Economic Theory **19**, 7–62 (2002)

Economic Theory

Epistemic logics and their game theoretic applications: Introduction★

Mamoru Kaneko

Institute of Policy and Planning Sciences, University of Tsukuba, Ibaraki 305-8573, JAPAN
(e-mail: kaneko@shako.sk.tsukuba.ac.jp)

Received: July 1, 2000; revised version: April 19, 2001

Summary. This paper is written as an introduction to epistemic logics and their game theoretic applications. It starts with both semantics and syntax of classical logic, and goes to the Hilbert-style proof-theory and Kripke-style model theory of epistemic logics. In these theories, we discuss individual decision making in some simple game examples. In particular, we will discuss the distinction between beliefs and knowledge, and how false beliefs play roles in game theoretic decision making. Finally, we discuss extensions of epistemic logics to incorporate common knowledge. In the extension, we discuss also false beliefs on common knowledge.

Keywords and Phrases: Classical logic, Epistemic logic, Common knowledge logic, Beliefs, Knowledge, Dominant strategy, Decision criterion, Epistemic depth of a formula.

JEL Classification Numbers: C60, C69, D89.

1 Introduction

1.1 Aim and some basic notions in logic

This paper is written for economists and/or game theorists as an introduction to epistemic logics and their applications to game theoretic problems. We believe

★ I thank Jeffrey J. Kline for many detailed comments on earlier drafts, and Nobu-Yuki Suzuki for helpful discussions on subjects covered in this paper. I am also grateful to the members, A. Kajii, A. Matsui, K. Tatamitani, S. Turnbull, T. Ui, of the logic study group here for helpful discussions and comments.
I am partially supported by Grant-in-Aids for Scientific Research No. 13430001, Japan Society for the Promotion of Sciences.

that epistemic interactions play key roles in social behavior of people, and also that symbolic expressions, manipulations and their interpretations are central in such epistemic interactions. For these beliefs, we adopt the research strategy to take and use basic concepts and results developed in mathematical logic.

The declaration of our ultimate aim may help readers understand our research attitude: Our logic approach might be regarded as the pursuit of foundations of extant game theories, as mathematical logic may be viewed as the study of foundations of mathematics. Contrary to this, we have little intentions of pursuing such foundations. A typical characteristic of an extant game theory is the pursuit of "rationality" in outcomes assuming, often unintentionally, a lot of transcendentalities. Although our approach shares target problems with extant theories, we avoid and/or examine consciously transcendentalities involved in extant theories. Instead, we view problems of finite nature as central in investigations of human behavior. From this point of view, we develop our logical approach.

We follow the standard bases of mathematical logic. For such bases, there are many entrance barriers which economists/game theorists may encounter, because of methodological differences between economics and logic. These barriers may inhibit future economic and game theoretic research, but some of their constituents may become important for future research itself. It is now timely to give a systematic introduction, emphasizing such methodological differences, to epistemic logics with some illustrations of game applications. We hope that this introduction will induce further developments of the logical approach.

Economics and game theory have the tradition that their mathematical methodology is based primarily on analysis such as topology, functional analysis and probability theory. Also in logic, we can treat these mathematical fields, but what we emphasize by the logical approach is the basic constructions of logic rather than direct applications of extant results in the field of logic. Logic has various unique constituents that are not found in other mathematical fields. This paper introduces such unique constituents to economists/game theorists. In the rest of this section, we mention several pairs of basic concepts unique to logic and particular to epistemic logic, which would help the reader understand the subjects better.

The very basic starting point of logic is the separation between symbolic expressions and their intended meanings. When we target human thinking seriously, this separation is unavoidable. It is stated in the terminology of logic as:

A1: Syntax vs. semantics

As a syntactical notion, we define a *formula* to be a symbolic expression based on given primitive symbols. As a semantical notion, we define a truth valuation of such a formula. This separation leads to two different theories:

A2: Proof theory vs. model theory

In the former, mathematical reasoning is captured as grammatical symbol manipulations from given axioms, while in the latter, mathematical models satisfying those axioms are considered. These theories are connected by the, so-called, *soundness-completeness* theorem. In this paper, we use the word “model theory” as almost synonymous to “semantics”.

The above connection is important conceptually as well as technically. Model theory talks about models, each of which is assumed to be a *complete* description of a target situation in the sense that any sentence is either *true* or *false*.[1] A proof theory talks about proofs, i.e., the provability of a formula from a given set of (nonlogical) axioms. The given set of axioms may be a *partial* description of the target situation. Since each model is complete, a single model is too much to capture the partial description. Therefore, we consider a *set* of candidate models for the partial description. If a formula is true for all the possible candidate models, it is said to be *valid*. The soundness-completeness theorem is a bridge between the syntactical provability and semantic validity.

We take the view that a player has only a partial description of a target situation. Therefore, we cannot adopt a single model as a description of the situation. We should treat a set of nonlogical axioms in a proof theoretic manner or a set of models in a model theoretic manner. This differs from the game theoretic literature of epistemic models since Aumann [1], where a single model is usually assumed to be a description of the target situation.[2]

Another relevant distinction here is:

A3: Object theorems vs. meta-theorems

A theorem whose provability and/or validity is discussed *inside* a logical system belongs to the former, and a theorem on a logical system belongs to the latter. Since our purpose is to investigate the players’ inferences required for decision making, meta-theorems on their decisions and inferences are our central concerns. A translation of a result in an extant game theory into a logical system is an instance of the former, and is not really our concern. The distinction will be clear when some examples are given.

It is important to notice that mathematical logic is a mathematical theory of mathematical theories. We use a standard mathematical method to handle a logical system. The mathematical method to handle a logical system is called *meta-mathematics*. This will be pointed out when it is relevant.[3]

In this paper, we will discuss epistemic logics, which are variants of modal logics originally targeting the investigation of “necessity” and “possibility”. We can borrow a lot from this literature.[4] In the case of epistemic logics, the above distinction A2 becomes.

[1] This completeness is assumed in the classical model theory, but not necessarily assumed in general. For example, a Kripke model for intuitionistic logic does not assume this completeness. See van Dalen [37].

[2] See Aumann [2] for some use of logical apparatuses from the viewpoint of the recent game theoretic tradition. For the recent game theoretical literature of epistemic models, see Bacharach-Mongin [3] and Bacharach et al. [4].

[3] See Kleene [21] for the distinction between object-mathematics and meta-mathematics.

[4] Hughes and Cresswell [12] and Chellas [6] are standard textbooks of modal logic. The modern literature of epistemic logic was started by Hintikka [10]. Fagin-Halpern-Moses-Verdi [7] and Meyer and van der Hoek [25] are found as textbooks on epistemic logics and other subjects.

A4: Hilbert-style proof theory vs. Kripke semantics

We discuss mainly these two theories in this paper. The Hilbert-style proof theory is convenient in its concise presentation. A strength of the logical approach is to show *unprovability*. In this respect, the Hilbert-style proof theory is difficult to handle, and the Kripke semantics is unavoidable. We will mention the other proof theory called *Gentzen-style*, which enables us to evaluate unprovability as well as provability. In this paper, however, we give only a brief explanation of it in Section 4.4.

Since the above two theories are deductively equivalent, which is stated by the soundness-completeness theorem, the reader may wonder why we adopt both theories. The answer is double-fold. First, technically speaking, since each theory has some merits and some demerits, it would be more powerful to have both theories. Second, conceptually speaking, both syntactical manipulations and semantical interpretations are important in investigations of human-thoughts in social contexts. Therefore, we keep the dualistic attitude. Note that there are other proof and model theories in the literature of logics (cf., Kleene [21] and Chellas [6]).

Another relevant distinction is

A5: Propositional logics vs. predicate logics

If a target problem has only a finite number of objects such as a finite game with pure strategies, a propositional logic would suffice, and if it has an infinite number of objects such as a game with mixed strategies, a predicate logic would be unavoidable. A predicate logic is an extension of a propositional logic so that it allows for quantifications, $\forall$ (for all) and $\exists$ (exists). When the problem treats only a finite set, the quantifications, $\forall$ and $\exists$, can be expressed by conjunction $\bigwedge$ (and) and disjunction $\bigvee$ (or), respectively. Therefore, as far as we confine ourselves to finite games, only propositional logics suffice.[5] In this paper, we discuss only propositional epistemic logics.

1.2 Logical approach to game theoretic problems

We now turn our attention from broad distinctions in logic to ones particular to epistemic logics and their game applications. The first distinction is:

B1: Classical logic vs. epistemic logics

Classical logic is the logic used in the standard mathematical practices. We adopt classical logic as the base logic for our epistemic logics. This means that the investigator (observer)'s reasoning ability is described by classical logic. The reasoning ability of each player is assumed to consist of the ability described by classical logic and the additional inference ability of (self-)introspection.

[5] Even if the problems have only finite objects, predicate logics may be relevant to some problems such as complexities of expressions (which are relevant for some situations, e.g., communication within language). Kleene [21] and Mendelson [24] are classical textbooks treating basics of predicate logic.

The second distinction is important for game theoretic applications:

B2: Logical axioms vs. nonlogical (mathematical) axioms

This distinction can be, more or less, arbitrary in classical logic, but it is crucial in epistemic logics (modal logics in general). We take the research strategy of separating game problems from logical problems as much as possible. To achieve this separation, we will describe game theoretic axioms as nonlogical axioms.

The third distinction is:

B3: Beliefs vs. knowledge

We define "knowledge" as "true beliefs", where truth is referred to the outside thinker. Beliefs are also divided into basic and inferred ones. A justification of a belief for a player is an argument (proof) for it from *basic* beliefs by himself. We do not discuss "justifications" of basic beliefs, which is a limit of the logical approach. We allow *false beliefs*, rather than discuss whether a player can obtain true beliefs.[6] False beliefs enable us to consider the emergence of beliefs from other sources such as individual experiences. We will treat the *truthfulness axiom*, which makes beliefs true, as a possible axiom rather than a basic axiom. We will discuss distinction B3 from the logical point of view in Section 6.

The following two game theoretical distinctions are important in our approach:

B4: Solution theory vs. performance-playability theory

A solution theory addresses what criteria are adopted for decision making, while performance - playability theory takes a solution theory as given and addresses how the theory performs and whether the player makes a decision. In this paper, we discuss both theories.

A related distinction is:

B5: Decision vs. prediction

In a game, each player makes a decision under predictions about other players' decisions. A decision is ultimately important for each player, and predictions about others' decisions are auxiliary. For some games, a decision may be made without predictions. We do not need to assume the same decision criteria for decision and prediction. These differ by nature. Traditional game theory has not distinguished between them. In this paper, we make this distinction, but will not have enough space to examine it fully. This distinction will be clearer in Kaneko and Suzuki [19].

1.3 Konnyaku Mondô (Jelly dialogue)

Before starting our discussions on the logical approach to game theory, we mention a Japanese traditional comic (rakugo) story suggestive for the distinctions

[6] We do not relate individual beliefs to *subjective probability*. The reader may understand this by reading the basic principles for beliefs in Section 4.2.

mentioned above. This story is indirectly related to the main body of the paper, but helps the reader understand the basic notions discussed in the rest of paper.

Konnyaku Mondô: A (devil's tongue) jelly maker lived in a Buddhist temple pretending to be a monk. A real Buddhist monk came to visit the temple to have a dialogue on Buddhism thoughts. The jelly maker first refused but eventually agreed to have a dialogue. Since the jelly maker did not know how he could communicate with the monk on Buddhism, he answered the monk's questions in gestures. The monk took this as a style of dialogue, and responded in gestures. After some exchanges of gestures, both thought that the jelly maker defeated the monk. After the dialogue, a witness asked the monk about the dialogue. The monk said that the jelly maker had a great Buddhism thought shown by his gestures and should be respected. Afterwards, the jelly maker was asked by the same witness and answered: the monk started talking badly about jelly products with his gestures, made the jelly maker angry, and thus the jelly maker beat the monk. Thus, each of them believed that they had perfectly meaningful dialogue and that it was common knowledge that the jelly maker defeated the monk in the dialogue. However, the monk believed that they had a Buddhism dialogue, while the jelly maker believed that they had discussed about jelly products (pp. 61–70 in [35]).

This story has various relevant points for the distinctions mentioned in this section. First, the gestures exchanged and the associated beliefs are distinguished as syntax and semantics – A1. Second, different people associate different interpretations with the same gestures, and they develop false beliefs, which corresponds to the distinction between beliefs and knowledge – B3. Third, the other person's mind is only imagined in the mind of a person. An implication is that prediction about the other's decision differs considerably from one's own decision – B5. The story also suggests the possibility that an individual person may develop a false belief of common knowledge from common observations. The last point will be discussed in a game theoretical context in Section 8.3.

This paper is organized as follows: Section 2 gives basic game theoretic concepts to be used for illustrations. Section 3 describes semantics and syntax of classical logic CL, and states the soundness-completeness theorem. Its proof will be given in the Appendix. Sections 4 and 5 give various epistemic logics and Kripke semantics, which are connected, again, by the soundness-completeness theorem. Section 6 discusses the relationship between "beliefs" and "knowledge". Section 7 discusses decision criteria. We show that in some case, a solution for decision making involves common knowledge, and that the epistemic logics introduced in Section 4 are incapable of treating such a problem. In Section 8, we consider an extension of an epistemic logic to incorporate common knowledge.

2 Basic game theoretic concepts

To exemplify logical constructs, we will refer to a 2-person finite noncooperative game $g = (g_1, g_2)$ in strategic form. Each player $i = 1, 2$ has ℓ_i pure strategies

($\ell_i \geq 2$). We assume throughout the paper that the players do not play mixed strategies. Player i's strategy space is denoted by $S_i := \{\mathbf{s}_{i1}, ..., \mathbf{s}_{i\ell_i}\}$ for $i = 1, 2$. His *payoff function* is a real-valued function g_i on $S := S_1 \times S_2$.

Let $(s_1, s_2) \in S$. We say that s_1 is a *best strategy to* s_2 iff $g_1(s_1, s_2) \geq g_1(t_1, s_2)$ for all $t_1 \in S_1$. We say that s_1 is a *dominant strategy* iff s_1 is a best strategy to s_2 for any $s_2 \in S_2$. Player 2's dominant strategy is defined in the parallel manner. A strategy pair $s = (s_1, s_2)$ is called a *Nash equilibrium* iff s_i is a best strategy to s_j for $i, j = 1, 2$ ($i \neq j$). We say that s_i is a *Nash strategy* for player i iff (s_1, s_2) is a Nash equilibrium for some s_j, where $i, j = 1, 2$ ($i \neq j$).

In the game $g^1 = (g_1^1, g_2^1)$ of Table 1 (Prisoner's Dilemma), the second strategy $\mathbf{s}_{i2}$ for each i is a dominant strategy. In the game $g^2 = (g_1^2, g_2^2)$ of Table 2 which is obtained from g^1 by changing the payoff 6 in the northeast corner to 2, only player 1 has a dominant strategy, $\mathbf{s}_{12}$. Either game has a unique Nash equilibrium, $(\mathbf{s}_{12}, \mathbf{s}_{22})$, which is marked with asterisk $*$.

Table 1. $g^1 = (g_1^1, g_2^1)$

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	(5,5)	(1,6)
$\mathbf{s}_{12}$	(6,1)	(3,3)*

Table 2. $g^2 = (g_1^2, g_2^2)$

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	(5,5)	(1,2)
$\mathbf{s}_{12}$	(6,1)	(3,3)*

Here, we briefly describe in the standard game theory language what decision criteria are candidates for these games. Later, we will see how such criteria are more accurately described in epistemic logics.

We start with the following simple decision criterion:

DC1: Player i should choose a dominant strategy.

In game $g^1 = (g_1^1, g_2^1)$, this criterion recommends a decision to either player. However, it recommends a strategy only to player 1 in game $g^2 = (g_1^2, g_2^2)$, since 2 has no dominant strategies in g^2. One way out for 2 is to predict 1's decision, assuming that 1 adopts DC1 for 1's choice. We write this criterion as follows:

DC2: Player i, predicting that player j ($j \neq i$) would choose a strategy following DC1 should choose a best strategy to his predicted strategy for player j.

This differs from DC1 in that it involves a prediction about the other player' decision making. The application of DC2 to player 2 in game g^2 states that 2 predicts that 1 would choose $\mathbf{s}_{12}$ as the dominant strategy and then 2 should choose $\mathbf{s}_{22}$ as the best strategy to $\mathbf{s}_{12}$. This argument may be regarded as a special case of the procedure so called the *iterated elimination of dominated strategies* (cf. Moulin [28] and Myerson [29]). The main concern in this literature is the consideration of a resulting outcome of such a procedure, but our concern is the considerations of required epistemic aspects for such a decision criterion and of its performance-playability relative to given beliefs. This point is clearer in Kaneko and Suzuki [19].

We consider other two criteria, the first of which is auxiliary.

DC3^0: Player i should choose a Nash strategy without thinking about j's beliefs.

This differs considerably from DC2 in that player i thinks about player j's beliefs in DC2 but not in DC3^0. The game, $g^3 = (g^3, g^3)$, of Table 3 is obtained from g^2 by adding one strategy to player 2. In this game, neither player has a dominant strategy. Hence, neither DC1 nor DC2 gives a decision to a player. DC3^0 makes a recommendation, but does not guarantees that player i believes that his prediction is taken by the other player. To guarantee each player to believe that his prediction would be taken by the other player, we change DC3^0 as follows:

DC3: (1) Player 1 should choose a best strategy to his prediction based on (2) below;

(2): player 2 should choose a best strategy to his prediction based on (1) above.

One problem is whether or not DC3 leads to a Nash equilibrium, but our main concern is to consider the epistemic structure involved in DC3. In fact, if we assume that player 1 adopts (1), that the player 2 in 1's mind adopts (2), and so on, then we would meet the infinite regress:

$$(1) \quad \leftarrow -- \quad ((2) \leftarrow -- ((1) \leftarrow -- ((2) \leftarrow -- \ldots$$

$$(2) \quad \leftarrow -- \quad ((1) \leftarrow -- ((2) \leftarrow -- ((1) \leftarrow -- \ldots$$

That is, player 1 believes that in 1's mind, player 2's prediction is based on (2), in 1's mind player 2 believes that in 2's mind, 1's prediction is based on (1), and so on. This infinite regress is closely related to the common knowledge of this criterion. When an individual player adopts this criterion, the infinite regress *appears in his mind*, and it takes only the form of an *individual belief* of common knowledge. Here, we would like to differentiate common knowledge from an individual belief of it. Here, the distinction B3 of Section 1 and the Konnyaku Mondô become relevant.

These criteria suggest that we could find a lot of decision criteria in that a decision criterion is genuinely subjective and belongs to each player's mind. Kaneko and Suzuki [19] discuss, stressing the bounded interpersonal introspections, the multitude of such decision criteria.

Here, we point out the difference of DC3 from the other criteria. Criteria DC1 and DC2 (as well as DC3^0) can be discussed in an (purely finitary) epistemic logic, but not DC3. To capture the infinite regress in DC3, we need an extension of an epistemic logic to incorporate common knowledge. Even if common knowledge is included, it may be possible to allow false beliefs. For example, player 1 believes the common knowledge of playing game g^3, while 2 does the common knowledge of playing the game g^4 of Table 4. To discuss those problems in meaningful manners, we need to develop epistemic logics carefully. These problems will be discussed in Sections 7 and 8.

Table 3. $g^3 = (g_1^3, g_2^3)$

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	(5,5)	(1,2)	(4,3)
$\mathbf{s}_{12}$	(6,1)	(3,3)*	(0,2)

Table 4. $g^4 = (g_1^4, g_2^4)$

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	(6,3)	(1,2)	(0,5)
$\mathbf{s}_{12}$	(5,2)	(3,3)*	(4,1)

3 Classical logic CL

In this section, we review classical logic CL and its semantics. The reader may feel that this is a detour to the logical approach to game theory. However, we will define epistemic logics as superstructures of classical logic CL, and will use a lot of concepts and results from CL in the development of epistemic logics as well as in applications to game theory. Therefore, the reader should become a bit familiar to CL before going to epistemic logics.

In Section 3.1, we define two sets of formulae. In Section 3.2, we give the classical semantics. In Section 3.3, we give one axiomatic presentation of classical logic CL, and state the soundness-completeness theorem for CL.

3.1 The sets of formulae: $\mathscr{P}$ and $\mathscr{P}^{\mathrm{n}}$

We start with the following list of symbols:

propositional variable symbols: $\mathbf{p}_0, \mathbf{p}_1, ...$;
logical connective symbols: $\neg$ (not), $\supset$ (implies), $\bigwedge$ (and), $\bigvee$ (or) ;
unary belief operator symbols: B_1, $\mathrm{B}_2, ...,$ B_n;
parentheses: (,) ; *braces*: { , }; and *comma*: , .

We stress that these are pure symbols and are used to be elements of more complex expressions, called formulae. We will associate the *intended* meanings, “not”, “implies”, “and”, “or”, with $\neg$, $\supset$, $\bigwedge$, $\bigvee$, respectively. The implication symbol $\supset$ should be distinguished from the set-theoretic inclusion $\subseteq$. These intended meanings will be defined operationally by logical axioms and inference rules. Unary belief operator symbol B_i is the belief operator of player i and is applied to each formula. We denote the set of players by $N = \{1, ..., n\}$. The set-theoretic brackets $\{,\}$ are used to express a finite set of formulae. It is assumed that the number of propositional variables is at least one and at most countable. The set of propositional variables is denoted by PV.

Based on the above list symbols, we define *formulae* inductively as follows:

F1: any $p \in PV$ is a formula;
F2: if A and B are formulae, so are $(\neg A), (A \supset B)$ and $\mathrm{B}_i(A)$ $(i \in N)$;
F3: if $\{A_0, A_1, ..., A_m\}$ is a finite set of formulae with $m \geq 0$, then $(\bigwedge\{A_0, A_1, ..., A_m\})$ and $(\bigvee\{A_0, A_1, ..., A_m\})$ are also formulae;
F4: every formula is obtained by a finite number of applications of F1, F2 and F3.[7]

[7] This definition deviates from the standard textbook definition of formulae in that conjunctive and disjunctive connectives $\bigwedge$ and $\bigvee$ are applied to a finite nonempty set of formulae, e.g.,

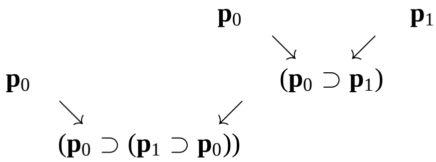

Figure 1

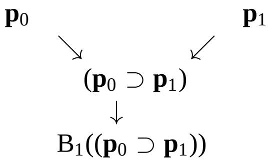

Figure 2

For example, $(\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0))$ is obtained by applications of F1 and F2 three times and twice, respectively, and is intended to mean that if $\mathbf{p}_0$ holds, then $\mathbf{p}_1$ implies $\mathbf{p}_0$. The construction of two formulae $(\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0))$ and $\mathrm{B}_1((\mathbf{p}_0 \supset \mathbf{p}_1))$ are described as Figures 1 and 2.

