R_1, ..., R_n), \tau)$ so that $(M, w) \nvDash A$ for some $w \in W$.

We denote the following set, by SI(A),

$$\cup_{i \in N} \{\mathbf{B}_i(C_i), \mathbf{B}_i \mathbf{Ir}_j(\mathbf{C}), \mathbf{Ir}_j(\mathbf{C}) : \mathbf{Ir}_i(\mathbf{C}) \text{ is a sub-formula of } A \text{ and } j \neq i\}, \tag{32}$$

and we let $\text{Sub}^o(A) = \{C : C \text{ is a sub-formula of } A\} \cup \text{SI}(A)$. Then, let $\text{Sub}(A) = \text{Sub}^o(A) \cup \{\neg C : C \in \text{Sub}^o(A)\}$. Now, we denote, by $\text{Con}(A) := \mathbb{W}(\text{Sub}(A))$, the set of maximally consistent subsets of Sub(A). Then, we can find a $w \in \text{Con}(A)$ so that

$$A \notin w. \tag{33}$$

Indeed, the supposition that $\nvdash A$ implies the consistency of $\neg A$; thus, we can find a w for (33) by Lemma 6.1.(1).

The following lemma is standard.

Lemma 6.2. . *For each $w \in Con(A)$, we have*
(1): *for any $\neg C \in Sub(A)$, either $C \in w$ or $\neg C \in w$;*
(2): *for any $(D \supset C) \in Sub(A)$, $(D \supset C) \in w$ if and only if $\neg D \in w$ or $C \in w$;*
(3): *for any $\wedge\Phi \in Sub(A)$, $\wedge\Phi \in w$ if and only if $B \in w$ for any $B \in \Phi$;*
(4): *for any $\vee\Phi \in Sub(A)$, $\vee\Phi \in w$ if and only if $B \in w$ for some $B \in \Phi$.*

We denote, by $u^{-\mathbf{B}_i}$, the set $\{C : \mathbf{B}_i(C) \in u\}$ for any set of formulae u. Now we define a model $M = (F, \tau) = ((W; R_1, ..., R_n), \tau)$ as follows:

M1: $W = \text{Con}(A)$;
M2: $R_i = \{(u, v) \in W^2 : u^{-\mathbf{B}_i} \subseteq v\}$ for all $i \in N$;
M3: for any $(w, p) \in W \times PV$, $\tau(w, p) = \top$ if and only if $p \in w$.

We show that $M = (F, \tau)$ is a model for the logic EIR^n.

Lemma 6.3. . *The relation R_i is serial.*

Proof. Let $u \in W$. Then, $u^{-\mathbf{B}_i}$ is consistent; hence there exists some $v \in \text{Con}(A)$ such that $u^{-\mathbf{B}_i} \subseteq v$, i.e., $(u, v) \in R_i$. Indeed, if $\vdash \wedge u^{-\mathbf{B}_i} \supset (\neg C \wedge C)$ for some C. then by Nec and Axiom K, $\vdash \wedge u \supset \mathbf{B}_i(\neg C \wedge C)$; so by Axiom D, u is inconsistent, a contradiction to $u \in W = \text{Con}(A)$.■

We claim that for any $C \in \text{Sub}(A)$ and any $w \in W$,

$$C \in w \text{ if and only if } (M, w) \models C. \tag{34}$$

By (33), we have some $w \in W$ with $A \notin w$. Once (34) is shown, we have $(M, w) \nvDash A$. Thus, it remains to show (34).

Now, (34) is shown by induction on the length of the formula C in Sub(A). By M3, (34) holds for any $p \in PV$. Let C be not a propositional variable. Suppose that (34) holds for any sub-formula of C. We consider three cases divided by the outermost connective of C.

(i) When C is expressed as $\neg D$, $D \supset D'$, $\wedge\Phi$ or $\vee\Phi$, (34) follows from Lemma 6.2.