In general, a formula has a *finite* tree structure where each terminal node corresponds to a propositional variable and each nonterminal node corresponds to a logical connective or a belief operator. This tree structure will be used to construct inductive proofs. We denote the *set of all formulae* by $\mathscr{P}$.

By a formula $\mathrm{B}_i(A)$, we intend to mean that player i believes formula A. The behavior of the belief operators is of our central interests. In classical logic CL, however, we first ignore these formulae, and later, will give a remark (Remark 3.5) on a somewhat nominal treatment of belief operators. We say that a formula A is *nonepistemic* iff A contains no $\mathrm{B}_1, ..., \mathrm{B}_n$. We denote the set of all nonepistemic formulae by $\mathscr{P}^{\mathrm{n}}$. Of course, $\mathscr{P}^{\mathrm{n}} \subseteq \mathscr{P}$.

In this paper, we do not fix exact rules of abbreviations of parentheses (,), but follow standard practices of abbreviations so that we could recover the original expressions when necessary. For example, $(\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0))$, $\mathrm{B}_1((\mathbf{p}_1 \supset \mathbf{p}_0))$ and $(\bigwedge \Phi)$ are abbreviated as $\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0)$, $\mathrm{B}_1(\mathbf{p}_1 \supset \mathbf{p}_0)$ and $\bigwedge \Phi$, respectively. We will also abbreviate $\bigwedge\{A, B\}, \bigvee\{A, B, C\}$ as $A \wedge B, A \vee B \vee C$, etc.[8] We denote $(A \supset B) \wedge (B \supset A)$ by $A \equiv B$.[9]

To discuss the game theoretic problems of Section 2, we adopt the economics practice to represent a payoff function in terms of preference relations. We start with:

strategy *symbols*: $\mathbf{s}_{11}, ..., \mathbf{s}_{1\ell_1}; \mathbf{s}_{21}, ..., \mathbf{s}_{2\ell_2}$;

4-ary *symbols*: P_1, P_2.

Strategy symbols are identical to those given in Section 2. By a 4-ary symbol P_i, we mean that the expression $P_i(s_1, s_2 : t_1, t_2)$ is allowed for $(s_1, s_2), (t_1, t_2) \in S$. These 4-ary expressions are called *atomic formulae,* and the set of them is denoted

$\bigwedge\{A_0, A_1, ..., A_m\}$, rather than to an ordered pair of formulae. We take this deviation to facilitate game theoretical applications. However, the resulting logical systems are equivalent (with respective to provabilities or validities defined in the systems). This formulation does not fit to a Gödel numbering. If one wants to take a Gödel numbering, then he should return to the standard formulation.

[8] In the definition of formulae, we presume the identity of a finite set. Hence, $(\bigwedge\{A_1, A_2\})$ is identical to $(\bigwedge\{A_2, A_1\})$ as a formula.

[9] We introduce four logical connectives, $\neg$, $\supset$, $\bigwedge$ and $\bigvee$. In fact, some of them are enough and the others can be defined as abbreviations. These abbreviations may be convenient for some presentation purposes, but not necessarily so for other purposes. This is rather a matter of taste.

by AF. When $\ell_1 = \ell_2 = 2$, AF consists of 32 atomic formulae. When we discuss game theoretic problems, we regard always AF as PV. Also, the sets of formulae $\mathscr{P}$ and $\mathscr{P}^{\mathrm{n}}$ are defined based on AF as the replacement of PV. We see presently how some game theoretic concepts are described. It should be noted that atomic formula $P_i(s_1, s_2 : t_1, t_2)$ is intended to be a *weak preference* for (s_1, s_2) over (t_1, t_2).

The statement that s_1 is a best strategy to s_2 is described as the formula $\bigwedge\{P_1(s_1, s_2 : t_1, s_2) : t_1 \in S_1\}$, which we denote by $\mathrm{Best}_1(s_1 \mid s_2)$. The statement that s_1 is a dominant strategy for 1 is expressed as $\bigwedge\{\mathrm{Best}_1(s_1 \mid s_2) : s_2 \in S_2\}$. This means that s_1 is the most preferable whatever 2 would choose. The formula $\bigwedge\{P_1(s_1, t_2 : t_1, t_2) : t_1 \in S_1 \text{ and } t_2 \in S_2\}$ is equivalent to $\bigwedge\{\mathrm{Best}_1(s_1 \mid s_2) : s_2 \in S_2\}$ in the logic we will define. We denote the former by $\mathrm{Dom}_1(s_1)$. In the parallel manner, we define the formulae, $\mathrm{Best}_2(s_2 \mid s_1)$ and $\mathrm{Dom}_2(s_2)$. The statement that (s_1, s_2) is a Nash equilibrium is described as $\mathrm{Best}_1(s_1 \mid s_2) \wedge \mathrm{Best}_2(s_2 \mid s_1)$, which is denoted by $\mathrm{Nash}(s_1, s_2)$. The statement that s_1 is a Nash strategy for player 1 is described as $\bigvee\{\mathrm{Nash}(s_1, s_2) : s_2 \in S_2\}$. This is abbreviated as $\bigvee_{s_2}\mathrm{Nash}(s_1, s_2)$.

Recall that a payoff function g_1 of player 1 was given as a real-valued function. Here, we express a payoff function g_1 by the following set of preferences:

$$\{P_1(s : t) : g_1(s) \geq g_1(t)\} \cup \{\neg P_1(s : t) : g_1(s) < g_1(t)\}. \tag{3.1}$$

This is a set of symbolic expressions and is denoted by $\hat{g}_1$. We can take the conjunction, $\bigwedge \hat{g}_1$, of this set. In the parallel manner, $\hat{g}_2$ and $\bigwedge \hat{g}_2$ are defined. Thus, the payoff functions for both players are described as the set $\hat{g} = \hat{g}_1 \cup \hat{g}_2$ or as the formula $\bigwedge(\hat{g}_1 \cup \hat{g}_2)$.

3.2 Classical semantics

So far, we have defined formulae expressing logical or game theoretic ideas, but we have not considered a way of evaluating them. In this subsection, we define semantical notions, “truth” and “falsity”. From these, we define another semantical notion, “validity”, which will be connected to a syntactical notion, “provability”.

First, we give the definition of a semantical valuation of each formula in $\mathscr{P}^{\mathrm{n}}$. In Remark 3.5, we mention the modification of this definition for $\mathscr{P}$.

A function $\kappa : PV \to \{\top, \bot\}$ is called an (classical) *assignment*, where $\top$ and $\bot$ are the symbols designating “true” and “false”. We extend each assignment κ to the function $V_\kappa : \mathscr{P}^{\mathrm{n}} \to \{\top, \bot\}$ by the following induction on the length (tree structure) of a formula:

C0: for any $p \in PV$, $V_\kappa(p) = \top$ iff $\kappa(p) = \top$;
C1: $V_\kappa(\neg A) = \top$ iff $V_\kappa(A) = \bot$;
C2: $V_\kappa(A \supset B) = \top$ iff $V_\kappa(A) = \bot$ or $V_\kappa(B) = \top$;
C3: $V_\kappa(\bigwedge \Phi) = \top$ iff $V_\kappa(A) = \top$ for all $A \in \Phi$;
C4: $V_\kappa(\bigvee \Phi) = \top$ iff $V_\kappa(A) = \top$ for some $A \in \Phi$.

By induction, the value $V_\kappa(A)$ is defined for every $A \in \mathscr{P}^{\mathrm{n}}$. That is, each step defines its left-hand side by the right-hand side. For example, when $PV = \{\mathbf{p}_0, \mathbf{p}_1\}$ and $\kappa(\mathbf{p}_0) = \top$ and $\kappa(\mathbf{p}_1) = \bot$, we calculate $V_\kappa(\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0))$ from the leaves of Figure 1, and obtain $V_\kappa(\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0)) = \top$. In fact, the valuation of this formula is $\top$ independent of κ. Such a formula is a classical tautology. More precisely, we say that $A \in \mathscr{P}^{\mathrm{n}}$ is a *classical tautology* iff $V_\kappa(A) = \top$ for all assignments κ. In fact, $A \supset (B \supset A)$ is a tautology for any formulae $A, B \in \mathscr{P}^{\mathrm{n}}$: Indeed, for any κ, if $V_\kappa(A) = \bot$, then $V_\kappa(A \supset (B \supset A)) = \top$ by C3, and if $V_\kappa(A) = \top$, then $V_\kappa(B \supset A) = \top$ by C3 and thus, $V_\kappa(A \supset (B \supset A)) = \top$.

To describe game theoretic assumptions, we will use *nonlogical* axioms. Let Γ be a subset of $\mathscr{P}^{\mathrm{n}}$. We say that κ is a *model of* Γ iff $V_\kappa(C) = \top$ for all $C \in \Gamma$. For any $A \in \mathscr{P}^{\mathrm{n}}$, we say that A is a *semantical consequence* of Γ iff $V_\kappa(A) = \top$ for all models κ of Γ, in which case we write $\Gamma \vDash A$. When Γ is empty, we write simply $\vDash A$. In this case, A is a classical tautology. We write $\Gamma \nvDash A$ iff not $\Gamma \vDash A$. Note that this differs from $\Gamma \vDash \neg A$. This means that the relation $\Gamma \vDash$ is not necessarily complete.

Let us return to game problems. Consider the game $g^1 = (g_1^1, g_2^1)$ of Table 1. Recall that the payoff functions g_1^1 and g_2^1 are described as $\hat{g}_1^1$ and $\hat{g}_2^1$ defined by (3.1). Since $\hat{g}_i^1$ contains either $P_i(s : t)$ or $\neg P_i(s : t)$ for each $P_i(s : t) \in AF$, the value of a model κ of $\hat{g}_i^1$ on $P_i(s : t)$ is uniquely determined, that is, $\hat{g}_1^1 \cup \hat{g}_2^1$ has a unique model κ, in which sense $\hat{g}_1^1 \cup \hat{g}_2^1$ is a *complete* description of game g^1 (up to the orderings determined by the payoff functions g_1 and g_2).

Consider a model κ of $\hat{g}_1^1$. In this case, since κ is arbitrary on $P_2(s : t)$, $\hat{g}_1^1$ allows 2^{16} models. For any model κ of $\hat{g}_1^1$, we have $V_\kappa(\mathrm{Dom}_1(\mathbf{s}_{12})) = \top$ and $V_\kappa(\neg\mathrm{Dom}_1(\mathbf{s}_{11})) = \top$. Thus,

$$\hat{g}_1^1 \vDash \mathrm{Dom}_1(\mathbf{s}_{12}) \text{ and } \hat{g}_1^1 \vDash \neg\mathrm{Dom}_1(\mathbf{s}_{11}). \tag{3.2}$$

That is, if $\hat{g}_1^1$ is assumed, $\mathrm{Dom}_1(\mathbf{s}_{12})$ and $\neg\mathrm{Dom}_1(\mathbf{s}_{11})$ are derived as unantical consequences of $\hat{g}_1^1$. Also, it holds that $\hat{g}_2^1 \vDash \mathrm{Dom}_2(\mathbf{s}_{22})$ and $\hat{g}_2^1 \vDash \neg\mathrm{Dom}_2(\mathbf{s}_{21})$. When $\hat{g}_1^1 \cup \hat{g}_2^1$ is assumed, it holds that

$$\hat{g}_1^1 \cup \hat{g}_2^1 \vDash \mathrm{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22}).$$

Unless $\hat{g}_2^1$ is assumed, $V_\kappa(\mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{12})) = \bot$ for some model κ of $\hat{g}_1^1$, and thus,

$$\hat{g}_1^1 \nvDash \mathrm{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22}). \tag{3.3}$$

That is, unless enough information is assumed, it is not concluded that $(\mathbf{s}_{12}, \mathbf{s}_{22})$ is a Nash equilibrium.

The above are, more or less, standard game theoretic arguments. Standard arguments can be expressed in the logical structure discussed so far. However, the current structure is limited in that it cannot address arguments like DC2 of Section 2. They involve beliefs about the decision of the other player. To adequately describe them, we need to introduce epistemic conditions on B_i which come in Section 4.

3.3 Classical logic CL and its provability $\vdash_0$

As yet we have discussed basic notions such as formulae and semantical consequences. Here, we introduce a proof-theoretic system of classical logic CL. The proof-theoretic system formulates mathematical inferences as pure symbol manipulations, while the consequence relation $\vDash$ is formulated by considering meanings or possibilities of symbolic expressions. We denote this proof-theoretic system also by CL. We find out soon that the provability and consequence relations are intimately related.

Classical logic CL consists of five axiom schemata and three inference rules, which describe the possible ways of manipulating formulae of $\mathscr{P}^{\mathrm{n}}$. The notion of a proof will be defined by means of such components. More concretely, those five *axiom schemata* and three *inference rules* are as follows: for any formulae A, B, C and finite nonempty set Φ of formulae in $\mathscr{P}^{\mathrm{n}}$,

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: $\bigwedge \Phi \supset A$, where $A \in \Phi$;

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

$$\frac{A \supset B \quad A}{B} \text{ (Modus Ponens)}$$

$$\frac{\{A \supset B : B \in \Phi\}}{A \supset \bigwedge \Phi} \left(\bigwedge\text{-Rule}\right) \qquad \frac{\{A \supset B : A \in \Phi\}}{\bigvee \Phi \supset B} \left(\bigvee\text{-Rule}\right).$$

Modus Ponens is abbreviated as MP. These axioms and rules are schemata in the sense that formulae, A, B, C and the set Φ can be arbitrary. A particular formula or inference rule of them is called an *instance*, for example, $\mathbf{p}_0 \supset (\mathbf{p}_1 \supset \mathbf{p}_0)$ is an instance of L1.[10]

Let A be a formula in $\mathscr{P}^{\mathrm{n}}$ and Γ a subset of $\mathscr{P}^{\mathrm{n}}$. A *proof* of A from Γ in CL is a finite tree with the following properties:

(1): a formula in $\mathscr{P}^{\mathrm{n}}$ is associated with each node;

(2): the formula associated with each leaf is an instance of the above axioms or is a formula in Γ;

(3): adjoining nodes together with their associated formulae form an instance of the above three inference rules;

(4): A is associated with the root node.[11]

[10] There are many other formulations of classical propositional logic (see Mendelson [24], pp. 37–38). The present axiomatization is given in Kaneko and Nagashima [16].

[11] More explicitly, a proof is given as a triple $(X, \prec: \varphi)$, where $(X, \prec)$ is a finite tree in the sense of graph theory and φ is a function from X to $\mathscr{P}^{\mathrm{n}}$ associating formula $\varphi(x)$ with a each node $x \in X$. An ending node x is called a *leaf*, and the initial node is called the *root*.

We say that A is said to be *provable from* Γ in CL, denoted by $\Gamma \vdash_0 A$, iff there is a proof of A from Γ. When Γ is empty, we write simply $\vdash_0 A$. We write $\Gamma \nvdash_0 A$ iff not $\Gamma \vdash_0 A$.

The following lemma gives simple examples of provable formulae.

Lemma 3.1.

(1): $\vdash_0 A \supset A$;
(2): $\{A \supset B, B \supset C\} \vdash_0 A \supset C$.

Proof.
(1): The following is a proof tree of $A \supset A$.

$$\dfrac{\dfrac{\overset{\text{L2}}{[A \supset ((A \supset A) \supset A)] \supset [(A \supset (A \supset A)) \supset (A \supset A)]} \quad \overset{\text{L1}}{A \supset ((A \supset A) \supset A)}}{(A \supset (A \supset A)) \supset (A \supset A)}\,\text{MP} \qquad \overset{\text{L1}}{A \supset (A \supset A)}}{A \supset A}\,\text{MP}$$

(2): The following is a proof of $A \supset C$ from $\{A \supset B, B \supset C\}$.

$$\dfrac{\overset{\text{Ass.}}{A \supset B} \qquad \dfrac{\dfrac{\overset{\text{Ass.}}{B \supset C} \quad \overset{\text{L1}}{(B \supset C) \supset (A \supset (B \supset C))}}{A \supset (B \supset C)}\,\text{MP} \qquad \overset{\text{L2}}{[A \supset (B \supset C)] \supset [(A \supset B) \supset (A \supset C)]}}{(A \supset B) \supset (A \supset C)}\,\text{MP}}{A \supset C}\,\text{MP}$$

□

The next lemma will be used without mentioning.

Lemma 3.2.

(1): $\Gamma' \subseteq \Gamma$ and $\Gamma' \vdash_0 A$ imply $\Gamma \vdash_0 A$;
(2): if $\Gamma \vdash_0 B$ for all $B \in \Gamma'$ and $\Gamma' \vdash_0 A$, then $\Gamma \vdash_0 A$.

Here, we find the distinction A3 stated in Section 1. The claims of Lemma 3.1 are object theorems, while those of Lemma 3.2 are meta-theorems on object theorems. The latter are not formulated in CL but in metamathematics.

We would like to have a bridge between the classical semantics of Section 3.2 and the syntactical system CL. To have this connection, we need a key concept: We say that a set Γ of formulae in $\mathscr{P}^{\mathrm{n}}$ is *inconsistent* in CL iff $\Gamma \vdash_0 \neg C \wedge C$ for some C, and that Γ is *consistent* in CL iff it is not inconsistent in CL.

Theorem 3.3 (Soundness-completeness for classical logic CL). Let Γ be a set of formulae and A a formula. Then

(1): $\Gamma \vdash_0 A$ if and only if $\Gamma \vDash A$;

(2): there is a model κ of Γ if and only if Γ is consistent in CL.

Assertions (1) and (2) (with the quantifications of all Γ and A for each) are actually equivalent. The *only-if part* of each is called *soundness*, and the *if* part is *completeness*. The term "soundness" means that the syntactical formulation of symbolic inferences provides nothing other than the semantical validity, and "completeness" means that the former captures the latter. Thus, the two approaches are equivalent. In this sense, our description of "logic" is complete. This gives the bridge between model theory and proof theory stated as A2.

The above completeness differs from the "completeness" in the sense that $\Gamma \vdash_0 A$ or $\Gamma \vdash_0 \neg A$. The former is the completeness of a logic, but the latter is the completeness of nonlogical axioms Γ.[12]

Theorem 3.3 is quite standard, and the above syntactical system turns out to be equivalent to the other formulations of classical logic (see Mendelson [24] for other formulations of CL). Nevertheless, the above axiomatization is modified to facilitate game theoretic arguments, and no textbook for this version is found. Thus, we should state its proof. The soundness part can be proved without much difficulty, following a proof in a textbook, but the completeness part is quite complicated. We give a proof of completeness in the Appendix.

By Theorem 3.3, we can use any tautologies as provable formulae in CL. For example, $\neg A \vee A$, $\neg(\neg A \wedge A)$, $\neg \bigvee \Phi \equiv \bigwedge\{\neg A : A \in \Phi\}$, $\neg \bigwedge \Phi \equiv \bigvee\{\neg A : A \in \Phi\}$ and $(A \supset B) \equiv (\neg B \supset \neg A)$ are provable in CL. In the subsequent sections, we use such tautologies without mentioning.

It follows from the soundness part of Theorem 3.3.(2), by taking the empty Γ, that CL is contradiction-free.

Corollary 3.4. There is no formula A such that $\vdash_0 \neg A \wedge A$.

Let us return to the game example. Using Theorem 3.3, (3.2) and (3.3) can be written as

$$\hat{g}_1^1 \vdash_0 \ \mathrm{Dom}_1(\mathbf{s}_{12}) \text{ and } \hat{g}_1^1 \vdash_0 \neg\mathrm{Dom}_1(\mathbf{s}_{11}). \tag{3.4}$$

$$\hat{g}_1^1 \nvdash_0 \ \mathrm{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22}). \tag{3.5}$$

In fact, (3.4) can directly be proved in CL, which is easier than (3.2). On the other hand, it is difficult to obtain (3.5) directly in CL. For a direct proof of (3.5), we should show that there is no proof P of $\mathrm{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22})$ from $\hat{g}_1^1$. This is a difficult task, since there are an infinite number of candidate proofs. However, Theorem 3.3 enables us to show unprovability by constructing a countermodel.

Remark 3.5. When we adopt $\mathscr{P}$ rather than $\mathscr{P}^{\mathrm{n}}$, we need no essential modification of syntactical system CL; just we replace $\mathscr{P}^{\mathrm{n}}$ by $\mathscr{P}$. On the other hand, the classical semantics should be slightly modified: the domain of each assignment κ becomes $PV \cup \{\mathrm{B}_i(A) : A \in \mathscr{P} \text{ and } i \in N\}$ with the same image $\{\top, \bot\}$. Accordingly, C0 is replaced by

C0*: for any $C \in PV \cup \{\mathrm{B}_i(A) : A \in \mathscr{P} \text{ and } i \in N\}$, $V_\kappa(C) = \top$ if and only if $\kappa(C) = \top$.

Then all the other definitions are the same. Here, $\mathrm{B}_i(A)$ is treated in the same manner as a propositional variable. Theorem 3.3 holds with these modifications. In the subsequent sections, we use this modified CL with the set of formulae $\mathscr{P}$.

[12] The latter is the completeness of a theory in the sense of the logic literature.

The following tautologies will be used, often without mentioning: for any $A, B, C \in \mathscr{P}$ and any finite nonempty subset Φ of $\mathscr{P}$,

(1): $\vdash_0 \left(A \bigwedge B \supset C\right) \equiv (A \supset (B \supset C))$;

(2): $\vdash_0 (A \supset B) \bigwedge (B \supset C) \supset (A \supset C)$;

(3): $\vdash_0 \neg \bigwedge \Phi \equiv \bigvee \{\neg A : A \in \Phi\}$ and $\vdash_0 \neg \bigvee \Phi \equiv \bigwedge \{\neg A : A \in \Phi\}$;

(4): $\vdash_0 (A \supset B) \equiv (\neg B \supset \neg A)$.

In the classical logic CL of Remark 3.5, it turns out that $\nvdash_0 \mathrm{B}_i(A)$ for any $A \in \mathscr{P}$. Thus, we cannot discuss how a player arrives at his beliefs or his reasoning ability. On the other hand, game theory is particularly interested in the construction of beliefs and the reasoning ability of a player, which is the subject of the next section.

4 Various epistemic logics: proof-theoretic approach

In this section, we present various epistemic logics from the proof-theoretic point of view. There are many possible logical systems and their different formulations. We discuss some of them, following the standard logic literature. However, the basic principles for epistemic logics are not be clearly seen in the standard formulation. Thus, we discuss general ideas for beliefs. In doing so, the logical system $\mathrm{KD4}^n$ emerges as central in various systems. For future purposes, we present also the sequent formulation $\mathrm{KD4}^n$ in Gentzen-style. The reader may skip Sections 4.2 – 4.4 for reading the rest of the paper.

4.1 Standard axiomatizations of epistemic logics

In this subsection, we follow the standard axiomatizations of epistemic logics. We obtain various logical systems determined by combinations of axioms, which are treated in a somewhat parallel manner.

We consider the following list of axiom schemata and inference rule, for whose names we follow the literature of modal logic: for any $i \in N$, any $A, C \in \mathscr{P}$ and any finite nonempty subset Φ of $\mathscr{P}$:

K: $\mathrm{B}_i(A \supset C) \supset (\mathrm{B}_i(A) \supset \mathrm{B}_i(C))$;[13]

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

T: $\mathrm{B}_i(A) \supset A$ – – – – – – – – – – truthfulness;

4: $\mathrm{B}_i(A) \supset \mathrm{B}_i\mathrm{B}_i(A)$ – – – – – – positive introspection;

5: $\neg\mathrm{B}_i(A) \supset \mathrm{B}_i(\neg\mathrm{B}_i(A))$ – – – negative introspection;

and

[13] K and D come from “*K*ripke” and “*d*eontic logic”. See Chellas [6].

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

By Remark 3.5.(1), Axiom K is equivalent to $\mathrm{B}_i(A \supset C) \wedge \mathrm{B}_i(A) \supset \mathrm{B}_i(C)$. This is interpreted as meaning that player i can use Modus Ponens. Axiom D means that player i does not have a contradictory belief. Axiom T means that beliefs are true to the outside thinker. Axiom T implies Axiom D. Axiom 4 means that if player i believes A, then he also believes that he believes A. Axiom 5 means that if he does not believe A, he believes that he does not. These two axioms look parallel, but we will argue in Section 4.2 that they differ substantively. The Necessitation rule states that if A is already proved, then $\mathrm{B}_i(A)$ is also provable. Some examples of proofs are given below.

We abbreviate the Necessitation rule as Nec. It is important to notice that Nec may be repeatedly applied even with different players, e.g.,

$$\cfrac{\cfrac{\cfrac{A}{\mathrm{B}_i(A)}}{\mathrm{B}_j\mathrm{B}_i(A)}}{\cdots}$$

Thus, once A is provable, then it becomes common knowledge *in effect*. Note that Nec differs from $A \supset \mathrm{B}_i(A)$ as an axiom. This difference will be clearer when we introduce nonlogical axioms. The treatment of nonlogical axioms differs from that in classical logic CL, which will be explained presently.

The most basic system is defined as

$\mathbf{K}^n$: CL + K + Nec (within $\mathscr{P}$).

A *proof P* in epistemic logic K^n is defined in the same manner as in CL except: (1) instances of K and Nec are allowed and (2) nonlogical axioms are not allowed in the proof.

Consider player 1's inference in the game g^1 of Table 1. Diagram 1 gives a proof of $\mathrm{B}_1(\bigwedge \hat{g}^1_1) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$ in K^n, where the uppermost formulae are instances of L4, the uppermost inference is $\bigwedge$-Rule, the second is Nec, the right-hand formula of the third line is an instance of Axiom K, and the last inference is MP. Note that $\bigwedge$-Rule has $|S| = 4$ upper formulae.

$$\cfrac{\cfrac{\cfrac{\{\bigwedge \hat{g}^1_1 \supset P_1(\mathbf{s}_{12}, s_2 : s_1, s_2) : (s_1, s_2) \in S\}}{\bigwedge \hat{g}^1_1 \supset \mathrm{Dom}_1(\mathbf{s}_{12})}}{\mathrm{B}_1\left(\bigwedge \hat{g}^1_1 \supset \mathrm{Dom}_1(\mathbf{s}_{12})\right)} \qquad \mathrm{B}_1\left(\bigwedge \hat{g}^1_1 \supset \mathrm{Dom}_1(\mathbf{s}_{12})\right) \supset (\mathrm{B}_1\left(\bigwedge \hat{g}^2_1\right) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12})))}{\mathrm{B}_1\left(\bigwedge \hat{g}^1_1\right) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12})).}$$

Diagram 1

That is, if 1 believes that his payoff function is $\bigwedge \hat{g}^1_1$, then he infers the belief that $\mathbf{s}_{12}$ is a dominant strategy. There is also a proof of $\mathrm{B}_1(\bigwedge \hat{g}^1_1) \supset \mathrm{B}_1(\neg\mathrm{Dom}_1(\mathbf{s}_{11}))$, which is derived from $\bigwedge \hat{g}^1_1 \supset \neg\mathrm{Dom}_1(\mathbf{s}_{11})$ in the same manner, but $\bigwedge \hat{g}^1_1 \supset \neg\mathrm{Dom}_1(\mathbf{s}_{11})$ needs a bit longer proof.

Various epistemic logics can be defined based on K^n by choices of some of the above axiom schemata. In this paper, we consider the following list of logics:

$\mathbf{K}^n$: CL + Nec + K; $\mathbf{KD}^n$: K^n + D; $\mathbf{KT}^n$: K^n + T;

$\mathbf{KD4}^n$: KD^n + 4; $\mathbf{S4}^n$: $\mathrm{K4}^n$ + T;

$\mathbf{KD45}^n$: $\mathrm{KD4}^n$ + 5; $\mathbf{S5}^n$: $\mathrm{S4}^n$ + 5.

Diagram 2

We will argue that $\mathrm{KD4}^n$ and KD^n take central positions in this list of logics. It will be discussed in Section 6 that a logic including Axiom T, say $\mathrm{S4}^n$, can be discussed in one without it.

Let $\mathscr{S}$ be a logic in the above list. A proof in $\mathscr{S}$ is defined in a similar manner to that in K^n. Let A be a formula in $\mathscr{P}$. We write $\vdash_{\mathscr{S}} A$ iff there is a proof of A in $\mathscr{S}$. The following is a simple observation, which will be used without mentioning: for any $A \in \mathscr{P}$,

$$\vdash_0 A \text{ implies } \vdash_{\mathscr{S}} A\,. \tag{4.1}$$

Note that the provability of $A \in \mathscr{P}$ in CL is mentioned in Remark 3.5.

In the above definition of a proof, we do not allow nonlogical axioms as initial formulae in proofs. To describe game theoretic assumptions, we introduce nonlogical axioms in a way different from in Section 3.3. Let Γ be a subset of $\mathscr{P}$ and $A \in \mathscr{P}$. We define $\Gamma \vdash_{\mathscr{S}} A$ iff $\vdash_{\mathscr{S}} A$ or $\vdash_{\mathscr{S}} \bigwedge \Phi \supset A$ for some finite nonempty subset Φ of Γ. For the reason to avoid a nonlogical axiom in a proof, see the remark about the Necessitation rule in Section 4.3.

In classical logic CL, for a nonempty finite set Γ of formulae, $\Gamma \vDash A$ is equivalent to $\vDash \bigwedge \Gamma \supset A$, which together with the soundness-completeness for CL implies $\Gamma \vdash_0 A$ is equivalent to $\vdash_0 \bigwedge \Gamma \supset A$. Hence it follows from (4.1) that

$$\Gamma \vdash_0 A \text{ implies } \Gamma \vdash_{\mathscr{S}} A. \tag{4.2}$$

The strengths of the provabilities of the above logics are described as follows:

$$\begin{array}{ccccc}
\mathrm{K}^n & \rightarrow & \mathrm{KD}^n & \rightarrow & \mathrm{KT}^n \\
 & & \downarrow & & \downarrow \\
 & & \mathrm{KD4}^n & \rightarrow & \mathrm{S4}^n \\
 & & \downarrow & & \downarrow \\
 & & \mathrm{KD45}^n & \rightarrow & \mathrm{S5}^n.
\end{array}$$

Diagram 3

where the expression, $\mathscr{S} \rightarrow \mathscr{S}'$, means that the provability of $\mathscr{S}'$ is stronger than that of $\mathscr{S}$, for example, $\vdash_{\mathrm{KD4}^n} A$ implies $\vdash_{\mathrm{S4}^n} A$. Diagram 1 is a legitimate proof in all $\mathscr{S}$'s.

The following are basic properties of the epistemic logics of the above list.

Lemma 4.1. For any $A, C \in \mathscr{P}$ and a nonempty finite subset Φ of $\mathscr{P}$,

(1): $\vdash_{\mathscr{S}} \mathrm{B}_i(A \supset C) \wedge \mathrm{B}_i(A) \supset \mathrm{B}_i(C)$;

(2): $\vdash_{\mathscr{S}} \bigvee \mathrm{B}_i(\Phi) \supset \mathrm{B}_i(\bigvee \Phi)$, where $\mathrm{B}_i(\Phi) = \{\mathrm{B}_i(A) : A \in \Phi\}$;

(3): $\vdash_{\mathscr{S}} \mathrm{B}_i(\bigwedge \Phi) \equiv \bigwedge \mathrm{B}_i(\Phi)$;

(4): $\vdash_{\mathscr{S}} \mathrm{B}_i(\neg A) \supset \neg \mathrm{B}_i(A)$, where $\mathscr{S}$ includes Axiom D;

(5): if $\Gamma \vdash_{\mathscr{S}} A$, then $\mathrm{B}_i(\Gamma) \vdash_{\mathscr{S}} \mathrm{B}_i(A)$.

Proof. We prove (1) – (4).

(1): This follows from Axiom K and Remark 3.5.(1).

(2): Let A be an arbitrary formula in Φ. Since $\vdash_{\mathscr{S}} A \supset \bigvee \Phi$ by L5, we have $\vdash_{\mathscr{S}} \mathrm{B}_i(A \supset \bigvee \Phi)$ by Nec. By K, $\vdash_{\mathscr{S}} \mathrm{B}_i(A) \supset \mathrm{B}_i(\bigvee \Phi)$. Since this holds for any $A \in \Phi$, we have $\vdash_{\mathscr{S}} \bigvee \mathrm{B}_i(\Phi) \supset \mathrm{B}_i(\bigvee \Phi)$ by $\bigvee$-Rule

(3): $\vdash_{\mathscr{S}} \mathrm{B}_i(\bigwedge \Phi) \supset \bigwedge \mathrm{B}_i(\Phi)$ is the dual of (2). We prove the converse only for $\Phi = \{A, C\}$. Since $\vdash_{\mathscr{S}} A \supset (C \supset A \wedge C)$, we have $\vdash_{\mathscr{S}} \mathrm{B}_i(A) \supset (\mathrm{B}_i(C) \supset \mathrm{B}_i(A \wedge C))$, using Nec, K and MP a few times. This is equivalent to $\vdash_{\mathscr{S}} \mathrm{B}_i(A) \wedge \mathrm{B}_i(C) \supset \mathrm{B}_i(A \wedge C)$. When Φ has more than two formulae, we should prove $\vdash_{\mathscr{S}} \bigwedge \mathrm{B}_i(\Phi) \supset \mathrm{B}_i(\bigwedge \Phi)$ by induction on the number of formulae in Φ.

(4): By (3), $\vdash_{\mathscr{S}} \neg \mathrm{B}_i(\neg A \wedge A) \equiv \neg(\mathrm{B}_i(\neg A) \wedge \mathrm{B}_i(A))$. By Axiom D, we have $\vdash_{\mathscr{S}} \neg(\mathrm{B}_i(\neg A) \wedge \mathrm{B}_i(A))$. This is equivalent to $\vdash_{\mathscr{S}} \mathrm{B}_i(\neg A) \supset \neg \mathrm{B}_i(A)$. □

We will use the following facts extensively without mentioning.

Lemma 4.2. For any $A, B, C \in \mathscr{P}$ and subsets Γ, Φ of $\mathscr{P}$,

(1): $\Gamma \vdash_{\mathscr{S}} A \supset B$ and $\Gamma \vdash_{\mathscr{S}} B \supset C$ imply $\Gamma \vdash_{\mathscr{S}} A \supset C$;

(2): $\Gamma \vdash_{\mathscr{S}} A$ for all $A \in \Phi$ if and only if $\Gamma \vdash_{\mathscr{S}} \bigwedge \Phi$, where Φ is a nonempty finite set.

Proof. (1) follows from Remark 3.5.(2). The *if* part of (2) follows from L4 and MP. The converse is proved by using $\bigwedge$-Rule and MP (L1 in the case of $\Gamma = \emptyset$). □

Let us see how the decision criteria DC1 and DC2 of Section 2 are expressed in epistemic logic $\mathscr{S}$.

Since $\vdash_{\mathscr{S}} \mathrm{B}_1(\bigwedge \hat{g}_1^1) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$ by Diagram 1, we have, by Lemma 4.1.(3), $\vdash_{\mathscr{S}} \bigwedge \mathrm{B}_1(\hat{g}_1^1) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$. Following our convention of nonlogical axioms in epistemic logic $\mathscr{S}$, we have

$$\mathrm{B}_1(\hat{g}_1^1) \vdash_{\mathscr{S}} \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12})). \tag{4.3}$$

Similarly, $\mathrm{B}_1(\hat{g}_1^1) \vdash_{\mathscr{S}} \mathrm{B}_1(\neg \mathrm{Dom}_1(\mathbf{s}_{11}))$. In logic $\mathscr{S}$ including Axiom D, we have, by Lemma 4.1.(4),

$$\mathrm{B}_1(\hat{g}_1^1) \vdash_{\mathscr{S}} \neg \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{11})). \tag{4.4}$$

Hence 1's belief on his own payoff function is enough to decide $\mathbf{s}_{12}$ to be a unique dominant strategy.

The counterpart of (3.5) in $\mathscr{S}$ is expected to be:

$$\mathrm{B}_1(\hat{g}_1^1) \nvdash_{\mathscr{S}} \mathrm{B}_1(\mathrm{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22})). \tag{4.5}$$

In fact, it would be difficult to prove this kind of unprovability within $\mathscr{S}$. One way of showing such an *unprovability* is to use semantics, which is discussed in Section 5.2.

As will be seen in Section 7.1, we formulate criterion DC2 as:

$$\hat{D}_{22}(s_2) = \mathrm{B}_2\left(\bigvee_{t_1}\mathrm{B}_1(\mathrm{Dom}_1(t_1))\right) \wedge \bigwedge_{t_1}\mathrm{B}_2\left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right).$$

That is, player 2 believes that 1 has a dominant strategy and that 2's decision is a best strategy to 1's dominant strategies. Consider the decision making of player 2 with this criterion in game g^2. Suppose that $\mathscr{S}$ includes Axiom D. Since $g_1^2 = g_1^1$, we have $\mathrm{B}_1(\hat{g}_1^2) \vdash_{\mathscr{S}} \bigvee_{t_1}\mathrm{B}_1(\mathrm{Dom}_1(t_1))$ by (4.3), and thus $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1^2) \vdash_{\mathscr{S}} \mathrm{B}_2(\bigvee_{t_1}\mathrm{B}_1(\mathrm{Dom}_1(t_1)))$ by Lemma 4.1.(5). Also, we have $\hat{g}_2^2 \vdash_{\mathscr{S}} \mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{12})$, and so $\hat{g}_2^2 \vdash_{\mathscr{S}} \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12})) \supset \mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{12})$. Hence $\mathrm{B}_2(\hat{g}_2^2) \vdash_{\mathscr{S}} \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12})) \supset \mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{12}))$ by Lemma 4.1.(5). It follows from (4.4) that $\mathrm{B}_1(\hat{g}_1^2) \vdash_{\mathscr{S}} \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{11})) \supset \mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{11})$, and thus $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1^2) \vdash_{\mathscr{S}} \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{11})) \supset \mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{11}))$. Combining all the results, we have

$$\mathrm{B}_2\mathrm{B}_1(\hat{g}_1^2), \mathrm{B}_2(\hat{g}_2^2) \vdash_{\mathscr{S}} \hat{D}_{22}(\mathbf{s}_{22}), \tag{4.6}$$

where we abbreviate $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1^2) \cup \mathrm{B}_2(\hat{g}_2^2)$ as $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1^2), \mathrm{B}_2(\hat{g}_2^2)$. That is, player 2 would choose $\mathbf{s}_{22}$ as a recommended strategy by DC2 under the beliefs $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1^2)$, $\mathrm{B}_2(\hat{g}_2^2)$.

4.2 Basic principles for beliefs

Here, we state the general principle for a belief about A in order to have a clearer view on what we would like to express by a belief about A. The reader may skip this subsection to go to the subsequent sections.

The general idea for the notion of "a belief about A" is stated as:

G: player i believes A if and only if he has an argument for A from his basic beliefs.

From the proof-theoretic point of view, we formulate "having an argument for A" as "having a proof of A." There are still various options for a formulation of "having a proof". Since our investigator (observer) is assumed to have the reasoning ability described by classical logic CL, we take the following:

G1: Player i has the reasoning ability described by classical logic CL.

That is, player i has at least the same reasoning *ability* as the investigator's. Hence, player i can infer what the investigator can infer. This does not imply that the players and investigator share the same basic beliefs. Since players' beliefs and reasoning abilities are described inside the investigator's logical system, some descriptions are made purely from the investigator's viewpoint and may not be shared with players, which will be seen below.

Besides G1, we add other three components to players' beliefs and reasoning abilities: (1) Basic beliefs taken as given; (2) Intrapersonal introspective abilities; and (3) Interpersonal introspective abilities. Basic beliefs of player i are given as nonlogical axioms in the terminology in Section 3, and their emergence is not discussed in this paper.[14] Regarding (2) and (3), intrapersonal introspection may be still regarded as a problem of individual inferences, but interpersonal ones are largely hypothetical. For these problems, there are a great spectrum of options. Here, we adopt simple and clear-cut options for these.

For the intrapersonal introspection ability, we assume:

G2: Player i has the introspection ability on his own abilities described by G1 and G2.

This requires player i only to be conscious of "having a proof". This consciousness of "having a proof" is now regarded as "having a proof of "having a proof"". Therefore, we regard G2 as forming part of basic principle G. However, some subtlety is involved in G2: G2 itself is involved in G2. To see the feasibility of G1–G2, we need an explicit mathematical formulation of them.

The nature of beliefs about other players and/or their beliefs considerably differs from that of beliefs about himself and/or his own. Beliefs about other players are based on projecting and are hypothetical constructs. There are many options for this problem. Here we consider only the following.

G3: When player i thinks about other players' beliefs, player i assume G1, G2 as well as G3 for the other players in a symmetric manner.

This means that a player imagined in the mind of a player (imagined in the mind of another player ...) follows G3. Thus, the situation is very complicated, but, we obtain the symmetric interpersonal beliefs by assuming the limit case of complications. Note that G3 appears in G3 itself, again.

To materialize our basic principle G, we need specific assumptions on the above three steps. We should keep the *remark* in mind: Principle G1 is the most basic, G2 is still basic, but G3 is one possible and convenient choice for our research strategy. More restrictive possibilities of G2 and G3 are discussed in Kaneko and Suzuki [19].

To formulate G1, it suffices to assume the beliefs about the instances of Logical Axioms L1–L5, and the axioms expressing the reasoning ability of player i corresponding to inference rules MP, $\bigwedge$-Rule and $\bigvee$-Rule:

B-L: $\mathrm{B}_i(A)$, where A is an instance of L1–L5;

B-MP: $\mathrm{B}_i(A \supset C) \wedge \mathrm{B}_i(A) \supset \mathrm{B}_i(C)$;

B-$\bigwedge$: $\bigwedge\{\mathrm{B}_i(A \supset C) : C \in \Phi\} \supset \mathrm{B}_i(A \supset \bigwedge \Phi)$;

B-$\bigvee$: $\bigwedge\{\mathrm{B}_i(C \supset A) : C \in \Phi\} \supset \mathrm{B}_i(\bigvee \Phi \supset A)$,

[14] The literature of belief revisions is related to the development of basic beliefs. For this literature, see Schulte [33] in this issue. Also, the development of basic beliefs is considered from the viewpoint of experiences in Kaneko and Matsui [15].

where A, C in B-MP, B-$\bigwedge$ and B-$\bigvee$ are any formulae and Φ is any finite nonempty set of formulae in $\mathscr{P}$. Note that B-MP is equivalent to Axiom K. We denote the set of all instances of B-L – B-$\bigvee$ by Δ_i^0. The following holds:

Theorem 4.3 (Classical reasoning ability). Let Γ be a set of formulae in $\mathscr{P}$ and A a formula in $\mathscr{P}$. Then $\Gamma \vdash_0 A$ if and only if $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \vdash_0 \mathrm{B}_i(A)$.

Sketch of a Proof. Consider the *only-if* part. Let P be a proof of A from Γ. We prove $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \vdash_0 \mathrm{B}_i(C)$ for any C occurring in P by induction on P from its leaves. It suffices to show that for any initial formula C in P, $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \vdash_0 \mathrm{B}_i(C)$, and that the probability relation $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \vdash_0$ goes down from the upper formulae of each inference rule to the lower formula.

The *if* part needs some new concept: We define the eraser ε_i of B_i: This goes into a formula and erases the outer B_i *only once* when ε_i meets B_i. Let P be a proof of $\mathrm{B}_i(A)$ from $\mathrm{B}_i(\Gamma) \cup \Delta_i^0$. Then it suffices to show by induction on the tree structure of P from its leaves that for any formula C in P, $\Gamma \vdash_0 \varepsilon_i C$.□

That is, A is provable from Γ in classical logic CL if and only if player i can derive A from his basic beliefs Γ with his reasoning ability described by Δ_i^0. Player i has given the same (potential) reasoning ability as the investigator's. Thus, we have succeeded in formulating G1 explicitly by this method.

For the connection of player i's inner logic to the investigator's, we add

B-D: $\neg \mathrm{B}_i(\neg A \wedge A)$.

This is Axiom D. It holds with B-D that Γ is consistent if and only if $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \cup$B-D is consistent, where B-D is now regarded as the set of instances of B-D. In this sense, Axiom B-D connects player i's inner logic with the investigator's logic up to their consistencies. Note that this axiom also enables us to show that $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \cup$B-D $\vdash_0 \mathrm{B}_i(\neg A)$ implies $\mathrm{B}_i(\Gamma) \cup \Delta_i^0 \cup$B-D $\vdash_0 \neg \mathrm{B}_i(A)$.

We formulate G2 by the following two axioms:

B-4: $\mathrm{B}_i(A) \supset \mathrm{B}_i \mathrm{B}_i(A)$;

B-I: $\mathrm{B}_i(A)$, where A is an instance of B-L, B-MP, B-$\wedge$, B-$\vee$, B-D and B-4.

Axiom B-4 is Axiom 4, and states that if player i believes A, then he believes that he believes A. Recalling basic principle G, this is described as that i is conscious of "having an argument for A from his basic beliefs". Axiom B-I states that i is conscious of the reasoning and introspective abilities described by B-L – B-4. We denote, by Δ_i, the set obtained from $\Delta_i^0 \cup$B-D by adding all the instances of B-4 and B-I. The assumption set Δ_i is the formulation of principles G1 and G2. This is essentially what an individual player is given. In the single player case, we have the following, whose proof is found in Kaneko and Nagashima [16].