(ii) Consider $C = \mathbf{B}_i(C')$. First, we show that $(M,w) \models \mathbf{B}_i(C')$ implies $\mathbf{B}_i(C') \in w$. Suppose $(M,w) \models \mathbf{B}_i(C')$. We claim that $w^{-\mathbf{B}_i} \cup \{\neg C'\}$ is inconsistent. Suppose it is consistent. Then there exists some $u \in W$ such that $w^{-\mathbf{B}_i} \cup \{\neg C'\} \subseteq u$; so $C' \notin u$. By the induction hypothesis, $(M,u) \not\models C'$. Since $w^{-\mathbf{B}_i} \subseteq u$, we have wR_iu, and hence $(M,w) \not\models \neg\mathbf{B}_i(C')$, a contradiction. Thus, $w^{-\mathbf{B}_i} \cup \{\neg C'\}$ is inconsistent; so $\vdash \wedge w^{-\mathbf{B}_i} \supset C'$. This implies $\vdash \wedge w \supset \mathbf{B}_i(C')$. Thus, $\mathbf{B}_i(C') \in w$.

Conversely, suppose that $\mathbf{B}_i(C') \in w$. We have $C \in u$ for any u with wR_iu, because $C \in w^{-\mathbf{B}_i} \subseteq u$. By the induction hypothesis, $(M,u) \models C$. Hence $(M,w) \models \mathbf{B}_i(C)$.

(iii) Here we show that (34) holds for the formula $C = \mathbf{Ir}_i(\mathbf{C}) = \mathbf{Ir}_i(C_1,...,C_n)$. The crucial part is the *if* statement, which is proved using Lemma 6.1.

(*Only-if*): Suppose $\mathbf{Ir}_i(\mathbf{C}) \in w$. Let $[(w_0,i_0),...,(w_\nu,i_\nu),w_{\nu+1}]$ be an alternating chain with $(w_0,i_0) = (w,i)$. We show, by induction, that A_{i_k} and $\mathbf{Ir}_{i_{k+1}}(\mathbf{C})$ are in w_{k+1} for all $0 \le k \le \nu$. This implies $(M,w) \models \mathbf{Ir}_i(\mathbf{C})$ since the chain is arbitrary.

Let $k = 0$. Since $\vdash \mathbf{Ir}_i(\mathbf{C}) \supset \mathbf{B}_i(C_i) \wedge \mathbf{B}_i(\wedge_{j\neq i}\mathbf{Ir}_j(\mathbf{C}))$ by IRA$_i$, we have $\mathbf{B}_i(C_i), \mathbf{B}_i\mathbf{Ir}_j(\mathbf{C}) \in w$ for all $j \neq i$. Because $(w,w_1) \in R_i$, $w^{-\mathbf{B}_i} \subseteq w_1$; so $C_i \in w_1$ and $\mathbf{Ir}_{i_1}(\mathbf{C}) \in w_1$. Suppose that $C_{i_k}, \mathbf{Ir}_{i_{k+1}}(\mathbf{C}) \in w_{k+1}$. By IRA$_{i_{k+1}}$, we have $\mathbf{B}_{i_{k+1}}(C_{i_{k+1}}), \mathbf{B}_{i_{k+1}}(\mathbf{Ir}_j(\mathbf{C})) \in w$ for all $j \neq i_{k+1}$. Again, since $(w_{k+1},w_{k+2}) \in R_{i_{k+1}}$, i.e., $w_{k+1}^{-\mathbf{B}_{i_{k+1}}} \subseteq w_{k+2}$, we have $C_{i_{k+1}} \in w_{i_{k+2}}$ and $\mathbf{Ir}_{i_{k+2}}(\mathbf{C}) \in w_{i_{k+2}}$. This concludes the induction argument. Thus, $C_{i_\nu} \in w_{\nu+1}$.