Theorem 4.4.(Reasoning ability described by G1 and G2). Let $n = 1$. For any $A \in \mathscr{P}$, $\Delta_1 \vdash_0 A$ if and only if $\vdash_{\mathrm{KD4}^1} A$.

Thus, $\mathrm{KD4}^1$ corresponds to the logic describing basic principles G1 and G2.[15]

[15] From Theorem 4.3 and Theorem 4.4, we can regard $\mathrm{KD4}^1$ as a logic of provability and introspection. However, this differs slightly from the logic so called *provability logic* in the literature (see Boolos [5]).

When $n \geq 2$, the set of axioms $\Delta_1 \cup ... \cup \Delta_n$ does not yet describe the interpersonal beliefs of players described in G3. Principle G3 requires interpersonal assumptions on reasoning and introspective abilities, e.g., player i believes that player j has the abilities described by Δ_j, etc. The entire assumption set is given as

$$\Delta^* = \{\mathrm{B}_{i_m}...\mathrm{B}_{i_1}(A) : A \in \Delta_1 \cup ... \cup \Delta_n,\ i_m, ..., i_1 \in N \text{ and } m \geq 0\}, \qquad (4.7)$$

where $\mathrm{B}_{i_m}...\mathrm{B}_{i_1}(A)$ is A itself if $m = 0$. This set states that the reasoning and introspection abilities of players described by $\Delta_1 \cup ... \cup \Delta_n$ are common knowledge. We show that this gives the same provability as epistemic logic $\mathrm{KD4}^n$. Its proof is also found in Kaneko and Nagashima [16].

Theorem 4.5.(Reasoning abilities described by G1, G2 and G3). Let $n \geq 2$. Then for any $A \in \mathscr{P}$, $\Delta^* \vdash_0 A$ if and only if $\vdash_{\mathrm{KD4}^n} A$.

Theorem 4.5 helps us understand the basic principle for our epistemic logics, especially, $\mathrm{KD4}^n$, but, it does not help us evaluate the axioms themselves. On the one hand, Theorem 4.5 would hold even if we add each or both of Axioms T and 5 to both sides. On the other hand, if we delete Axiom 4 from both sides, then the left-hand side needs to be modified for the above equivalence. In this case, G2 has a slightly different content. To evaluate various axioms, we should return to the original principle G.

4.3 Evaluations of epistemic axioms

Let us return to the formulations of epistemic logics in Section 4. Our purpose is to consider epistemic aspects of decision making in game situations. For this, it does not suffice to consider only mathematical properties of such logics. We would like to choose some logics as more appropriate than others. We adopt the Inference Rule Nec, Axioms K and D as very basic. We reject Axiom 5 as inappropriate: this rejection is made by recalling our basic principle G. We avoid Axiom T to allow false beliefs, but can treat Axiom 4 inside our logics.

Here, we give remarks on Nec and Axioms D, T, 4, 5.

Necessitation. When nonlogical axioms are involved and $\bigwedge \Phi \supset A$ is provable, Nec is applied to the whole formula $\bigwedge \Phi \supset A$ and yields $\mathrm{B}_i(\bigwedge \Phi \supset A)$. Since Nec can be applied arbitrarily many times, $\bigwedge \Phi \supset A$ becomes common knowledge in effect in the sense that $\mathrm{B}_{i_m}...\mathrm{B}_{i_1}(\bigwedge \Phi \supset A)$ are all derived. Nevertheless, this does not means that the assumption $\bigwedge \Phi$ becomes common knowledge, but that only the implication $\bigwedge \Phi \supset A$ becomes common knowledge. This note is related to the reason for the introduction of nonlogical axioms Φ in the present form. On the other hand, if $\bigwedge \Phi$ is assumed as an initial formula in a proof, then $\bigwedge \Phi$ becomes common knowledge. To avoid this, we have introduced nonlogical axioms as the antecedent of $\bigwedge \Phi \supset A$.

Axiom K changes $\mathrm{B}_i(\bigwedge \Phi \supset A)$ into $\mathrm{B}_i(\bigwedge \Phi) \supset \mathrm{B}_i(A)$. The former states that i believes $\bigwedge \Phi \supset A$, while the latter states that if i believes $\bigwedge \Phi$, he believes

A, too. Thus, Axiom K transforms a statement from i's viewpoint into one from the investigator's.

The difference between Nec and $A \supset B_i(A)$ as a logical axiom may be now clear. If $A \supset B_i(A)$ is assumed as a logical axiom, then from $\vdash_{\mathscr{S}} \bigwedge \Phi \supset A$, we obtain $\vdash_{\mathscr{S}} \bigwedge \Phi \supset B_i(A)$, and similarly $\vdash_{\mathscr{S}} \bigwedge \Phi \supset B_{i_m}...B_{i_1}(A)$ for all $i_m, \ldots, i_1 \in N$. Hence, any logical consequence from $\bigwedge \Phi$ could be common knowledge. Without $A \supset B_i(A)$ as an logical axiom, however, we can obtain only $\vdash_{\mathscr{S}} B_{i_m}...B_{i_1}(\bigwedge \Phi) \supset B_{i_m}...B_{i_1}(A)$. Hence

$$\{B_{i_m}...B_{i_1}(C) : C \in \Phi, i_m, ..., i_1 \in N \text{ and } m \geq 0\} \vdash_{\mathscr{S}} B_{j_k}...B_{j_1}(A)$$

for any $j_k, ..., j_1 \in N$. Thus, if $\bigwedge \Phi$ is common knowledge, then any consequence from Φ is common knowledge.

Axiom D: The basic principle G states that player i should not have a proof of a contradiction. If his beliefs are contradictory, then he would have a proof of a contradiction. Thus, Axiom D excludes contradictory basic beliefs.

Axiom T: This distinguishes knowledge from beliefs. With this axiom, beliefs are always true relative to the thinker (ultimately to the investigator), and without it, beliefs may be false. It is important to discuss the truth and/or falsity of beliefs in the future studies of economics and game theory. Axiom T prohibits the possibility of talking about the falsity of beliefs. In Section 6, we argue that Axiom T can be captured in an epistemic logic $\mathscr{S}$ without it.

Axiom 4: Although we have adopted this axiom to describe a part of G2, we do not think that this is so basic as Nec, K and D. It will be used once in game theoretic arguments in Section 7, but is avoidable with a slightly longer argument. The reasons for this reservation are: First, if we want to examine the role of self-consciousness, we should do it in a logic without Axiom 4. Also, a logic without Axiom 4 would be easier to handle in meta-theoretic respects. So far, we do not have enough developments in theory and applications to give clear distinctions between logics with and without Axiom 4.[16]

Axiom 5: We do not take this as a basic axiom. Axiom 5 is equivalent to $B_i(A) \vee B_i(\neg B_i(A))$ for any A, which is easier to be evaluated. According to the basic principle G, this states: for any A,

(*): player i has either a proof of A or a proof that he has no proof of A from his basic beliefs.

That is, when he has no proof objectively, he has also a proof of "no proof of A". This does not allow the *third* possibility that there is no proof objectively but player i does not notice it. Unless his basic beliefs are very rich, we expect

[16] We treated intrapersonal and interpersonal introspections separately in Section 4.2. In game theoretical practices, assumptions of interpersonal beliefs play significant roles but not much intrapersonal introspection. In the human history, however, self-consciousness might be evolved as the ability to derive interpersonal beliefs (Mithen [26], pp. 217–219). This may give a hint to reconsider the role of Axiom 4 and/or the principle G2.

that this third possibility could be the case. We would like to allow the third possibility as natural. Thus, we exclude Axiom 5 from the basic axioms.

Finally, we note that if Axiom T is additionally assumed, then $\neg \mathrm{B}_i(\neg \mathrm{B}_i(A))$ and $\mathrm{B}_i(A)$ are equivalent.

4.4 Gentzen-style formulation of KD4^n

This subsection is written for the reader who wants to go further to advances in the logic approach. We give the sequent calculus formulation of KD4^n in the Gentzen-style. Some other papers in this issue adopt this style. Although it is deductively equivalent to the Hilbert-style formulation, it adds another sense of "logical reality". The following is a very brief introduction. The interested reader may consult some textbooks such as Kleene [21] and Takeuti [34] (in the case of classical logic). Nevertheless, the best introduction may be still Gentzen's [11] original article.

We introduce the concept of a sequent. Let Γ, Θ be finite (possibly empty) subsets of $\mathscr{P}$. Using auxiliary symbols [,], and $\rightarrow$, we introduce a new expression $\Gamma \rightarrow \Theta$, which we call a *sequent*. We abbreviate $\Gamma \cup \Delta \rightarrow \Lambda \cup \Theta$ and $\{A\} \cup \Gamma \rightarrow \Theta \cup \{C\}$ as $\Gamma, \Delta \rightarrow \Lambda, \Theta$ and $A, \Gamma \rightarrow \Theta, C$, etc.

A sequent $\Gamma \rightarrow \Theta$ is associated with each node in a proof. Here, Γ is a set of nonlogical axioms. Thus, a set of nonlogical axioms appears in every step in a proof. The counterpart of $\Gamma \rightarrow \Theta$ in the Hilbert-style formulation is $\bigwedge \Gamma \supset \bigvee \Theta$, where $\bigwedge \emptyset$ and $\bigvee \emptyset$ are meant to be $\neg p \vee p$ and $\neg p \wedge p$, respectively. For a moment, the reader may interpret $\Gamma \rightarrow \Theta$ as $\Gamma \vdash_{\mathrm{KD4}^n} \bigvee \Theta$ in the previous formulation. We will explain presently the relationship of the present formulation with the Hilbert-style epistemic logic KD4^n.

In the following, $\Gamma, \Theta, \Delta, \Lambda, \Phi$ are finite sets of formulae, A, B formulae and Φ is also assumed to be nonempty.

Axiom (Initial sequent): $A \rightarrow A$,

Structural rules:

$$\frac{\Gamma \rightarrow \Theta}{\Delta, \Gamma \rightarrow \Theta, \Lambda}\ (\mathrm{Th}) \qquad\qquad \frac{\Gamma \rightarrow \Theta, A \qquad A, \Delta \rightarrow \Lambda}{\Gamma, \Delta \rightarrow \Theta, \Lambda}\ (\mathrm{Cut})\ .$$

Operational rules:

$$\frac{\Gamma \rightarrow \Theta, A}{\neg A, \Gamma \rightarrow \Theta}\ (\neg \rightarrow) \qquad\qquad \frac{A, \Gamma \rightarrow \Theta}{\Gamma \rightarrow \Theta, \neg A}\ (\rightarrow \neg)$$

$$\frac{\Gamma \rightarrow \Theta, A \qquad B, \Gamma \rightarrow \Theta}{A \supset B, \Gamma \rightarrow \Theta}\ (\supset \rightarrow) \qquad\qquad \frac{A, \Gamma \rightarrow \Theta, B}{\Gamma \rightarrow \Theta, A \supset B}\ (\rightarrow \supset)$$

$$\frac{A, \Gamma \rightarrow \Theta}{\bigwedge \Phi, \Gamma \rightarrow \Theta}\ \left(\bigwedge \rightarrow\right) \text{ where } A \in \Phi \qquad\qquad \frac{\{\Gamma \rightarrow \Theta, A : A \in \Phi\}}{\Gamma \rightarrow \Theta, \bigwedge \Phi}\ (\rightarrow \bigwedge)$$

$$\frac{\{A, \Gamma \to \Theta : A \in \Phi\}}{\bigvee \Phi, \Gamma \to \Theta} \ (\bigvee \to) \qquad \frac{\Gamma \to \Theta, A}{\Gamma \to \Theta, \bigvee \Phi} \ (\to \bigvee), \text{ where } A \in \Phi \ .$$

Epistemic Rule (Necessitation rule):

$$\frac{\Gamma, \mathrm{B}_i(\Delta) \to \Theta}{\mathrm{B}_i(\Gamma), \mathrm{B}_i(\Delta) \to \mathrm{B}_i(\Theta)} \ (\mathrm{B}_i \to \mathrm{B}_i), \text{ where } |\Theta| \leq 1 \text{ and } i \in N,$$

where $|\Theta|$ is the cardinality of Θ.

A *proof* P of $\Gamma \to \Theta$ in the present system is defined to be a tree in the same manner as in the previous Hilbert-style. That is, a sequent is associated with each node of P, the sequent associated with each leaf of P is an instance of the axiom sequent, some instances of the inference rules connect nodes of P, and $\Gamma \to \Theta$ is associated with the root of P. We say that $\Gamma \to \Theta$ is *provable in* KD4^n, denoted by $\vdash_{\mathrm{KD4}^n} \Gamma \to \Theta$, iff there is a proof P of $\Gamma \to \Theta$.

We may regard sequent $\mathrm{B}_1(\hat{g}_1^1) \to \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$ as a counterpart of $\mathrm{B}_1(\bigwedge \hat{g}_1^1) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$ of (4.4), which is proved as follows:

$$\frac{\left\{ \dfrac{P_1(\mathbf{s}_{12}, s_2 : s_1, s_2) \to P_1(\mathbf{s}_{12}, s_2 : s_1, s_2)}{\hat{g}_1^1 \to P_1(\mathbf{s}_{12}, s_2 : s_1, s_2)} \ (\mathrm{Th}) \right\}_{(s_1, s_2)}}{\dfrac{\hat{g}_1^1 \to \mathrm{Dom}_1(\mathbf{s}_{12})}{\mathrm{B}_1(\hat{g}_1^1) \to \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))} \ (\mathrm{B}_1 \to \mathrm{B}_1)} \ (\to \bigwedge).$$

From the last sequent, we can derive sequent $\to \mathrm{B}_1(\bigwedge \hat{g}_1^1) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$, which is also regarded as a counterpart of $\mathrm{B}_1(\bigwedge \hat{g}_1^1) \supset \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$.

The relation between the Gentzen-style of and Hilbert-style of KD4^n is as follows.

Theorem 4.6 (Relation to KD4^n in the Hilbert-style). Let Γ and Θ be finite sets of formulae. Then $\vdash_{\mathrm{KD4}^n} \Gamma \to \Theta$ if and only if $\vdash_{\mathrm{KD4}^n} \bigwedge \Gamma \supset \bigvee \Theta$. Recall that $\bigvee \emptyset$ and $\bigwedge \emptyset$ are $\neg p \wedge p$ and $\neg p \vee p$.

Thus, $\vdash_{\mathrm{KD4}^n} \Gamma \to \Theta$ corresponds to $\Gamma \vdash_{\mathrm{KD4}^n} \bigvee \Theta$ in the previous formulation.

Note that $\vdash_{\mathrm{KD4}^n} \Gamma \to$ is equivalent to $\vdash_{\mathrm{KD4}^n} \Gamma \to \neg p \wedge p$. This is proved by (Cut) and the fact that $\vdash_{\mathrm{KD4}^n} \neg p \wedge p \to$.

The following *cut-elimination theorem* is the main theorem for the Gentzen-style formulation KD4^n. It makes the system meta-theoretically different from the Hilbert-style formulation of KD4^n.

Theorem 4.7 (Cut-elimination). If $\vdash_{\mathrm{KD4}^n} \Gamma \to \Theta$, then there is a cut-free proof P of $\Gamma \to \Theta$.

The cut-elimination theorem was first proved for classical logic by Gentzen [11], and then it was proved for many other systems. The above one is a variant of the cut-elimination theorem for S4^1 and some others given by Ohnishi-Matsumoto [30]. One remark is that it is not successful to have cut-elimination for S5^n.

The cut-elimination theorem states that if a sequent is provable, then we can find a proof of the same endsequent without using (Cut). All the inference rules given above except (Cut) add new symbols from the upper sequent(s) to the lower sequent. Therefore, in a cut-free proof P, any formula occurring in some sequent in P also occurs as a subformula in the endsequent. This is called the *subformula property*. On the other hand, the Hilbert-style formulation has Modus Pones, at which one formula is eliminated. In this case, we cannot trace back from a given provable formula what have happened in a proof.

To illustrate the cut-elimination theorem with an example, we use the following stronger version $(\supset\rightarrow)^*$ of $(\supset\rightarrow)$, which is, in fact, equivalent to $(\supset\rightarrow)$ with the presence of (Th) :

$$\frac{\Gamma \rightarrow \Theta, A \qquad B, \Delta \rightarrow \Lambda}{A \supset B, \Gamma, \Delta \rightarrow \Theta, \Lambda} \ (\supset\rightarrow)^*$$

The fact that $(\supset\rightarrow)^*$ is admissible in the above sequent calculus is proved as follows:

$$\frac{\dfrac{\Gamma \rightarrow \Theta, A}{\Gamma, \Delta \rightarrow \Theta, A}\ (\mathrm{Th}) \qquad \dfrac{B, \Delta \rightarrow \Lambda}{B, \Gamma, \Delta \rightarrow \Theta, \Lambda}\ (\mathrm{Th})}{A \supset B, \Gamma, \Delta \rightarrow \Theta, \Lambda} \ (\supset\rightarrow).$$

The following is a proof $A, A \supset B, B \supset C \rightarrow C$ with (Cut).

$$\frac{\dfrac{A \rightarrow A \qquad B \rightarrow B}{A, A \supset B \rightarrow B}\ (\supset\rightarrow)^* \qquad \dfrac{B \rightarrow B \qquad C \rightarrow C}{B, B \supset C \rightarrow C}\ (\supset\rightarrow)^*}{A, A \supset B, B \supset C \rightarrow C} \ (\mathrm{Cut})$$

This can be proved without (Cut).

$$\frac{\dfrac{A \rightarrow A \qquad \dfrac{B \rightarrow B}{B \rightarrow C, B}(\mathrm{Th})}{A, A \supset B \rightarrow C, B}(\supset\rightarrow)^* \qquad \dfrac{A \rightarrow A \qquad \dfrac{C \rightarrow C}{B, C \rightarrow C}\ (\mathrm{Th})}{A, A \supset B, C \rightarrow C}\ (\supset\rightarrow)^*}{A, A \supset B, B \supset C \rightarrow C} \ (\supset\rightarrow)^*$$

5 Kripke semantics: model-theoretic approach

In this section, the basic principle G: "having an argument for A from basic beliefs" is formulated as "A is a true in all the possible models of the basic beliefs". Mathematically, this is formulated in the Kripke semantics. Here, the basic principle G1 of Section 4.1 is clear-cut, but G2, G3 are less clear-cut. Nevertheless, this model-theoretic approach has some technical advantages over the proof-theoretic approach given in Sections 4.1 and 4.2. As already stated, it is difficult to prove unprovability assertions such as (4.5) directly in epistemic logic $\mathscr{S}$. The more complex a formula is, the more difficult to evaluate provability is in $\mathscr{S}$. The Kripke semantics enables us to evaluate such unprovabilities.[17].

[17] Chellas [6] and Hughes and Cresswell [12] are good textbooks on Kripke semantics, which treat uni-modal logics. For multi-modal epistemic logics, see Fagin, *et al* [7] and Meyer and van der Hoek [25].

5.1 Kripke models and completeness

We say that an $n+1$ tuple $(W; R_1, ..., R_n)$ is a *Kripke frame* iff W is an arbitrary nonempty set and each R_i is a binary relation on W. Each element w in W is called a *possible world*, and each R_i an *accessibility relation*. In each world w, we assume the classical truth valuation, i.e., the logical connectives $\neg, \supset, \bigwedge, \bigvee$ are valuated in the manner CV0–CV4. This means that each player has the reasoning ability described by the classical logic in each possible world, which expresses the basic principle G1 of Section 4.2. The other principles G2 and G3 are expressed by the relationships to other possible worlds defined by $R_1, ..., R_n$. Each R_i describes the accessible worlds from each world w, and the truthfulness of $\mathrm{B}_i(A)$ is evaluated by referring to the truthfulness of A in the accessible worlds from w.

An *assignment* σ *in* a Kripke frame $\mathscr{K} = (W; R_1, ..., R_n)$ is a function from $W \times PV$ to $\{\top, \bot\}$. A pair $(\mathscr{K}, \sigma)$ of a Kripke frame $\mathscr{K}$ and an assignment σ is called a *Kripke model*. We define the *valuation relation* $(\mathscr{K}, \sigma, w) \vDash$ and its negation $(\mathscr{K}, \sigma, w) \nvDash$ for each $w \in W$ by induction on the length of a formula:

K0: for each $p \in PV$, $(\mathscr{K}, \sigma, w) \vDash p$ iff $\sigma(w, p) = \top$;

K1: $(\mathscr{K}, \sigma, w) \vDash \neg A$ iff $(\mathscr{K}, \sigma, w) \nvDash A$;

K2: $(\mathscr{K}, \sigma, w) \vDash A \supset B$ iff $(\mathscr{K}, \sigma, w) \nvDash A$ or $(\mathscr{K}, \sigma, w) \vDash B$;

K3: $(\mathscr{K}, \sigma, w) \vDash \bigwedge \Phi$ iff $(\mathscr{K}, \sigma, w) \vDash A$ for all $A \in \Phi$;

K4: $(\mathscr{K}, \sigma, w) \vDash \bigvee \Phi$ iff $(\mathscr{K}, \sigma, w) \vDash A$ for some $A \in \Phi$;

K5: $(\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(A)$ iff $(\mathscr{K}, \sigma, u) \vDash A$ for all u with wR_iu.

The above inductive definition works simultaneously over the possible worlds. We say that A is *true at world* w in $(\mathscr{K}, \sigma)$ iff $(\mathscr{K}, \sigma, w) \vDash A$. This valuation is complete in the sense that for any $w \in W$ and $A \in \mathscr{P}$,

$$\text{either } (\mathscr{K}, \sigma, w) \vDash A \text{ or } (\mathscr{K}, \sigma, w) \vDash \neg A. \tag{5.1}$$

Step K5 expresses the idea that the truth of $\mathrm{B}_i(A)$ in world w is defined by referring to the truth of A in the accessible worlds from w. Accessibility relation R_i describes the possibilities that player i can imagine at each w. We note that when $(\mathscr{K}, \sigma, u) \vDash \neg A$ for some u with wR_iu, we have $(\mathscr{K}, \sigma, u) \nvDash \mathrm{B}_i(A)$ by K5, and then $(\mathscr{K}, \sigma, u) \vDash \neg\mathrm{B}_i(A)$ by K1.

Note that the following does not necessarily hold:

$$\text{either } (\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(A) \text{ or } (\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(\neg A). \tag{5.2}$$

If (5.2) was assumed, then player i would have determinate beliefs about every aspect of the model. The consideration of possibilities in a Kripke frame $\mathscr{K} = (W; R_1, ..., R_n)$ enables us to avoid (5.2). However, (5.1) still implies that either $(\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(A)$ or $(\mathscr{K}, \sigma, w) \vDash \neg\mathrm{B}_i(A)$. Thus, the third possibility is excluded in each model. To avoid this completeness, we consider the validity defined by a *set* of Kripke models. This enables us to have a connection to the proof theoretic approach as well as to erase the superfluous information included in each model. Such a set of Kripke models is defined by conditions on the accessibility relations $R_1, ..., R_n$. Then we have an explicit connection to each syntactical system in Section 4 to the Kripke semantics.

Specifically, we consider the following conditions on $R_1, ..., R_n$, each of which corresponds to an epistemic axiom:

no condition	$\longleftrightarrow$	**K**
seriality : for any $w \in W$, there is some $u \in W$ with wR_iu	$\longleftrightarrow$	**D**
reflexivity : wR_iw for all $w \in W$	$\longleftrightarrow$	**T**
transitivity : for all $w, u, v \in W$, wR_iu and uR_iv imply wR_iv	$\longleftrightarrow$	**4**
euclidean : for all $w, u, v \in W$, wR_iu and wR_iv imply uR_iv	$\longleftrightarrow$	**5.**

We postpone seeing the reasons for these correspondences after the main result. Let $\mathscr{S}$ be an epistemic logic in Diagram 4.2. Then $\mathscr{S}^*$ is the set of all Kripke frames satisfying the conditions on the accessibility relations corresponding to $\mathscr{S}$. For example, if $\mathscr{S}$ is $KD4^n$, $\mathscr{S}^*$ is the set of Kripke frames $\mathscr{K} = (W; R_1, ..., R_n)$ whose $R_1, ..., R_n$ satisfy seriality and transitivity.[18] Using this notation, we can state the following soundness-completeness theorem. It can be regarded as a special case of Theorem 8.2, which will be proved in Section 9. Here, we discuss only the soundness part of Theorem 5.1.

Theorem 5.1 (Soundness-completeness). Let $\mathscr{S}$ be an epistemic logic in Diagram 2, and A a formula in $\mathscr{P}$**.**

(1): $\vdash_{\mathscr{S}} A$ if and only if $(\mathscr{K}, \sigma, w) \vDash A$ for all Kripke frames $\mathscr{K} = (W; R_1, ..., R_n)$ in $\mathscr{S}^*$, all assignments σ and all $w \in W$.

(2): There is a Kripke frame $\mathscr{K} = (W; R_1, ..., R_n)$ in $\mathscr{S}^*$, an assignment σ and a world $w \in W$ satisfying $(\mathscr{K}, \sigma, w) \vDash A$ if and only if A is consistent in logic $\mathscr{S}$.

Note that the consistency of A in $\mathscr{S}$ means $A \supset \neg C \wedge C$ for any $C \in \mathscr{P}$.

Let Γ be a finite nonempty set. Since $\Gamma \vdash_{\mathscr{S}} A$ is defined by $\vdash_{\mathscr{S}} \bigwedge \Gamma \supset A$, Theorem 5.(1) implies

(1^A): $\Gamma \vdash_{\mathscr{S}} A$ if and only if for all $\mathscr{K} = (W; R_1, ..., R_n)$ in $\mathscr{S}^*, \sigma$ and $w \in W$, $(\mathscr{K}, \sigma, w) \vDash C$ for all $C \in \Gamma$ imply $(\mathscr{K}, \sigma, w) \vDash A$.

In the present context, a model of Γ is $(\mathscr{K}, \sigma)$ making all assumptions in Γ true at some world w. Using this terminology, (2) is written: there is a Kripke model of A in $\mathscr{S}^*$ if and only if A is consistent in logic $\mathscr{S}$.

---

18 We find the reason for the popularity of $S5^n$ among game theorists. Reflexivity and euclidean imply symmetry. If $\mathscr{S}$ is $S5^n$, then each R_i in $(W; R_1, ..., R_n)$ in $\mathscr{S}^*$ becomes an equivalence relation. Hence the quotient space $W/R_i = \{\{w \in W : wR_iu\} : u \in W\}$ is a partition of W. The $n+1$-tuple $(W; W/R_1, ..., W/R_n)$ may be regarded as an information partition model (of Aumann [1]). However, a Kripke model describes the possibilities perceived by players but not information processing.

By Theorem 5.1, we have the equivalence between the provability $\vdash_{\mathscr{S}}$ defined by means of symbol manipulations and the consequence relation $\vDash$ defined in terms of a set of possible models. This equivalence is important and useful not only in understanding player i's belief on A as "having an argument for A from basic beliefs" but also in investigating the properties of $\vdash_{\mathscr{S}}$ and $\vDash$. In fact, without the Kripke semantics, we would not go much further than the results in Section 4. The usefulness of the Kripke semantics will be shown by some examples and some other results below.

Let us see the reasons for the correspondences between the epistemic axioms and the conditions on the accessibility relation R_i. In the Kripke semantics, the basic idea is that the truthfulness of $\mathrm{B}_i(A)$ at w is determined by looking at the truthfulness of A in all possible worlds accessible from w. Also, in each world, the classical valuation is assumed. Axiom K: $(\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(A \supset C) \wedge \mathrm{B}_i(A) \supset \mathrm{B}_i(C)$ is derived only by this fact, that is, in any referred world, the truthfulness is closed under Modus Ponens.

The other axioms impose some restrictions on accessibility relations. Axiom D: $(\mathscr{K}, \sigma, w) \vDash \neg \mathrm{B}_i(\neg A \wedge A)$ is derived by the fact that a contradictory formula is not true in any accessible world from w, but needs at least one world, which needs seriality. Axiom T: $(\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(A) \supset A$ requires that the accessible worlds from w include w itself: otherwise, the truthfulness of $\mathrm{B}_i(A)$ should be independent of that of A. Axiom 4: $(\mathscr{K}, \sigma, w) \vDash \mathrm{B}_i(A) \supset \mathrm{B}_i \mathrm{B}_i(A)$ requires that the accessible worlds be closed with R_i in the sense that if u is accessible from w by finite steps of R_i, u is already accessible directly by R_i. In the same manner, the corresponding conditions to Axiom 5 is understood.

Now we exemplify the above theorem by proving the unprovability assertion (4.5) with Kripke models. The game (g_1^5, g_2^5) of Table 5, called the *Matching Pennies*, has no Nash equilibrium, and the game (g_1^1, g_2^5) of Table 6 has the unique Nash equilibrium $(\mathbf{s}_{12}, \mathbf{s}_{21})$.

Table 5. (g_1^5, g_2^5)

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	$(1, -1)$	$(-1, 1)$
$\mathbf{s}_{12}$	$(-1, 1)$	$(1, -1)$

Table 6. (g_1^1, g_2^5)

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	$(5, -1)$	$(1, 1)$
$\mathbf{s}_{12}$	$(6,1)$	$(3, -1)$

$$\circlearrowleft_{1,2}\ w_1 : \hat{g}_1^1 \cup \hat{g}_2^5$$

$$\uparrow_{1,2}$$

$$w_0 : \hat{g}_1^5 \cup \hat{g}_2^5$$

Diagram 4. $(\mathscr{K}, \sigma)$

Consider the Kripke model $(\mathscr{K}, \sigma)$ described as Diagram 4. It is read as follows: Each arrow indexed by i connects the possible worlds with R_i, i.e., $R_i = \{(w_0, w_1), (w_1, w_1)\}$ for $i = 1, 2$, and the assignment σ is determined by the set associated with each world, i.e., for any atomic formula $p \in AF$, $\sigma(w_0, p) = \top$ iff $p \in \hat{g}_1^5 \cup \hat{g}_2^5$ and $\sigma(w_1, p) = \top$ iff $p \in \hat{g}_1^1 \cup \hat{g}_2^5$. This $\mathscr{K}$ is serial and transitive. Since $(\mathscr{K}, \sigma, w_1) \vDash \neg\text{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22})$, we have $(\mathscr{K}, \sigma, w_0) \nvDash \neg\text{B}_1(\text{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22}))$. Also, $(\mathscr{K}, \sigma, w_0) \vDash \text{B}_1(\hat{g}_1^1)$. Hence, $\text{B}_1(\hat{g}_1^1) \nvdash_{\text{KD4}^n} \text{B}_1(\text{Nash}(\mathbf{s}_{12}, \mathbf{s}_{22}))$ by (1^A). On the other hand, $(\mathscr{K}, \sigma, w_0) \vDash \text{B}_1(\text{Nash}(\mathbf{s}_{12}, \mathbf{s}_{21}))$ but $(\mathscr{K}, \sigma, w_0) \nvDash \text{Nash}(\mathbf{s}_{12}, \mathbf{s}_{21})$. Hence, player 1 derives the false belief from his beliefs $\hat{g}_1^1 \cup \hat{g}_2^5$ that $(\mathbf{s}_{12}, \mathbf{s}_{21})$ is a Nash equilibrium.

The two claims of Theorem 5.1 are equivalent. The *if* part of either claim is called *completeness*, which will be proved in Section 9. The *only-if* part, called *soundness*, is proved as follows:

Lemma 5.2. Let $\mathscr{S}$ be an epistemic logic in Diagram 2, $\mathscr{K} = (W; R_1, ..., R_n)$ in $\mathscr{S}^*$ and σ any assignment in $\mathscr{K}$.

(1): Let A be an instance of L1–L5 or an instance of an epistemic axiom for $\mathscr{S}$. Then $(\mathscr{K}, \sigma, w) \vDash A$ for any world $w \in W$.

(2): For any $w \in W$, $(\mathscr{K}, \sigma, w) \vDash$ satisfies MP, $\bigwedge$-Rule, and $\bigvee$-Rule, e.g., MP: if $(\mathscr{K}, \sigma, w) \vDash A \supset B$ and $(\mathscr{K}, \sigma, w) \vDash A$, then $(\mathscr{K}, \sigma, w) \vDash B$.

(3): If $(\mathscr{K}, \sigma, w) \vDash A$ for all $w \in W$, then $(\mathscr{K}, \sigma, w) \vDash \text{B}_i(A)$ for all $w \in W$.

Proof. (1): Consider $(\mathscr{K}, \sigma, w) \vDash \text{B}_i(A \supset C) \supset (\text{B}_i(A) \supset \text{B}_i(C))$. Suppose $(\mathscr{K}, \sigma, w) \vDash \text{B}_i(A \supset C)$ and $(\mathscr{K}, \sigma, w) \vDash \text{B}_i(A)$. These imply $(\mathscr{K}, \sigma, u) \vDash A \supset C$ and $(\mathscr{K}, \sigma, u) \vDash A$ for all u with wR_iu. Hence, $(\mathscr{K}, \sigma, u) \vDash C$ for all u with wR_iu. I.e., $(\mathscr{K}, \sigma, u) \vDash \text{B}_i(C)$.

Next, consider (3). Let $(\mathscr{K}, \sigma, v) \vDash A$ for all $v \in W$. Let w be any world in W. Since $(\mathscr{K}, \sigma, u) \vDash A$ for all u with wR_iu, we have $(\mathscr{K}, \sigma, w) \vDash \text{B}_i(A)$. □

Proof of the only-if part of (1) of Theorem 5.1. Let P be a proof of A. Then we show, by induction on the tree structure of P from its leaves, the assertion that for any C occurring in P, $(\mathscr{K}, \sigma, w) \vDash C$ for any $\mathscr{K}$ in $\mathscr{S}^*$, assignment σ and world w in $\mathscr{K}$. Each of the inductive steps is verified by Lemma 5.2. □

The contradiction-freeness of $\mathscr{S}$ follows from the soundness part of Theorem 5.1.(1).

Theorem 5.3 (Contradiction-freeness). Each $\mathscr{S}$ in Diagram 2 is contradiction-free, i.e., there is no formula A in $\mathscr{P}$ such that $\vdash_{\mathscr{S}} \neg A \wedge A$.

Proof. Suppose $\vdash_{\mathscr{S}} \neg A \wedge A$ for some A. By Theorem 5.1.(1), we have $(\mathscr{K}, \sigma, w) \vDash \neg A \wedge A$ for all $w \in W$ in $\mathscr{K}$, all σ in $\mathscr{K}$ and all $\mathscr{K}$ in $\mathscr{S}^*$. However, this is impossible by (5.1). □

As already stated, if $\Gamma \vdash_0 A$, then $\Gamma \vdash_{\mathscr{S}} A$. When Γ and A are nonepistemic, the converse follows from the definition of the semantical valuation and Theorem 5.1. In this sense, $\mathscr{S}$ is said to be a *conservative extension* of CL.

Theorem 5.4 (Conservativity of $\mathscr{S}$ upon CL). Let $\Gamma \subseteq \mathscr{P}^{\text{n}}$ and $A \in \mathscr{P}^{\text{n}}$. Then $\Gamma \vdash_{\mathscr{S}} A$ if and only if $\Gamma \vdash_0 A$.

5.2 Decision criterion DC1 with false beliefs

Here, let us return to a game problem. Specifically, consider DC1 of Section 2 from the viewpoint of false beliefs. In DC1, neither player predicts the other's decision, and each player's own decision making is relevant for $i = 1, 2$. Here we adopt the dominant strategy criterion, $\mathrm{B}_i(\mathrm{Dom}_i(s_i))$, which we denote by $\hat{D}_{ii}(s_i)$. Let us apply $\hat{D}_{ii}(s_i)$ to the game g^1 of Table 1. By combining (4.3) and (4.4), we have

$$\mathrm{B}_1(\hat{g}_1^1), \mathrm{B}_2^1(\hat{g}_2^1) \vdash_{\mathscr{S}} \bigwedge_i \left(\neg \hat{D}_{ii}(\mathbf{s}_{i1}) \wedge \hat{D}_{ii}(\mathbf{s}_{i2})\right). \tag{5.3}$$

That is, each player i infers from the belief on his payoff function that his decision is his second strategy.

The objectivity of the payoff functions is not included in (5.3). If epistemic logic $\mathscr{S}$ includes Axiom T, beliefs $\mathrm{B}_1(\hat{g}_1^1)$,$\mathrm{B}_2^1(\hat{g}_2^1)$ imply that the game to be played is $g^1 = (g_1^1, g_2^1)$. Now, let us adopt $\mathrm{KD4}^n$ as $\mathscr{S}$. Then it is possible to add any payoff function g_i on $\{\mathbf{s}_{11}, \mathbf{s}_{12}\} \times \{\mathbf{s}_{12}, \mathbf{s}_{22}\}$ to (5.3) as an objective one. In other words, the beliefs $\mathrm{B}_1(\hat{g}_1^1)$, $\mathrm{B}_2^1(\hat{g}_2^1)$ are false in most cases. For example, we can assume that the true game is $g^2 = (g_1^2, g_2^2)$ rather than $g^1 = (g_1^1, g_2^1)$:

$$\hat{g}^2, \mathrm{B}_1(\hat{g}_1^1), \mathrm{B}_2^1(\hat{g}_2^1) \vdash_{\mathscr{S}} \bigwedge_i \left(\neg \hat{D}_{ii}(\mathbf{s}_{i1}) \wedge \hat{D}_{ii}(\mathbf{s}_{i2})\right). \tag{5.4}$$

In this case, 1's belief is true but 2's is false.

Assertion (5.4) is meaningful only if $\hat{g}^2 \cup \mathrm{B}_1(\hat{g}_1^1) \cup \mathrm{B}_2^1(\hat{g}_2^1)$ is consistent in $\mathrm{KD4}^n$. This consistency can be proved by constructing a model of $\hat{g}^2 \cup \mathrm{B}_1(\hat{g}_1^1) \cup \mathrm{B}_2^1(\hat{g}_2^1)$:

$$\circlearrowleft_{1,2}\ w_1 : \hat{g}^1$$

$$\uparrow_{1,2}$$

$$w_0 : \hat{g}^2$$

Diagram 5. $(\mathscr{K}, \sigma)$

where σ assigns $\top$ to each atomic formula included in $\hat{g}^1$ at w_1 and to one in $\hat{g}^2$ at w_0. This $\mathscr{K} = (W; R_1, R_2)$ is serial and transitive frame. Then $(\mathscr{K}, \sigma, w_0) \vDash A$ for any $A \in \hat{g}^2 \cup \mathrm{B}_1(\hat{g}_1^1) \cup \mathrm{B}_2^1(\hat{g}_2^1)$. Hence $\hat{g}^2 \cup \mathrm{B}_1(\hat{g}_1^1) \cup \mathrm{B}_2^1(\hat{g}_2^1)$ is consistent by (1^A) after Theorem 5.1. Since this frame $\mathscr{K}$ does not satisfy reflexivity, $(\mathscr{K}, \sigma)$ is not a model in $\mathrm{S4}^n$. Hence the above consistency proof cannot be converted to $\mathrm{S4}^n$.

6 Beliefs vs. knowledge

As mentioned as B3 in Section 1, we adopt the distinction between *beliefs and knowledge* that knowledge is a true belief, while a belief may be false. Here, truth is referred to the outside thinker, ultimately, the investigator. For example, in $B_i(A)$, the thinker i *only* believes the truth of A, and the investigator determines the truth of A. In $B_j B_i(A)$, thinker j can determine the truth of player i's belief A by referring to j's belief about A. In this case, the investigator may determine the truths of A, $B_i(A)$ as well as $B_j B_i(A)$.[19] Once Axiom T is assumed, beliefs are true to any outside thinkers. To take human interpersonal epistemic interactions seriously, we would like to allow false beliefs. The situation with Axiom T must be a special case.

In fact, Axiom T can be treated inside an epistemic logic without it. Let $\mathscr{S}$ be an epistemic logic without Axioms T and 5, and $\mathscr{S}'$ the logic obtained from $\mathscr{S}$ by adding Axiom T. Then it holds that $\mathscr{S}'$ is faithfully embedded into $\mathscr{S}$. This result guarantees that we capture the distinction between beliefs and knowledge in $\mathscr{S}$. In this section, we show the embedding theorem for $S4^n$ into $KD4^n$. This embedding would not hold with the presence of Axiom 5.

In epistemic logic $\mathscr{S}$, we denote formula $B_i(A) \wedge A$ by $B_i^+(A)$. It means that player i believes A and that A is true to the outside thinker. For example, in $B_j B_i^+(A) = B_j(B_i(A) \wedge A)$, player j thinks about the truth of i's belief comparing with his belief on A. The formula $B_i^+(A)$ satisfies

T$^+$: $\vdash_{\mathscr{S}} B_i^+(A) \supset A$;

K$^+$: $\vdash_{\mathscr{S}} B_i^+(A \supset C) \supset (B_i^+(A) \supset B_i^+(C))$;

4$^+$: $\vdash_{\mathscr{S}} B_i^+(A) \supset B_i^+ B_i^+(A)$, when Axiom 4 is included in $\mathscr{S}$;

Nec$^+$: if $\vdash_{\mathscr{S}} A$, then $\vdash_{\mathscr{S}} B_i^+(A)$.

Thus, if $\mathscr{S}$ includes Axiom 4, the operator $B_i^+(\cdot)$ behaves like an operator in $S4^n$. In this definition, however, only the outermost $B_i(A)$ is replaced by $B_i^+(A)$, but A may include other B_j. Hence, we cannot yet regard B_i^+ exactly as the operator in $S4^n$. To have the exact relationship, we need a more accurate translation. Now, we focus on the case of $S4^n$ and $KD4^n$.

To avoid confusions, we differentiate the formulae in $S4^n$ from those in $KD4^n$. We denote, by $\mathscr{P}_K$, the set of all formulae generated by the same list of symbols of Section 3 except for the replacements of operator symbols $B_1, ..., B_n$ by new ones $K_1, ..., K_n$. Here $K_i(A)$ is intended to mean that player i knows A.

Now we define the translator $\psi : \mathscr{P}_K \to \mathscr{P}$ by the following induction:

T0: for any $p \in PV$, $\psi(p) = p$;

T1: $\psi(\neg A) = \neg\psi A$;

T2: $\psi(A \supset C) = \psi A \supset \psi C$;

T3: $\psi(\bigwedge \Phi) = \bigwedge\{\psi A : A \in \Phi\}$; and $\psi(\bigvee \Phi) = \bigvee\{\psi A : A \in \Phi\}$;

[19] According to philosophical literature, knowledge is defined as "justified true beliefs". In this definition, justification needs some different sources of authority such as experiences or community (see Moser [27] for debates on justifications of beliefs). Objects targeted by epistemic logics are beliefs inferred from basic beliefs. Justifications for inferred beliefs are traced back to those on the basic beliefs. However, when we consider justifications for basic beliefs, we cannot go further to any other sources. To have such justifications, we need a general framework including experiences.

T4: $\psi(\mathrm{K}_i(A)) = \mathrm{B}_i^{+}(\psi A)$ for $i = 1, ..., n$.

That is, any formula A in $\mathscr{P}_{\mathrm{K}}$ is translated into the corresponding formula in $\mathscr{P}$ which is obtained from A by replacing all occurrences of subformulae of the form $\mathrm{K}_i(C)$ by $\mathrm{B}_i^{+}(C^*)$, where C^* is obtained from C by the same principle. For example, $\psi(\mathrm{K}_2\mathrm{K}_1(\mathrm{Dom}_1(\mathbf{s}_{12})) = \mathrm{B}_2^{+}\mathrm{B}_1^{+}(\mathrm{Dom}_1(\mathbf{s}_{12}))$, which is equivalent to $\mathrm{B}_2\mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))\wedge \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))\wedge \mathrm{B}_2(\mathrm{Dom}_1(\mathbf{s}_{12}))\wedge \mathrm{Dom}_1(\mathbf{s}_{12})$.

We have the following theorem, which will be proved in the end of this section.

Theorem 6.1 (Faithful embedding of S4^n into KD4^n). For any $A \in \mathscr{P}_{\mathrm{K}}$, $\vdash_{\mathrm{S4}^n} A$ if and only if $\vdash_{\mathrm{KD4}^n} \psi A$.

Thus, we can discuss logic S4^n inside KD4^n. When we can forget false beliefs for some game theoretic problems, discussions in S4^n are simpler than in KD4^n in that Axiom T can be used in S4^n. Such results in S4^n can be translated into more general discussions with the presence of false beliefs in KD4^n by Theorem 6.1. Conversely, Theorem 6.1 may be used to translate some meta-theorems obtained for KD4^n into S4^n: KD4^n is easier than S4^n from the meta-theoretic point of view. One example will be mentioned in Section 7.2. The same embedding assertion holds between KD^n (also K^n) and KT^n. Over all, S4^n and KT^n can be considered inside KD4^n and KD^n, respectively.

The above embedding theorem fails with the presence of Axiom 5. For example, S5^n cannot be embedded into KD45^n. A counterexample is $\nvdash_{\mathrm{KD45}^n} \neg\mathrm{B}_1^{+}(\neg p) \supset \mathrm{B}_1^{+}(\neg\mathrm{B}_1^{+}(\neg p)))$, where $p \in PV$. This unprovability is proved by constructing a Kripke model.