(*If*): Suppose $(M,w) \models \mathbf{Ir}_i(\mathbf{C})$. We define $W^j_{\mathbf{C}} = \{u : (M,u) \models \mathbf{Ir}_j(\mathbf{C})\}$ for $j \in N$. Recall the definition $\varphi_u = \wedge u$ for $u \in W$. Let $\chi_{W^j_{\mathbf{C}}} = \vee\{\varphi_u : u \in W^j_{\mathbf{C}}\}$. We show that for $j \in N$,

$$\vdash \chi_{W^j_{\mathbf{C}}} \supset \mathbf{B}_j(C_j) \wedge \mathbf{B}_j\left(\wedge_{k\neq j}\chi_{W^k_{\mathbf{C}}}\right). \tag{35}$$

Once this is proved, using IRI$_i$, we have $\vdash \chi_{W^i_{\mathbf{C}}} \supset \mathbf{Ir}_i(\mathbf{C})$, which together with $\vdash \varphi_w \supset \chi_{W^i_{\mathbf{C}}}$ implies $\vdash \varphi_w \supset \mathbf{Ir}_i(\mathbf{C})$. Thus, $\mathbf{Ir}_i(\mathbf{C}) \in w$. Now, we prove (35).

We first show that $\vdash \chi_{W^j_{\mathbf{C}}} \supset \mathbf{B}_j(C_j)$. Let $u \in W^j_{\mathbf{C}}$. Since $(M,u) \models \mathbf{Ir}_j(\mathbf{C})$, we have, by V6, $(M,u') \models C_j$ for any u with uR_ju'. By the induction hypothesis for (34), $C_j \in u'$. Since $u^{-\mathbf{B}_j} \subseteq u'$, $\mathbf{B}_j(C_j) \in u$; so $\vdash \varphi_u \supset \mathbf{B}_j(C_j)$. Since this holds for any $u \in W^j_{\mathbf{C}}$, we have $\vdash \chi_{W^j_{\mathbf{C}}} \supset \mathbf{B}_j(C_j)$.

Now we show that $\vdash \chi_{W^j_{\mathbf{C}}} \supset \mathbf{B}_j(\chi_{W^k_{\mathbf{C}}})$ for any $k \neq j$. Let $u \in W^j_{\mathbf{C}}$. It suffices to show that $\vdash \varphi_u \supset \mathbf{B}_j(\chi_{W^k_{\mathbf{C}}})$. It follows from Lemma 6.1.(1) that $\vdash \varphi_{v'} \supset \neg\varphi_v$ for any distinct $v', v \in W$; this together with Lemma 6.1.(3) implies that $\vdash \chi_{W^k_{\mathbf{C}}} \equiv \neg(\vee_{v\in W-W^k_{\mathbf{C}}}\varphi_v)$. By this equivalence, it is enough to show

$$\vdash \varphi_u \supset \mathbf{B}_j(\neg\varphi_v) \text{ for any } v \in W - W^k_{\mathbf{C}}. \tag{36}$$

Suppose that (36) does not hold for some $v \in W - W^k_{\mathbf{C}}$; that is, φ_u and $\neg\mathbf{B}_j(\neg\varphi_v)$ are consistent.

Now, we consider the two cases: (A): $u^{-\mathbf{B}_j} \subseteq v$; and (B) $u^{-\mathbf{B}_j} \nsubseteq v$. We show that neither is the case; so we would have (36), and (35), too.

Consider (A), i.e., uR_jv. We see $(M,v) \models \mathbf{Ir}_k(\mathbf{C})$; indeed, let $[(w_1,i_1),...,(w_\nu,i_\nu),w_{\nu+1}]$ be an alternating chain with $(w_1,i_1) = (v,k)$. Then, since uR_jv, $[(u,j),(w_1,i_1),...,(w_\nu,i_\nu),w_{\nu+1}]$

is also an alternating chain. Since $(M,u) \models \mathbf{Ir}_j(\mathbf{C})$, we have $(M,w_{\nu+1}) \models C_{i_\nu}$. This implies $(M,v) \models \mathbf{Ir}_k(\mathbf{C})$, which is a contradiction to the choice of v from $W - W^k_{\mathbf{C}}$.