Logic KD4^n is capable of distinguishing between knowledge and beliefs, while S4^n is not. For example, the following holds:

$$\Gamma \vdash_{\mathrm{KD4}^n} \mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12})) \wedge \mathrm{B}_2\mathrm{B}_1^{+}(\mathrm{Dom}_1(\mathbf{s}_{12})) \wedge \mathrm{B}_2(\mathrm{Best}_2(\mathbf{s}_{22} \mid \mathbf{s}_{12})).$$

where $\Gamma = \hat{g}^5 \cup \mathrm{B}_1(\hat{g}_1^2) \cup \mathrm{B}_2(\hat{g}_2^2) \cup \mathrm{B}_2\mathrm{B}_1^{+}(\hat{g}_1^2)$ and g^5 is the game of Table 5. In Γ, each player i believes that his payoff function is g_i^2, and player 2 believes that player 1 has true beliefs on 1's payoff function g_1^2. Nevertheless, these basic beliefs of both players are all false objectively, since the true game is g^5. Accordingly, player 2 believes that 1's inferred belief, $\mathrm{B}_1(\mathrm{Dom}_1(\mathbf{s}_{12}))$, is true, but this inferred belief is also false objectively.

Proof of Theorem 6.1. The *only-if* part of this theorem can be proved by induction on a proof in S4^n from its leaves, using the above T$^+$, K$^+$, 4$^+$ and Nec$^+$.

The *if* part needs two steps. We define another translator $\varphi : \mathscr{P} \to \mathscr{P}_{\mathrm{K}}$ by T0–T3 and T4$'$: $\varphi(\mathrm{B}_i(C)) = \mathrm{K}_i(\varphi C)$ for $i = 1, ..., n$. That is, φ is the operator which simply substitutes K_i for all occurrences of B_i in A. Then, for any $C \in \mathscr{P}$,

$$\vdash_{\mathrm{KD4}^n} C \text{ implies } \vdash_{\mathrm{S4}^n} \varphi(C). \tag{6.1}$$

This can be proved by induction on a proof in KD4^n.

The second step is the following assertion: for any $A \in \mathscr{P}_{\mathrm{K}}$,

$$\vdash_{S4^n} \varphi \cdot \psi(A) \equiv A. \tag{6.2}$$

This is now proved by induction on the structure of a formula. For any $p \in PV$, $\varphi \cdot \psi(p)$ is p itself, which implies $\vdash_{S4^n} \varphi \cdot \psi(p) \equiv p$.

Suppose the induction hypothesis that (6.2) holds for any immediate subformulae of A. We should consider the cases: $\neg, \supset, \bigwedge, \bigvee$ and K_i. Here we consider only $\neg$ and K_i.

($\neg$): Let $A = \neg C$. By the induction hypothesis, we have $\vdash_{S4^n} \varphi \cdot \psi(C) \equiv C$. Then $\vdash_{S4^n} \neg\varphi \cdot \psi(C) \equiv \neg C$. Since $\neg\varphi \cdot \psi(C)$ is $\varphi \cdot \psi(\neg C)$ by the definitions of φ and ψ, we have $\vdash_{S4^n} \varphi \cdot \psi(\neg C) \equiv \neg C$.

(K_i): Let $A = K_i(C)$. By the induction hypothesis, we have $\vdash_{S4^n} \varphi \cdot \psi(C) \equiv C$. Hence $\vdash_{S4^n} K_i(\varphi \cdot \psi(C)) \equiv K_i(C)$ by Nec, MP and Axiom K. By the definition of φ and ψ, we have $\varphi \cdot \psi(K_i(C)) = \varphi(B_i(\psi(C)) \wedge \psi(C)) = K_i(\varphi \cdot \psi(C)) \wedge \varphi \cdot \psi(C)$. Since $\vdash_{S4^n} K_i(\varphi \cdot \psi(C)) \wedge \varphi \cdot \psi(C) \equiv K_i(\varphi \cdot \psi(C))$, we have $\vdash_{S4^n} \varphi \cdot \psi(K_i(C)) \equiv K_i(C)$.

It follows from (6.1) that for any $A \in \mathscr{P}_K$, $\vdash_{KD4^n} \psi A$ implies $\vdash_{S4^n} \varphi \cdot \psi(A)$, and the latter is equivalent to $\vdash_{S4^n} A$ by (6.2). $\square$

7 Solution theories for DC2 and DC3

As yet our game theoretic consideration was about performance-playability relative to a player's beliefs, taking criterion DC1 as given. Here, we discuss solution theories for DC2 and DC3. In the game theoretic terminology, these are axiomatic considerations of decision making. From the descriptive point of view, there must be a lot of possible prediction-decision criteria in that a lot of arbitrary structures may be considered in such criteria, particularly, in predictions about others' decision making. In this section, we consider only D2 and DC3. More criteria are discussed in Kaneko and Suzuki [19].

Although we give somewhat tedious proofs of an axiomatic characterization of DC2, the point is neither in the characterization nor in the proof, but is in the comparisons with DC3. For DC3, we meet a difficulty caused by an infinite regress of beliefs. This difficulty leads us to an extension of epistemic logic $\mathscr{S}$ to incorporate common knowledge, which is the subject of Section 8.

Throughout this section, we assume $\mathscr{S} = KD4^n$. Let $\{D_{ij}(s_j) : s_j \in S_j$ and $i, j = 1, 2\}$ be a given set of formulae indexed by $s_j \in S_j$ and $i, j = 1, 2$. For $j = i$, $D_{ii}(s_i)$ is intended to mean that s_i is player i's decision, and for $j \neq i$, $D_{ij}(s_j)$ that i predicts that s_j would be a decision of player j.

7.1 Decision criterion DC2

Let $i = 2$ and $j = 1$. Only player 2 predicts 1's decision. We assume that 1's decision criterion is $\hat{D}_{11}(s_1) = B_1(\mathrm{Dom}_1(s_1))$. We require 2's decision and prediction to satisfy:

DC21$_2$: $\bigwedge_{s_1,s_2} \left(D_{22}(s_2) \wedge D_{21}(s_1) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1))\right)$;

DC22$_2$: $\bigwedge_{s_1} (D_{21}(s_1) \supset \mathrm{B}_2(D_{21}(s_1))) \wedge \bigwedge_{s_2} (D_{22}(s_2) \supset \mathrm{B}_2(D_{22}(s_2)))$;

DC23$_2$: $\bigvee_{s_2} D_{22}(s_2) \equiv \bigvee_{s_1} D_{21}(s_1)$;

DC24$_2$: $\bigwedge_{s_1} \left(D_{21}(s_1) \supset \mathrm{B}_2(\hat{D}_{11}(s_1))\right)$.

The first states that if 2 predicts that 1 would choose s_1, then 2 believes that his decision s_2 is a best strategy against s_1. The second requires player 2 to be conscious of his prediction and own decision. The third states that 2's decision is possible if and only if so is 2's prediction, and the last that 2's prediction implies his belief of $\hat{D}_{11}(s_1) = \mathrm{B}_1(\mathrm{Dom}_1(s_1))$.

Our problem is to find appropriate formulae $D_{21}(s_1)$ and $D_{22}(s_2)$ ($s_1 \in S_1$ and $s_2 \in S_2$) satisfying the above four requirements. In fact, contradictory formulae, i.e., the *deductively strongest* formulae, satisfy these requirements in the sense that if we substitute $\neg p \wedge p$ for $D_{22}(s_2)$ and $D_{21}(s_1)$, these would be provable in $\mathscr{S} = \mathrm{KD4}^n$, where p is any atomic formula. However, we would like to find the *deductively weakest* formulae satisfying these requirements, since they have no additional properties other than what the requirements describe. One set of candidates is given as:

$\hat{D}_{21}(s_1) := \mathrm{B}_2\mathrm{B}_1(\mathrm{Dom}_1(s_1))$;

$\hat{D}_{22}(s_2) := \mathrm{B}_2(\bigvee_{t_1} \mathrm{B}_1(\mathrm{Dom}_1(t_1))) \wedge \bigwedge_{t_1} \mathrm{B}_2 \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right)$.

The formula $\hat{D}_{22}(s_2)$ states that 2 has a prediction t_1 about 1's decision, and that whatever his prediction t_1 is, 2's decision s_2 is a best response to t_1. Note that in $\mathscr{S}$, $\hat{D}_{22}(s_2)$ is equivalent to

$$\mathrm{B}_2 \left(\bigvee_{t_1} \mathrm{B}_1(\mathrm{Dom}_1(t_1)) \wedge \bigwedge_{t_1} \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right) \right).$$

These $\hat{D}_{21}(s_1)$ and $\hat{D}_{22}(s_2)$ ($s_1 \in S_1$ and $s_2 \in S_2$) satisfy the above requirements DC21$_2, \ldots,$ DC24$_2$ in the following sense. First, we denoted, by DC21$_2(\hat{D}), \ldots,$ DC24$_2(\hat{D})$, the formulae obtained from DC21$_2, \ldots,$ DC24$_2$ by plugging $\hat{D}_{21}(s_1)$ and $\hat{D}_{22}(s_2)$ to $D_{21}(s_1)$ and $D_{22}(s_2)$. Theorem 7.1 is proved below.

Theorem 7.1.

(1): $\vdash_{\mathscr{S}}$ DC21$_2(\hat{D}) \wedge$DC22$_2(\hat{D}) \wedge$DC24$_2(\hat{D})$.

(2): Let $g = (g_1, g_2)$ be any 2-person game having a unique dominant strategy for player 1. Then $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1), \mathrm{B}_2(\hat{g}_2) \vdash_{\mathscr{S}}$ DC23$_2(\hat{D})$.

Thus, $\hat{D}_{21}(s_1)$ and $\hat{D}_{22}(s_2)$ are candidate formulae for the axioms DC21$_2, \ldots,$ DC24$_2$. Conversely, the following theorem states that $\hat{D}_{21}(s_1)$ and $\hat{D}_{22}(s_2)$ are the deductively weakest formulae satisfying DC21$_2, \ldots,$ DC24$_2$, under the assumption that 1 has a unique dominant strategy. Hence, $\hat{D}_{21}(s_1)$ and $\hat{D}_{22}(s_2)$ are what we look for. This theorem is proved also below.

Theorem 7.2 (Personalized characterization of DC2). Let $g = (g_1, g_2)$ be any 2-person game having a unique dominant strategy for player 1. If $\mathrm{B}_2(\hat{g}_2)$,$\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$ $\vdash_{\mathscr{S}}$ $\mathrm{DC21}_2 \bigwedge \ldots \bigwedge \mathrm{DC24}_2$, then

$$\mathrm{B}_2(\hat{g}_2), \mathrm{B}_2\mathrm{B}_1(\hat{g}_1) \vdash_{\mathscr{S}} \bigwedge_{s_1} \left(D_{21}(s_1) \supset \hat{D}_{21}(s_1)\right) \wedge \bigwedge_{s_2} \left(D_{22}(s_2) \supset \hat{D}_{22}(s_2)\right). \quad (7.1)$$

It follows from Theorems 7.1 and 7.2 that the deductively weakest formulae $\{D_{2j}(s_j) : s_j \in S_j \text{ and } j = 1, 2\}$ satisfying $\mathrm{DC21}_2 \wedge \ldots \wedge \mathrm{DC24}_2$ are *uniquely* determined to be $\{\hat{D}_{2j}(s_j) : s_j \in S_j \text{ and } j = 1, 2\}$.

The assumption that a game $g = (g_1, g_2)$ has a unique dominant strategy for player 1 can be relaxed as follows:

$$\hat{g} \vdash_0 \bigwedge_{t_1, t_1'} \bigwedge_{t_2} \left(\mathrm{Dom}_1(t_1) \wedge \mathrm{Best}_2(t_2 \mid t_1) \wedge \mathrm{Dom}_1(t_1') \supset \mathrm{Best}_2(t_2 \mid t_1')\right). \quad (7.2)$$

In this case, though player 2 predicts multiple dominant strategies for player 1, this multiplicity would cause no problem for player 2 in $g = (g_1, g_2)$. If (7.2) is violated, then we should modify player 2's decision criterion or add more information to guarantee him to have a decision. Condition (7.2) corresponds to the interchangeability condition of Nash [31].[20]

Since the above characterization is purely personalized, this axiomatization is compatible with false beliefs discussed in the previous sections relative to the investigator as well as to the other player.

Proof of Theorem 7.1.
(1): We prove only $\vdash_{\mathscr{S}}$ $\mathrm{DC21}_2(\hat{D})$. Let (s_1, s_2) be any strategy pair. Then $\vdash_{\mathscr{S}}$ $\hat{D}_{22}(s_2) \supset \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(s_1)) \supset \mathrm{Best}_2(s_2 \mid s_1))$ by the definition of $\hat{D}_{22}(s_2)$. Hence $\vdash_{\mathscr{S}}$ $\hat{D}_{22}(s_2) \supset (\mathrm{B}_2\mathrm{B}_1(\mathrm{Dom}_1(s_1)) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1)))$. This is equivalent to $\vdash_{\mathscr{S}}$ $\hat{D}_{22}(s_2) \wedge \mathrm{B}_2\mathrm{B}_1(\mathrm{Dom}_1(s_1)) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1))$. Thus, $\vdash_{\mathscr{S}}$ $\mathrm{DC21}_2(\hat{D})$.

(2): We prove $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$,$\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\bigvee_{s_1} \hat{D}_{21}(s_1) \supset \bigvee_{s_2} \hat{D}_{22}(s_2)$. Let t_1^* be the unique dominant strategy for player 1 in g, and t_2^* a best response to t_1^*. Since $\mathrm{B}_1(\hat{g}_1)$ $\vdash_{\mathscr{S}}$ $\mathrm{B}_1(\mathrm{Dom}_1(t_1^*))$ and $\mathrm{B}_1(\hat{g}_1)$ $\vdash_{\mathscr{S}}$ $\neg\mathrm{B}_1(\mathrm{Dom}_1(t_1))$ for any $t_1 \neq t_1^*$, we have $\mathrm{B}_1(\hat{g}_1), \hat{g}_2$ $\vdash_{\mathscr{S}}$ $\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2^* \mid t_1)$ for all t_1. Thus, $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$, $\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2^* \mid t_1))$ for all t_1. Hence, $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$, $\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\bigwedge_{t_1} \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2^* \mid t_1))$. Hence $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$,$\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\bigvee_{s_1} \hat{D}_{21}(s_1) \supset \bigwedge_{t_1} \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2^* \mid t_1))$. Since $\vdash_{\mathscr{S}}$ $\bigvee_{s_1} \hat{D}_{21}(s_1) \supset \mathrm{B}_2(\bigvee_{s_1} \mathrm{B}_1(\mathrm{Dom}_1(s_1))$, we have $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$,$\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\bigvee_{s_1} \hat{D}_{21}(s_1) \supset \mathrm{B}_2(\bigvee_{s_1} \mathrm{B}_1(\mathrm{Dom}_1(s_1)) \wedge \bigwedge_{t_1} \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2^* \mid t_1))$, i.e., $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$,$\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\bigvee_{s_1} \hat{D}_{21}(s_1) \supset \hat{D}_{22}(t_2^*)$. Hence $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$,$\mathrm{B}_2(\hat{g}_2)$ $\vdash_{\mathscr{S}}$ $\bigvee_{s_1} \hat{D}_{21}(s_1) \supset \bigvee_{s_2} \hat{D}_{22}(s_2)$.

Since $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1) \vdash_{\mathscr{S}} \bigvee_{s_1} \hat{D}_{21}(s_1)$, we have $\mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$,$\mathrm{B}_2(\hat{g}_2) \vdash_{\mathscr{S}} \bigvee_{s_2} \hat{D}_{22}(s_2) \supset \bigvee_{s_1} \hat{D}_{21}(s_1)$, using L1 and MP. □

[20] See Kaneko [13] for such modifications in the case of the decision criterion of a common knowledge Nash strategy.

Proof of Theorem 7.2. The former half follows from $\mathrm{DC24}_2$. Let $\Gamma = \mathrm{B}_2(\hat{g}_2) \cup \mathrm{B}_2\mathrm{B}_1(\hat{g}_1)$. For the latter, it suffices to show $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \supset \hat{D}_{22}(s_2)$, where s_2 is any strategy for 2. By the former half, $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \supset \bigvee_{s_1} \hat{D}_{21}(s_1)$. This further implies $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \supset \mathrm{B}_2(\bigvee_{s_1} \mathrm{B}_1(\mathrm{Dom}_1(s_1))$ by Lemma 4.1.(2). Hence, it remains to show $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \supset \bigwedge_{t_1} \mathrm{B}_2(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1))$. Let t_1^* be the dominant strategy for 1.

Since $\hat{g}_2, \mathrm{B}_1(\hat{g}_1) \vdash_{\mathscr{S}} \mathrm{Best}_2(t_2 \mid t_1^*) \supset \bigwedge_{t_1} \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2 \mid t_1)\right)$, we have $\mathrm{B}_2(\hat{g}_2), \mathrm{B}_2\mathrm{B}_1(\hat{g}_1) \vdash_{\mathscr{S}} \mathrm{B}_2(\mathrm{Best}_2(t_2 \mid t_1^*)) \supset \bigwedge_{t_1} \mathrm{B}_2 (\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2 \mid t_1))$. Then $\Gamma \vdash_{\mathscr{S}} \mathrm{B}_2(\mathrm{Best}_2(t_2 \mid s_1)) \wedge \mathrm{B}_2\mathrm{B}_1(\mathrm{Dom}_1(s_1)) \supset \bigwedge_{t_1} \mathrm{B}_2 (\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(t_2 \mid t_1))$. Since $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \bigwedge D_{21}(s_1) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1)) \wedge \mathrm{B}_2\mathrm{B}_1(\mathrm{Dom}_1(s_1))$ by $\mathrm{DC21}_2$ and $\mathrm{DC24}_2$, we have $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \wedge D_{21}(s_1) \supset \bigwedge_{t_1} \mathrm{B}_2 \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right)$. This is written as $\Gamma \vdash_{\mathscr{S}} D_{21}(s_1) \supset (D_{22}(s_2) \supset \bigwedge_{t_1} \mathrm{B}_2 \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right))$, and then $\Gamma \vdash_{\mathscr{S}} \bigvee_{s_1} D_{21}(s_1) \supset (D_{22}(s_2) \supset \bigwedge_{t_1} \mathrm{B}_2 \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right))$. Since $\vdash_{\mathscr{S}} D_{22}(s_2) \supset \bigvee_{s_2} D_{22}(s_2)$ by L5 and $\Gamma \vdash_{\mathscr{S}} \bigvee_{s_2} D_{22}(s_2) \supset \bigvee_{s_1} D_{21}(s_1)$ by $\mathrm{DC23}_2$, we have $\Gamma \vdash_{\mathscr{S}} D_{22}(s_2) \supset \bigwedge_{t_1} \mathrm{B}_2 \left(\mathrm{B}_1(\mathrm{Dom}_1(t_1)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right)$. □

7.2 Decision criterion DC3 for player 2

Recall that criterion DC2 does not recommend a decision in the game g^3 of Table 3. Suppose that player 2 starts believing that his criterion DC2 is inadequate, and that he adopts DC3 for his criterion. In the following, we consider criterion DC3 from the viewpoint of player 2. Though we treat also 1's prediction-decision making, it occurs in the mind of player 2.

In DC3, we assume that prediction-decision making is reciprocal between 1 and 2, and thus, $D_{11}(s_1), D_{12}(s_2), D_{21}(s_1), D_{22}(s_2)$ are all relevant now. We can keep the requirements for player i corresponding to $\mathrm{DC21}_2$ – $\mathrm{DC23}_2$, which are denoted by $\mathrm{DC31}_i$ – $\mathrm{DC33}_i$, but we modify the fourth one into

$\mathrm{DC34}_1$: $\bigwedge_{s_2} (D_{12}(s_2) \supset \mathrm{B}_1(D_{22}(s_2))) \wedge \bigwedge_{s_1} (D_{11}(s_1) \supset \mathrm{B}_1(D_{21}(s_1)))$;

$\mathrm{DC34}_2$: $\bigwedge_{s_1} (D_{21}(s_1) \supset \mathrm{B}_2(D_{11}(s_1))) \wedge \bigwedge_{s_2} (D_{22}(s_2) \supset \mathrm{B}_2(D_{12}(s_2)))$.

The former conjunct of each means that player i's prediction is based on his belief about the decision criterion for j, and the latter that i believes that his decision is predicted by j. The reciprocity of DC3 stated in Section 2 is involved here.

We focus on the prediction-decision criterion for player 2. For his own criterion, we assume $\mathrm{DC31}_2, \ldots, \mathrm{DC34}_2$. Since player 2 makes also his prediction about 1's prediction-decision, we assume $\mathrm{B}_2(\mathrm{DC31}_1), \ldots, \mathrm{B}_2(\mathrm{DC34}_1)$. Our problem is to find formulae $\{D_{ij}(s_j) : s_j \in S_j \text{ and } i, j = 1, 2\}$ satisfying the requirements $\mathrm{DC31}_2, \ldots, \mathrm{DC34}_2$ and $\mathrm{B}_2(\mathrm{DC31}_1), \ldots, \mathrm{B}_2(\mathrm{DC34}_1)$. In fact, we would meet a serious difficulty. The following theorem states that only trivial formulae would be candidates for $\mathrm{DC34}_2 \wedge \mathrm{B}_2(\mathrm{DC34}_1)$, which is proved in Section 7.3.

Theorem 7.3 (Reciprocal failure). Suppose $\vdash_{\mathscr{S}} \mathrm{DC34}_2 \wedge \mathrm{B}_2(\mathrm{DC34}_1)$. Then for each $(s_1, s_2) \in S$, $\vdash_{\mathscr{S}} \neg(D_{21}(s_1) \wedge D_{22}(s_2))$ or $\vdash_{\mathscr{S}} D_{21}(s_1) \wedge D_{22}(s_2)$.

Thus, as soon as we assume $\mathrm{DC34}_2 \wedge \mathrm{B}_2(\mathrm{DC34}_1)$, $D_{21}(s_1) \wedge D_{22}(s_2)$ becomes a trivial formula. Hence, the above axiomatic system is a failure as a description of prediction-decision making. In fact, this failure is not caused for a game theoretic reason but for a reason in our logical system. It is closely related to the problem of the infinite regress involved in DC3 pointed out in Section 2. Epistemic logic $\mathscr{S}$ is incapable of capturing the reciprocity or infinite regress involved in $\mathrm{DC34}_2 \wedge \mathrm{B}_2(\mathrm{DC34}_1)$. This leads us to an extension of epistemic logic $\mathscr{S}$ to incorporate the concept of common knowledge, which is the subject of Section 8.

Notice that Theorem 7.3 is a meta-theorem evaluating the axiomatic system $\mathrm{DC34}_2 \wedge \mathrm{B}_2(\mathrm{DC34}_1)$ in logic $\mathscr{S} = \mathrm{KD4}^n$, while Theorems 7.1 and 7.2 are object theorems in $\mathscr{S}$.

Before going to the next subsection, we mention one lemma (cf., Chellas [6], p.99). To illustrate a model-theoretic argument, we give a proof of this lemma.

Lemma 7.4. For any A, $\vdash_{\mathscr{S}} \mathrm{B}_i(A)$ if and only if $\vdash_{\mathscr{S}} A$.