Consider (B). Then, $C \in u^{-\mathbf{B}_j}$ but $C \notin v$ for some C. Then, if $\neg C \in \mathrm{Sub}(A)$, then $\neg C \in v$; and if $\neg C \notin \mathrm{Sub}(A)$, then $C = \neg C'$ and $C' \in v$. In either case, $\vdash \varphi_v \supset \neg C$, equivalently, $\vdash C \supset \neg\varphi_v$. Hence, $\vdash \mathbf{B}_j(C) \supset \mathbf{B}_j(\neg\varphi_v)$. Since $\mathbf{B}_j(C) \in u$, we have $\vdash \varphi_u \supset \mathbf{B}_j(\neg\varphi_v)$, which implies that φ_u and $\neg\mathbf{B}_j(\neg\varphi_v)$ are inconsistent.■

7 Conclusions

We have developed the theory of the epistemic infinite-regress logic EIR^n. The logic EIR^n is built for studies of prediction/decision making in interdependent game situations. We gave the completeness theorems for EIR^n with respect to the Kripke semantics. Based on these completeness results, we presented various meta-theorems; the entire discourse is based on the choice of the KD-type epistemic logic, which allow us to treat subjective thinking separately for each player. We also showed applications to game theoretic prediction/decision making.

As expressed in (1) in Section 1, the epistemic infinite regress $\mathbf{Ir}_i(A_1,...,A_n)$ is an infinitary concept, though in this paper it is captured as a fixed-point concept. Another way to capture epistemic infinite regresses is the infinitary logic approach. In this approach, logics are typically very large such as Karp [12] (see Heifetz [4] for infinitary epistemic logics). It would be informative to look at the infinite-regress logic EIR^n from the viewpoint of the infinitary logic approach, which leads to small infinitary logics. This will be discussed in another paper (Hu-Kaneko-Suzuki [6]), and its relationship to the μ-calculus will be studied.

In Section 5, we showed applications of our theory on EIR^n to *ex ante* prediction/decision making and their interactions with *ex post* observations. The relationship between these two viewpoints should be studied in a more general manner including different prediction/decision criteria. This may give new insights to the game theory *per se,* as well as logic in a wider sense including induction.

Our development raises many new questions; some questions such as a relationship to the infinitary logic are rather purely logical (or philosophical in the mathematical sense), and some others indicate interactive relations between subjective thinking and objective observations. Studies on these problems will help us better understand human rational thinking and behavior.

References

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

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

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

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

[5] Hu, T.-W., and M. Kaneko, (2014), Game Theoretic Decidability and Undecidability, Working Paper Series No. E1410, IRCPE, Waseda University.

[6] Hu, T., M. Kaneko, and N.-Y. Suzuki, (2015), Small Infinitary Epistemic Logics and Some Fixed-Point Logics, to be completed in 2015.

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

[8] Kaneko, M., and T. Nagashima, (1997), Game logic and its applications II, *Studia Logica* 58, 273–303.

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

[10] Kaneko, M., (2004), *Game Theory and Mutual Misunderstanding*, Springer, Heidelberg.

[11] Kaneko,M., Suzuki, N.-Y., (2003), Epistemic models of shallow depths and decision making in games:Horticulture. *Journal of Symbolic Logic* 68, 163–186.

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

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

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

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

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

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

[18] Suzuki, N.-Y., (2013), Semantics for intuitionistic epistemic logics of shallow depths for game theory. *Economic Theory* 53, 85-110.

[19] A. S. Troelstra, A. S., and Schwichtenberg, (2000), H. *Basic Proof Theory,* Cambridge Tracts in Theoretical Computer Science, Cambridge.

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

[21] 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.

[22] Venema, Y. (2008), Lectures on μ-calculus. University of Amsterdam.