Proof. The *if* part follows from the Necessitation rule. We prove the contrapositive of the *only-if* part. Suppose $\nvdash_{\mathscr{S}} A$. By Theorem 5.1, there is a serial and transitive Kripke frame $\mathscr{K} = (W; R_1, ..., R_n)$ and an assignment σ such that $(\mathscr{K}, \sigma, u) \nvDash A$ for some $u \in W$. Let w_0 be a symbol not in W. We extend $(\mathscr{K}, \sigma)$ into $(\mathscr{K}', \sigma') = ((W'; R_1', ..., R_n'), \sigma')$ as follows:

(1): $W' = W \cup \{w_0\}$;

(2): $R_i' = R_i \cup \{(w_0, w) : w \in W\}$ for $i \in N$;

(3): for any $p \in PV$, $\sigma'(w, p) = \sigma(w, p)$ for all $w \in W$ and $\sigma'(w_0, p)$ is arbitrary.

Each R_i' is serial and transitive. It holds also that for any $w \in W$ and any formula C, $(\mathscr{K}, \sigma, w) \vDash C$ if and only if $(\mathscr{K}', \sigma', w) \vDash C$. Hence $(\mathscr{K}', \sigma', u) \nvDash A$. Since $w_0 R_i u$, we have $(\mathscr{K}', \sigma', w_0) \nvDash \mathrm{B}_i(A)$. By Theorem 5.1, we have $\nvdash_{\mathscr{S}} A$. □

Thus, $\vdash_{\mathscr{S}} \mathrm{B}_2(\mathrm{DC34}_1)$ is equivalent to $\vdash_{\mathscr{S}} \mathrm{DC34}_1$. After all, the assumption of Theorem 7.3 is equivalent to $\vdash_{\mathscr{S}} \mathrm{DC34}_2 \wedge \mathrm{DC34}_1$.

7.3 Epistemic depths and the depth lemma

To prove Theorem 7.3, we consider the epistemic depth of a formula. First, let $N^{<\omega>} := \{(i_m, ..., i_1) : i_m, ..., i_1 \in N$ and $i_{k+1} \neq i_k$ for $k = 1, ..., m-1\}$, where we stipulate that $N^{<\omega>}$ includes the null sequence ϵ, i.e., the sequence of length 0. For $e = (i_m, ..., i_1) \in N^{<\omega>}$, $\mathrm{B}_{i_m} ... \mathrm{B}_{i_1}(A)$ is denoted by $\mathrm{B}_e(A)$, and $\mathrm{B}_\epsilon(A)$ is stipulated to be A. We define the following concatenation: for $i \in N$ and $e = (i_m, ..., i_1) \in N^{<\omega>}$, let $i * e = (i, i_m, ..., i_1)$ if $i \neq i_m$ and $i * e = (i_m, ..., i_1)$ if $i = i_m$. Also, we let $i * \epsilon = (i)$.

Let $A \in \mathscr{P}$. We define the (*epistemic*) *depth* $\delta(A)$ of A by induction on the length of a formula:

D0: $\delta(p) = \{\epsilon\}$ for any $p \in PV$;

D1: $\delta(\neg C) = \delta(C)$;

D2: $\delta(C \supset D) = \delta(C) \cup \delta(D)$;

D3: $\delta(\bigwedge \Phi) = \delta(\bigvee \Phi) = \bigcup_{C \in \Phi} \delta(C)$;

D4: $\delta(\mathrm{B}_i(C)) = \{i * e : e \in \delta(A)\}$.

For example, $\delta(\mathbf{p}_0 \supset \mathrm{B}_2\mathrm{B}_3(\mathbf{p}_1)) = \delta(\mathbf{p}_0) \cup \delta(\mathrm{B}_2\mathrm{B}_3(\mathbf{p}_1)) = \{\epsilon, (2,3)\}$. We define $\delta(\Gamma) = \bigcup_{C \in \Gamma} \delta(C)$. The following theorem is due to in Kaneko and Suzuki [18] and a weaker form was given by Kaneko and Nagashima [17].

Theorem 7.5 (Depth lemma for KD4^n). Let Γ be a subset of $\mathscr{P}$ and $\mathrm{B}_e(A) = \mathrm{B}_{i_m}...\mathrm{B}_{i_1}(A)$ a formula in $\mathscr{P}$. Suppose $e \in N^{<\omega>}$ and $e \notin \delta(\Gamma)$. Then $\Gamma \vdash_{\mathscr{S}} \mathrm{B}_e(A)$ if and only if Γ is inconsistent in $\mathscr{S}$ or $\vdash_{\mathscr{S}} A$.

When $\Gamma = \{C\}$, the assertion is written as: $\vdash_{\mathscr{S}} C \supset \mathrm{B}_e(A)$ if and only if $\vdash_{\mathscr{S}} \neg C$ or $\vdash_{\mathscr{S}} A$.

The reader may find that Theorem 7.5 (hence, Theorem 7.3, too) is translated into S4^n by Theorem 6.1.

Now, let us return to the proof Theorem 7.3.

Lemma 7.6. Suppose the assumption of Theorem 7.3. Then, for any odd m and $(i_m, ..., i_1) \in N^{<\omega>}$ with $i_m = 2$, $\vdash_{\mathscr{S}} D_{21}(s_1) \wedge D_{22}(s_2) \supset \mathrm{B}_{i_m}...\mathrm{B}_{i_1}(D_{21}(s_1) \wedge D_{22}(s_2))$.

Proof. The claim for $m = 1$ follows from DC34$_2$. Suppose the claim for odd m. Then $\vdash_{\mathscr{S}} \mathrm{B}_1(D_{21}(s_1) \wedge D_{22}(s_2)) \supset \mathrm{B}_1\mathrm{B}_{i_m}... \mathrm{B}_{i_1}(D_{21}(s_1) \wedge D_{22}(s_2))$. Hence, $\vdash_{\mathscr{S}} \mathrm{B}_2\mathrm{B}_1(D_{21}(s_1) \wedge D_{22}(s_2)) \supset \mathrm{B}_2\mathrm{B}_1\mathrm{B}_{i_m}...\mathrm{B}_{i_1}(D_{21}(s_1) \wedge D_{22}(s_2))$. It remains to show $\vdash_{\mathscr{S}} D_{21}(s_1) \wedge D_{22}(s_2) \supset \mathrm{B}_2\mathrm{B}_1(D_{21}(s_1) \wedge D_{22}(s_2))$. Since $\vdash_{\mathscr{S}} D_{11}(s_1) \wedge D_{12}(s_2) \supset \mathrm{B}_1(D_{21}(s_1) \wedge D_{22}(s_2))$ by $\vdash_{\mathscr{S}}$ DC34$_1$ by Lemma 7.4, we have $\vdash_{\mathscr{S}} \mathrm{B}_2(D_{11}(s_1) \wedge D_{12}(s_2)) \supset \mathrm{B}_2\mathrm{B}_1(D_{21}(s_1) \wedge D_{22}(s_2))$. Since $\vdash_{\mathscr{S}} D_{21}(s_1) \wedge D_{22}(s_2) \supset \mathrm{B}_2(D_{11}(s_1) \wedge D_{12}(s_2))$ by DC34$_2$, we have $\vdash_{\mathscr{S}} D_{21}(s_1) \wedge D_{22}(s_2) \supset \mathrm{B}_2\mathrm{B}_1(D_{21}(s_1) \wedge D_{22}(s_2))$. □

Proof of Theorem 7.3. Take an odd m large enough so that $e = (i_m, ..., i_1) \notin \delta(D_{21}(s_1) \wedge D_{22}(s_2))$ with $i_m = 2$. Applying Theorem 7.5 to the assertion of Lemma 7.6, we have $\vdash_{\mathscr{S}} \neg(D_{21}(s_1) \wedge D_{22}(s_2))$ or $\vdash_{\mathscr{S}} D_{21}(s_1) \wedge D_{22}(s_2)$. □

8 Common knowledge logic $\mathscr{CS}$

The difficulty we met in DC3 is caused by the limited capability of epistemic logic $\mathscr{S}$ to express common knowledge. In $\mathscr{S}$, the common knowledge of a formula A is expressed by as the set $\mathrm{C}^*(A) := \{\mathrm{B}_e(A) : e \in N^{<\omega>}\}$. However, when this is used as nonlogical axioms, e.g., $\mathrm{C}^*(A) \vdash_{\mathscr{S}} B$, only a finite subset of $\mathrm{C}^*(A)$ is used as initial formulae in a proof of B from $\mathrm{C}^*(A)$. The entirety of $\mathrm{C}^*(A)$ is never captured in $\mathscr{S}$, which is the reason for the difficulty. In this section, we extend $\mathscr{S}$ to capture the entirety of $\mathrm{C}^*(A)$. We find two approaches in the logic literature: the *fixed-point* and *infinitary approaches*. In this section, we follow the former due to Halpern and Moses [8]. We also discuss decision making criterion DC3 in the extension.[21]

8.1 Common knowledge logic

We add one more unary operator symbol, C, to the list of primitive symbols given in Section 3.1. In F2 of the definition of formulae, we allow C(A) to be a formula. We denote this extended set of formulae by $\mathscr{P}_{\mathrm{C}}$. Note that this C(A) is a formula and is syntactically different from $\mathrm{C}^*(A)$.

Let $\mathscr{S}$ be an epistemic logic in Diagram 2. We define the extension $\mathscr{CS}$ as follows:

$\mathscr{CS}$: $\mathscr{S}$ + CA + CI within $\mathscr{P}_{\mathrm{C}}$,

where CA and CI are the following axiom schema and inference rule:

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

CI: $\dfrac{D \supset A \wedge \mathrm{B}_1(D) \wedge ... \wedge \mathrm{B}_n(D)}{D \supset \mathrm{C}(A)}$,

where A and D are any formulae. The provability of $\mathscr{CS}$ is denoted by $\vdash_{\mathscr{CS}}$. Lemmas 4.1 and 4.2 hold also for $\mathscr{CS}$.

Axiom CA is the *fixed-point property* that if A is common knowledge, then A holds and each player believes the common knowledge of A. Using CA, Nec and MP a finitely many times, we have

$$\vdash_{\mathscr{CS}} \mathrm{C}(A) \supset \mathrm{B}_e(A) \text{ for all } e = (i_m, ..., i_1) \in N^{<\omega>}. \tag{8.1}$$

Thus, the single formula C(A) includes all the content of $\mathrm{C}^*(A) = \{\mathrm{B}_e(A) : e \in N^{<\omega>}\}$ in $\mathscr{CS}$. On the other hand, CI states that if formula D has the fixed-point property of the same form as CA, then D includes the common knowledge of A. Thus, C(A) is the deductively weakest formula having the fixed-point property.

The following fact may help understand the term "fixed-point": By CA and CI, we have, reading both D and A as C(A) in CI,

$$\vdash_{\mathscr{CS}} \mathrm{C}(A) \supset \mathrm{CC}(A). \tag{8.2}$$

The converse is provable, too.

To see that C(A) captures really the entirety of $\mathrm{C}^*(A) = \{\mathrm{B}_e(A) : e \in N^{<\omega>}\}$ in $\mathscr{CS}$ *with no* superfluous properties, we prove the following lemma using semantics after stating the completeness of $\mathscr{CS}$:

[21] For the fixed-point approach, see also Fagin-Halpern-Moses-Vardi [7], Lismont-Mongin [22],[23] and Meyer and van der Hoek [25]. For the infinitary approach, see Kaneko and Nagashima [16] (including the predicate case) and also Heifetz [9]. Kaneko [14] proved in the propositional case that these approaches can be regarded as equivalent as far as the definition of common knowledge is concerned. However, Wolter [36] proved that this equivalence does not hold in the predicate case. See Kaneko *et al* [20] for a map of common knowledge propositional and predicate logics.

Lemma 8.1.
(1): If $\vdash_{\mathscr{CS}} D \supset \mathrm{B}_e(A)$ for all $e \in N^{<\omega>}$, then $\vdash_{\mathscr{CS}} D \supset \mathrm{C}(A)$.

(2): If $\vdash_{\mathscr{CS}} D \supset \mathrm{B}_i\mathrm{B}_e(A)$ for all $e \in N^{<\omega>}$, then $\vdash_{\mathscr{CS}} D \supset \mathrm{B}_i\mathrm{C}(A)$.

We would like to regard C(A) effectively as equivalent to the conjunction of $\mathrm{C}^*(A) = \{\mathrm{B}_e(A) : e \in N^{<\omega>}\}$, though the infinitary conjunction is not allowed in $\mathscr{CS}$. This effective equivalence can be seen by comparing $\mathscr{CS}$ with classical logic CL. In CL, $\bigwedge \Phi$ is determined by Axiom L4 and $\bigwedge$-Rule. Fact (8.1) corresponds to L4, and Lemma 8.1.(1) does to $\bigwedge$-Rule. Thus, $\mathscr{CS}$ succeeds in having the parallel structure to that of CL. In order to prove Lemma 8.1, however, we need to extend the Kripke semantics and to prove the completeness theorem. The present author has no direct syntactical proof of Lemma 8.1.

Now, consider the semantical counterpart of $\mathscr{CS}$. For $\mathscr{CS}$, we can use the same Kripke semantics, and need only to extend the valuation relation $(\mathscr{K}, \sigma, w) \models$ to $\mathscr{P}_{\mathrm{C}}$ from $\mathscr{P}$.

Let $(\mathscr{K}, \sigma) = ((W; R_1, ..., R_n), \sigma)$ be a Kripke model. We extend the valuation relation $(\mathscr{K}, \sigma, w) \models$ from $\mathscr{P}$ to $\mathscr{P}_{\mathrm{C}}$ by K0–K5 and

K6: $(\mathscr{K}, \sigma, w) \models \mathrm{C}(A)$ if and only if $(\mathscr{F}, \sigma, u) \models A$ for all u reachable from w,

where u is *reachable from* w iff there is a sequence $w_0 = w, w_1, ..., w_m = u$ such that for each $k = 0, 1, ..., m-1$, $w_k R_j w_{k+1}$for some $j \in N$. K6 is equivalent to

K6*: $(\mathscr{K}, \sigma, w) \models \mathrm{C}(A)$ if and only if $(\mathscr{F}, \sigma, w) \models \mathrm{B}_e(A)$ for all $e \in N^{<\omega>}$.

This equivalence can be proved by induction on the length of e.

We state the soundness-completeness of $\mathscr{CS}$, which is proved in Section 9.

Theorem 8.2 (Soundness-completeness of $\mathscr{CS}$). Let $\mathscr{S}$ be an epistemic logic in Diagram 2, and $\mathscr{S}^*$ the set of the Kripke frames satisfying the corresponding conditions on the accessibility relations. Let A be a formula in $\mathscr{P}_{\mathrm{C}}$**.**

(1): $\vdash_{\mathscr{CS}} A$ if and only if $(\mathscr{K}, \sigma, w) \models A$ for all Kripke frames $\mathscr{K}$ in $\mathscr{S}^*$, all assignments σ and all $w \in W$.

(2): There is a Kripke frame $\mathscr{K}$ in $\mathscr{S}^*$, an assignment σ and a world $w \in W$ satisfying $(\mathscr{K}, \sigma, w) \models A$ if and only if A is consistent in $\mathscr{CS}$.

Claims (1) and (2) are equivalent as in Theorem 5.1. The *only-if* part of (1) is proved by modifying the corresponding proof for Theorem 5.1 adding the following steps to Lemma 5.2.

Lemma 8.3.
(1): $(\mathscr{K}, \sigma, w) \models \mathrm{C}(A) \supset A \wedge \mathrm{B}_1\mathrm{C}(A) \wedge ... \wedge \mathrm{B}_n\mathrm{C}(A)$ for any $\mathscr{K} = (W; R_1, ..., R_n)$ in $\mathscr{S}^*$, assignment σ in $\mathscr{K}$ and $w \in W$.

(2): Let $\mathscr{K}$ be any frame in $\mathscr{S}^*$ and σ any assignment in $\mathscr{K}$. If $(\mathscr{K}, \sigma, w) \models D \supset A \wedge \mathrm{B}_1(D) \wedge ... \wedge \mathrm{B}_n(D)$ for any $w \in W$, then $(\mathscr{K}, \sigma, w) \models D \supset \mathrm{C}(A)$ for any $w \in W$.

Proof. We prove only (2). Let u be any world. Suppose $(\mathscr{K}, \sigma, u) \models D$. Then $(\mathscr{K}, \sigma, u) \models A$ and $(\mathscr{K}, \sigma, u) \models \mathrm{B}_i(D)$ for $i = 1, ..., n$. Let u_m be any world so that it is reachable by m steps ($m \geq 0$). We assume the inductive hypothesis that $(\mathscr{K}, \sigma, u_m) \vDash A$ and $(\mathscr{K}, \sigma, u_m) \vDash \mathrm{B}_i(D)$ for $i = 1, ..., n$. Now let $u_m R_i u_{m+1}$. Then $(\mathscr{K}, \sigma, u_{m+1}) \vDash D$. Since $(\mathscr{K}, \sigma, w) \vDash D \supset A \wedge \mathrm{B}_1(D) \wedge ... \wedge \mathrm{B}_n(D)$ for any $w \in W$, we have $(\mathscr{K}, \sigma, u_{m+1}) \vDash A$ and $(\mathscr{K}, \sigma, u_{m+1}) \vDash \mathrm{B}_i(D)$ for $i = 1, ..., n$. Thus, we have proved $(\mathscr{K}, \sigma, v) \vDash A$ for any v reachable from u. By K6, $(\mathscr{K}, \sigma, u) \vDash \mathrm{C}(A)$. □

One possible use of the above soundness-completeness is to prove Lemma 8.1.

Proof of Lemma 8.1. We prove only (2). Suppose $\vdash_{\mathscr{CS}} D \supset \mathrm{B}_i \mathrm{B}_e(A)$ for all $e \in N^{<\omega>}$. Let $(\mathscr{K}, \sigma)$ be any Kripke model. Then $(\mathscr{K}, \sigma, w) \vDash D \supset \mathrm{B}_i \mathrm{B}_e(A)$ for any world w. Then let u be any world. Suppose $(\mathscr{K}, \sigma, u) \vDash D$. Then $(\mathscr{K}, \sigma, u) \vDash \mathrm{B}_i \mathrm{B}_e(A)$. Let u' be any world with $u R_i u'$. Then $(\mathscr{K}, \sigma, u') \vDash \mathrm{B}_e(A)$. Since this holds for any $e \in N^{<\omega>}$, we have $(\mathscr{K}, \sigma, u') \vDash \mathrm{C}(A)$ by K6*. Since this holds for any u' with $u R_i u'$, we have $(\mathscr{K}, \sigma, u) \vDash \mathrm{B}_i \mathrm{C}(A)$. We proved $(\mathscr{K}, \sigma, u) \vDash D \supset \mathrm{B}_i \mathrm{C}(A)$. By Theorem 8.2, we have $\vdash_{\mathscr{CS}} D \supset \mathrm{B}_i \mathrm{C}(A)$. □

The following is an immediate but important consequence from Theorem 8.2.

Theorem 8.4 (Conservativity of $\mathscr{CS}$ upon $\mathscr{S}$). Let A be a formula in $\mathscr{P}$. Then $\vdash_{\mathscr{S}} A$ if and only if $\vdash_{\mathscr{CS}} A$.

This theorem guarantees that for $\mathscr{S} = \mathrm{KD4}^n$, the Reciprocal Failure Theorem and Depth Lemma (Theorems 7.1 and 7.3) can be converted into common knowledge logic $\mathscr{CS}$ with the restrictions of the target formulae to $\mathscr{P}$.

Remark 8.5. In $\mathscr{CS}$, the common knowledge operator C enjoys the properties K, T, 4 and Nec. Hence, $\mathscr{CS}$ may be regarded as $\mathrm{S4}^1$ with respect to C.[22] Therefore, the assertions of Lemma 4.1 hold for C in $\mathscr{CS}$, e.g., (3) of Lemma 4.1 becomes: $\vdash_{\mathscr{CS}} \mathrm{C}(\bigwedge \Phi) \equiv \bigwedge \mathrm{C}(\Phi)$ for a finite nonempy set Φ of formulae in $\mathscr{P}_{\mathrm{C}}$.

8.2 Solution theory for decision criterion DC3

In Section 7.2, we formulated Axioms $\mathrm{DC31}_2, ..., \mathrm{DC34}_2, \mathrm{B}_2(\mathrm{DC31}_1), ..., \mathrm{B}_2(\mathrm{DC34}_1)$ for decision criterion DC3 for player 2, and showed that Axioms $\mathrm{DC34}_2$ and $\mathrm{B}_2(\mathrm{DC34}_1)$ lead to the reciprocal failure. By the conservativity of the extension $\mathscr{CS}$ upon $\mathscr{S}$, the Reciprocal Failure Theorem (Theorem 7.3) still holds for $\mathscr{P}$ in $\mathscr{CS}$, where $\mathscr{S} = \mathrm{KD4}^n$. Now, however, we can look for candidates in $\mathscr{P}_{\mathrm{C}}$ rather than in $\mathscr{P}$. Here, we show that DC3 can be a meaningful criterion in the common knowledge extension $\mathscr{CS}$.

To define the candidate formulae, we first modify $\mathrm{Nash}(s_1, s_2)$ into

[22] In $\mathscr{CS}$, we are treating common *knowledge,* rather than common beliefs, in the sense that it has the property: $\vdash_{\mathscr{CS}} \mathrm{C}(A) \supset A$. As in Section 6, common beliefs can be expressed in common knowledge logic $\mathscr{CS}$. Specifically, the common belief of A is defined as $\bigwedge_i \mathrm{B}_i \mathrm{C}(A)$ in $\mathscr{CS}$. See Kaneko et al. [20]. However, it will be seen in the following that the individual belief of common knowledge plays an important role in our game theoretical application rather than common beliefs.

$$\mathrm{B}_1(\mathrm{Best}_1(s_1 \mid s_2)) \wedge \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1)), \tag{8.3}$$

which we denote by $\mathrm{Nash}^*(s_1, s_2)$. This differs from $\mathrm{Nash}(s_1, s_2)$ in that each individual payoff function is taken up to his beliefs in $\mathrm{Nash}^*(s_1, s_2)$. Then we define the candidate formulae: for $(s_1, s_2) \in S$,

$$\tilde{D}_{11}(s_1) = \mathrm{B}_1\left(\bigvee_{s_2} \mathrm{C}(\mathrm{Nash}^*(s_1, s_2))\right) \text{ and } \tilde{D}_{12}(s_2) = \mathrm{B}_1\left(\bigvee_{s_1} \mathrm{C}(\mathrm{Nash}^*(s_1, s_2))\right) \tag{8.4}$$

$$\tilde{D}_{21}(s_1) = \mathrm{B}_2\left(\bigvee_{s_2} \mathrm{C}(\mathrm{Nash}^*(s_1, s_2))\right) \text{ and } \tilde{D}_{22}(s_2) = \mathrm{B}_2\left(\bigvee_{s_1} \mathrm{C}(\mathrm{Nash}^*(s_1, s_2))\right). \tag{8.5}$$

Now, we have to see that these formulae satisfy Axioms $\mathrm{DC34}_2$ and $\mathrm{B}_2(\mathrm{DC34}_1)$. We denote by $\mathrm{DC34}_1(\tilde{D})$ and $\mathrm{DC34}_2(\tilde{D})$, the formulae obtained by plugging $\tilde{D}_{ij}(s_j)$ into $D_{ij}(s_j)$ in $\mathrm{DC34}_1$ and $\mathrm{DC34}_2$. The proof of the following lemma will be given in the end of this subsection.

Lemma 8.6. $\vdash_{\mathscr{CS}} \mathrm{DC34}_2(\tilde{D}) \wedge \mathrm{B}_2(\mathrm{DC34}_1(\tilde{D}))$.

In fact, requirements $\bigwedge\{\mathrm{DC31}_2, ..., \mathrm{DC34}_2\}$ and $\mathrm{B}_2(\bigwedge\{\mathrm{DC31}_1, ..., \mathrm{DC34}_1\})$ hold for the above candidates $\{\tilde{D}_{ij}(s_j) : s_j \in S_j, i, j = 1, 2\}$. Only $\mathrm{DC31}_2$ and $\mathrm{B}_2(\mathrm{DC31}_1)$ need a game theoretic assumption, which is proved below.

Theorem 8.7. Let $g = (g_1, g_2)$ be a game with a unique Nash equilibrium. Then

(a): $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{DC31}_2(\tilde{D}) \wedge \mathrm{B}_2(\mathrm{DC31}_1(\tilde{D}))$;

(b): $\vdash_{\mathscr{CS}} \bigwedge\{\mathrm{DC32}_2(\tilde{D}), ..., \mathrm{DC34}_2(\tilde{D})\} \wedge \mathrm{B}_2(\bigwedge\{\mathrm{DC32}_1(\tilde{D}), ..., \mathrm{DC34}_1(\tilde{D})\})$.

The next theorem states that formulae $\tilde{D}_{ij}(s_j), i, j = 1, 2$ are the deductively weakest formulae satisfying our requirements. The proof is given below.

Theorem 8.8 (Personalized characterization of DC3). Let $g = (g_1, g_2)$ be a 2-person game with a unique Nash equilibrium. Then $D_{ij}(s_j), i, j = 1, 2$ satisfy our requirements in the sense of (a) and (b) of Theorem 8.7. Then $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \bigwedge_{s_1}\left(D_{21}(s_1) \supset \tilde{D}_{21}(s_1)\right) \wedge \bigwedge_{s_2}\left(D_{22}(s_2) \supset \tilde{D}_{22}(s_2)\right)$.

These theorems correspond to Theorems 7.1 and 7.2 for DC2. As in the case of DC2, Theorems 8.7 and 8.8 imply that the deductively weakest formulae $\tilde{D}_{21}(s_1)$, $\tilde{D}_{22}(s_2), (s_1, s_2) \in S_1 \times S_2$ are uniquely determined.

We have succeeded in avoiding the reciprocal failure by incorporating common knowledge into epistemic logic $\mathscr{S} = \mathrm{KD4}^n$. Nevertheless, we should recognize that epistemic logic $\mathscr{CS}$ involves two levels of infinities: first, $\mathscr{S}$ allows formulae of any epistemic (finite) depths, e.g., $\mathrm{B}_e(A) = \mathrm{B}_{i_m}...\mathrm{B}_{i_1}(A)$ for any m, and second, $\mathscr{CS}$ allows $\mathrm{C}(A)$ to capture the entirety of $\mathrm{C}^*(A) = \{\mathrm{B}_e(A) : e \in N^{<\omega>}\}$. After all, common knowledge is an infinitary concept of an idealization, though CA and CI avoid infinitary treatments in an ingenious way.

For the analysis of human decision making, it seems more natural to avoid such infinitary concepts. Probably, this depends upon a situation. If the rules of the game including payoff functions are visible for the players and if they are standing face-to-face, then these constituents may be regarded as common knowledge between these players. This is due to the special property of vision: Since the speed of light can be regarded as infinity, the mutual verification can be made almost instantaneously (it is reminiscent of the *Analogy of the Sun* of Plato [32], Book VI). On the other hand, payoff functions for games are usually not visible but belong to individual subjectivity. Visions do not help to obtain the common knowledge of payoff functions. This was already seen in the Konnyaku Mondô in Section 1.3: the exchanged gestures were common knowledge between the jelly maker and monk, but the subjective interpretations of the gestures were very different.

The above characterization is made from the viewpoint of player 2. This personalized characterization makes sense in $\mathscr{CS}$ with $\mathscr{S}$ = KD4^n. If we adopt $\mathscr{S}$ = S4^n, then $B_i C(\text{Nash}^*(s_1, s_2))$ is equivalent to $C(\text{Nash}(s_1, s_2))$ in $\mathscr{CS}$. Furthermore,

$$\vdash_{\mathscr{CS}} \tilde{D}_{ij}(s_j) \equiv \bigvee_{s_t} C(\text{Nash}(s_1, s_2)), \text{ where } t \neq j. \tag{8.6}$$

This is what is described often informally by some game theorists.[23] In this case, subjectivity disappears, *a fortiori*, false beliefs cannot be discussed. The false beliefs on common knowledge will be discussed in a game theoretic example in Section 8.3.

Proof of Lemma 8.6. We prove $\vdash_{\mathscr{CS}}$ DC34$_1(\tilde{D})$, which implies $\vdash_{\mathscr{CS}} B_2$(DC34$_1$ $(\tilde{D})$). Recall DC34$_1(\tilde{D})$ is $\bigwedge_{s_1} \left(\tilde{D}_{11}(s_1) \supset B_1(\tilde{D}_{21}(s_1))\right) \wedge \bigwedge_{s_2} \left(\tilde{D}_{12}(s_2) \supset B_1(\tilde{D}_{22}(s_2))\right)$. We prove only the first half. Since $\vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \supset B_1 C(\text{Nash}^*(s_1, s_2))$ by CA, we have $\vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \supset \bigvee_{s_2} B_1 C(\text{Nash}^*(s_1, s_2))$ by L5, which implies $\vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \supset B_1 \left(\bigvee_{s_2} C(\text{Nash}^*(s_1, s_2))\right)$ by Lemma 4.1.(2). I.e., $\vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \supset \tilde{D}_{11}(s_1)$. Since this holds for any s_2, we have $\vdash_{\mathscr{CS}} \bigvee_{s_2} C(\text{Nash}^*(s_1, s_2)) \supset \tilde{D}_{11}(s_1)$ by $\bigvee$-Rule, and $\vdash_{\mathscr{CS}} B_2(\bigvee_{s_2} C(\text{Nash}^*(s_1, s_2))) \supset B_2(\tilde{D}_{11}(s_1))$. I.e., $\vdash_{\mathscr{CS}} \tilde{D}_{21}(s_1) \supset B_2(\tilde{D}_{11}(s_1))$. □

Proof of Theorem 8.7. We prove $B_2 C(\hat{g}) \vdash_{\mathscr{CS}}$ DC31$_2(\tilde{D})$. Since $B_1(\hat{g}_1), B_2(\hat{g}_2) \vdash_{\mathscr{CS}} \text{Nash}^*(s_1, s_2) \wedge \text{Nash}^*(t_1, t_2) \supset \text{Nash}^*(t_1, s_2)$ using the uniqueness of a Nash equilibrium, we have $CB_1(\hat{g}_1), CB_2(\hat{g}_2) \vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \wedge C(\text{Nash}^*(t_1, t_2)) \supset C(\text{Nash}^*(t_1, s_2))$ by Remark 8.5. Since $C(\hat{g}) \vdash_{\mathscr{CS}} CB_i(A)$ for all $A \in \hat{g}_i$ and $i = 1, 2$, we have $C(\hat{g}) \vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \bigwedge C(\text{Nash}^*(t_1, t_2)) \supset C(\text{Nash}^*(t_1, s_2))$. Hence $C(\hat{g}) \vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \wedge C(\text{Nash}^*(t_1, t_2)) \supset B_2(\text{Best}_2(s_2 \mid t_1))$. Since $C(\hat{g}) \vdash_{\mathscr{CS}} C(\text{Nash}^*(r_1, r_2))$ if and only if $\hat{g} \vdash_{\mathscr{CS}} \text{Nash}(r_1, r_2)$ for any (r_1, r_2), we have $C(\hat{g}) \vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \wedge C(\text{Nash}^*(t_1, t_2)) \supset \text{Best}_2(s_2 \mid t_1)$. We can introduce $\bigvee$ to the premise: indeed, the following is equivalent to the last formula,

$$\vdash_{\mathscr{CS}} C(\text{Nash}^*(s_1, s_2)) \supset \left(\bigwedge C(\hat{g}) \wedge C(\text{Nash}^*(t_1, t_2)) \supset \text{Best}_2(s_2 \mid t_1)\right).$$

Since s_1 is arbitrary, we have, applying $\bigvee$-Rule,

[23] This set of decision criteria is considered in Kaneko and Nagashima [16] with the mixed strategies, and is considered in Kaneko [13] with pure strategies.

$$\vdash_{\mathscr{CS}} \bigvee_{s_1} \mathrm{C}(\mathrm{Nash}^*(s_1,s_2)) \supset \left(\bigwedge \mathrm{C}(\hat{g}) \wedge \mathrm{C}(\mathrm{Nash}^*(t_1,t_2)) \supset \mathrm{Best}_2(s_2 \mid t_1)\right).$$

This is equivalent to $\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \bigvee_{s_1}\mathrm{C}(\mathrm{Nash}^*(s_1,s_2) \wedge \mathrm{C}(\mathrm{Nash}^*(t_1,t_2)) \supset \mathrm{Best}_2(s_2 \mid t_1)$. Using a similar argument, we have $\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \bigvee_{s_1}\mathrm{C}(\mathrm{Nash}^*(s_1,s_2)) \wedge \bigvee_{t_2}\mathrm{C}(\mathrm{Nash}^*(t_1,t_2)) \supset \mathrm{Best}_2(s_2 \mid t_1)$. Hence, $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(\bigvee_{s_1}\mathrm{C}(\mathrm{Nash}^*(s_1,s_2))) \wedge \mathrm{B}_2(\bigvee_{t_2}\mathrm{C}(\mathrm{Nash}^*(t_1,t_2))) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid t_1))$, i.e., $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \tilde{D}_{22}(s_2) \wedge \tilde{D}_{21}(t_1) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid t_1))$. Thus, $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{DC31}_2(\tilde{D})$.

Similarly, $\mathrm{B}_1\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \tilde{D}_{11}(s_1) \wedge \tilde{D}_{12}(t_2) \supset \mathrm{B}_1(\mathrm{Best}_1(s_1 \mid t_2))$. By CA, we have $\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \tilde{D}_{11}(s_1) \wedge \tilde{D}_{12}(t_2) \supset \mathrm{B}_1(\mathrm{Best}_1(s_1 \mid t_2))$. Thus, $\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{DC31}_1(\tilde{D})$. Hence $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(\mathrm{DC31}_1(\tilde{D}))$. □

For the proof of Theorem 8.8, we first prove the following lemma.

Lemma 8.9. $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_{21}(s_1) \wedge D_{22}(s_2) \supset \mathrm{B}_2\mathrm{C}(\mathrm{Nash}^*(s_1,s_2))$ for $i = 1,2$.

Proof. Denote $D_{i1}(s_1) \wedge D_{i2}(s_2)$ by $D_i(s_1,s_2)$ for $i = 1,2$. First, by $\mathrm{DC31}_2$, $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1))$. Hence $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2\mathrm{B}_2(\mathrm{Best}_2(s_2 \mid s_1))$. Second, by $\mathrm{B}_2(\mathrm{DC31}_1)$, $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(D_1(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1(\mathrm{Best}_1(s_1 \mid s_2))$. Since $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2(D_1(s_1,s_2))$ by $\mathrm{DC34}_2$, we have $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2\mathrm{B}_1(\mathrm{Best}_1(s_1 \mid s_2))$. Thus, $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2(\mathrm{Nash}^*(s_1,s_2))$.

In the same manner, we have $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(D_1(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1(\mathrm{Nash}^*(s_1,s_2))$. For this, we use $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(D_1(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1(D_2(s_1,s_2))$.

Suppose the induction hypothesis that $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2\mathrm{B}_e(\mathrm{Nash}^*(s_1,s_2))$ for all $e = (i_m,\ldots,i_1)$ and $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(D_1(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1\mathrm{B}_e(\mathrm{Nash}^*(s_1,s_2))$. We prove these for $(i_{m+1},i_m,\ldots,i_1)$. Since $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2(D_1(s_1,s_2))$, we have $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2\mathrm{B}_1\mathrm{B}_e(\mathrm{Nash}^*(s_1,s_2))$ as well as $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_2(s_1,s_2) \supset \mathrm{B}_2\mathrm{B}_2\mathrm{B}_e(\mathrm{Nash}^*(s_1,s_2))$. Finally, since $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(D_1(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1(D_2(s_1,s_2))$, and $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2\mathrm{B}_1(D_2(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1\mathrm{B}_2\mathrm{B}_e(\mathrm{Nash}^*(s_1,s_2))$, we have $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \mathrm{B}_2(D_1(s_1,s_2)) \supset \mathrm{B}_2\mathrm{B}_1\mathrm{B}_2\mathrm{B}_e(\mathrm{Nash}^*(s_1,s_2))$. □

Proof of Theorem 8.8. Since Lemma 8.9 is equivalent to $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_{21}(s_1) \supset (D_{22}(s_2) \supset \mathrm{B}_2\mathrm{C}(\mathrm{Nash}^*(s_1,s_2)))$. Thus $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_{21}(s_1) \supset (D_{22}(s_2) \supset \bigvee_{t_1}\mathrm{B}_2\mathrm{C}(\mathrm{Nash}^*(t_1,s_2)))$, and then $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} \bigvee_{s_1} D_{21}(s_1) \supset (D_{22}(s_2) \supset \bigvee_{t_1}\mathrm{B}_2\mathrm{C}(\mathrm{Nash}^*(t_1,s_2)))$. By $\mathrm{DC33}_2$, $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_{22}(s_2) \supset \bigvee_{t_1}\mathrm{B}_2\mathrm{C}(\mathrm{Nash}^*(t_1,s_2))$. By Lemma 4.1.(2), we have $\mathrm{B}_2\mathrm{C}(\hat{g}) \vdash_{\mathscr{CS}} D_{22}(s_2) \supset \mathrm{B}_2(\bigvee_{t_1}\mathrm{C}(\mathrm{Nash}^*(t_1,s_2))$. □

8.3 Konnyaku Mondô phenomena: mutual misunderstanding of common understanding in DC3

We have relativized the concept of common knowledge to individual beliefs of common knowledge. In particular, we have the individual belief of common knowledge of a Nash strategy, $\tilde{D}_{ij}(s_j) \equiv \mathrm{B}_i(\bigvee_{s_t}\mathrm{C}(\mathrm{Nash}^*(s_1,s_2)))$, where $t \neq j$, as decision and prediction criteria. This relativization enables us to discuss the phenomenon in a game situation like the Konnyaku Mondô mentioned in Section

1.3. That is, each of two players falsely believes that a different game is common knowledge between the players.

Consider the assumption set $\mathrm{B}_1\mathrm{C}(\hat{g}^3)\cup\ \mathrm{B}_2\mathrm{C}(\hat{g}^4)$, where g^3 and g^4 are the games of Tables 3 and 4. That is, player 1 believes that it is common knowledge that game g^3 is played, while 2 believes that it is common knowledge that g^4 is played. We can prove that $\mathrm{B}_1\mathrm{C}(\hat{g}^3)\cup\ \mathrm{B}_2\mathrm{C}(\hat{g}^4)$ is consistent in $\mathscr{CS}$ with $\mathscr{S}$ = $\mathrm{KD4}^n$, and that

$$\mathrm{B}_1\mathrm{C}(\hat{g}^3), \mathrm{B}_2\mathrm{C}(\hat{g}^4) \vdash_{\mathscr{CS}} \tilde{D}_{11}(\mathbf{s}_{12}) \wedge \tilde{D}_{12}(\mathbf{s}_{22}) \wedge \tilde{D}_{21}(\mathbf{s}_{12}) \wedge \tilde{D}_{22}(\mathbf{s}_{22}).$$

This states that both players' predictions are *behaviorally* correct. Nevertheless, decisions and predictions are based on the mutual misunderstanding of common understanding. Neither player would find this misunderstanding by seeing the resulting choice of the other player, since the predictions are behaviorally correct. This argument cannot be done in logic $\mathscr{CS}$ with $\mathscr{S}$ = $\mathrm{S4}^n$, since $\mathrm{B}_1\mathrm{C}(\hat{g}^3)\cup\mathrm{B}_2\mathrm{C}(\hat{g}^4)$ becomes inconsistent.

The consistency of $\mathrm{B}_1\mathrm{C}(\hat{g}^3)\cup\mathrm{B}_2\mathrm{C}(\hat{g}^4)$ in $\mathscr{CS}$ with $\mathscr{S}$ = $\mathrm{KD4}^n$ is verified by constructing the following model:

$$\begin{array}{ccc} \circlearrowright_{12}\ w_1 : \hat{g}^3 & & \circlearrowright_{12}\ w_2 : \hat{g}^4 \\ & \nwarrow_1 \quad \nearrow_2 & \\ & w_0 & \end{array}$$

Diagram 6

Then, since $(\mathscr{K}, \sigma, w_1) \vDash \mathrm{C}(\bigwedge \hat{g}^3)$ and $(\mathscr{K}, \sigma, w_2) \vDash \mathrm{C}(\bigwedge \hat{g}^4)$, we have $(\mathscr{K}, \sigma, w_0) \vDash \mathrm{B}_1\mathrm{C}(\bigwedge \hat{g}^3) \wedge \mathrm{B}_2\mathrm{C}(\bigwedge \hat{g}^4)$.

The above mutual misunderstanding of common understanding may be observed in our life. The point here is the possibility that each player develops a false and different belief of the common knowledge of the situation. This is exactly the point suggested by the Konnyaku Mondô of Section 1.3.

9 Proof of the completeness of $\mathscr{CS}$

Here we prove the completeness part for Theorem 8.2 for $\mathscr{S}$ = $\mathrm{KD4}^n$. In other cases, we need some modifications (see Halpern and Moses [8]). Let A be a formula, which is assumed to be consistent in $\mathscr{CS}$. We are going to show that there is a serial and transitive Kripke frame $\mathscr{K} = (W; R_1, ..., R_n)$ and an assignment σ such that for some $w \in W$, $(\mathscr{K}, \sigma, w) \vDash A$.

If common knowledge operator C does not occur in A, then the following proof becomes also a proof of the completeness of Theorem 5.2.(2) for $\mathscr{S}$, in which case we can ignore the step (5) of the induction proof of (9.1). We suggest that the reader who is not familiar with proofs in logic should read the proof of the completeness for CL given in the Appendix before this subsection.

We denote $\{D : D$ is a subformula of $A\} \cup \bigcup_{i \in N} \{\mathrm{B}_i(D), \mathrm{B}_i\mathrm{C}(D) : \mathrm{C}(D)$ is a subformula of $A\}$ by Sub(A).[24] Then we define $\mathrm{Sub}^+(A) = \mathrm{Sub}(A) \cup \{\neg D : D \in \mathrm{Sub}(A)\}$. We say that a subset Γ of $\mathrm{Sub}^+(A)$ is *maximally consistent* iff it is consistent in $\mathscr{CS}$ and $\Gamma \cup \{D\}$ is inconsistent for any $D \in \mathrm{Sub}^+(A) - \Gamma$. We denote the set $\{\Gamma : \Gamma$ is a maximally consistent subset of $\mathrm{Sub}^+(A)$ in $\mathscr{CS}\}$ by Con(A).

Each maximally consistent set Γ of $\mathrm{Sub}^+(A)$ has the following properties.

Lemma 9.1. Let $\Gamma \in \mathrm{Con}(A)$. Then

(1): for any $\neg B$ in $\mathrm{Sub}^+(A)$, either $B \in \Gamma$ or $\neg B \in \Gamma$;
(2): for any B in $\mathrm{Sub}^+(A)$, $\neg B \in \Gamma \Leftrightarrow B \notin \Gamma$;
(3): for any $B \supset C$ in $\mathrm{Sub}^+(A)$, $B \supset C \in \Gamma \Leftrightarrow \neg B \in \Gamma$ or $C \in \Gamma$;
(4): for any $\bigwedge \Phi$ in $\mathrm{Sub}^+(A)$, $\bigwedge \Phi \in \Gamma \Leftrightarrow B \in \Gamma$ for all $B \in \Phi$;
(5): for any $\bigvee \Phi$ in $\mathrm{Sub}^+(A)$, $\bigvee \Phi \in \Gamma \Leftrightarrow B \in \Gamma$ for some $B \in \Phi$.

For a set u of formulae, we write $u^{\mathrm{B}_i} := \{\mathrm{B}_i(C) : \mathrm{B}_i(C) \in u\}$ and $u^{\backslash \mathrm{B}_i} := \{C : \mathrm{B}_i(C) \in u\}$. We define a Kripke model $\mathscr{K} = (W; R_1, ..., R_n)$ and an assignment σ by

M1: $W = \mathrm{Con}(A)$;
M2: $R_i = \{(u, v) \in W \times W : u^{\backslash \mathrm{B}_i} \cup u^{\mathrm{B}_i} \subseteq v\}$ for all $i \in N$;
M3: for any $(w, p) \in W \times PV$, $\sigma(w, p) = \top$ iff $p \in w$.

First, we verify that each R_i is serial and transitive.

Lemma 9.2.
R_i is serial and transitive.

Proof. Consider seriality. Let $u \in W$. Consider the set $u^{\backslash \mathrm{B}_i} \cup u^{\mathrm{B}_i}$. We prove that this is a consistent set. Suppose not. Then there is a finite subset $\{C_1, ..., C_\ell$, $\mathrm{B}_i(C_{\ell+1}), ..., \mathrm{B}_i(C_k)\}$ of $u^{\backslash \mathrm{B}_i} \cup u^{\mathrm{B}_i}$ such that $\vdash_{\mathscr{CS}} C_1 \wedge ... \wedge C_\ell \wedge \mathrm{B}_i(C_{\ell+1}) \wedge ... \wedge \mathrm{B}_i(C_k) \supset \neg D \wedge D$. Then $\vdash_{\mathscr{CS}} \mathrm{B}_i(C_1 \wedge ... \wedge C_\ell) \wedge \mathrm{B}_i(C_{\ell+1}) \wedge ... \wedge \mathrm{B}_i(C_k) \supset \mathrm{B}_i(\neg D \wedge D)$. However, by Axiom D, $\vdash_{\mathscr{CS}} \neg \mathrm{B}_i(\neg D \wedge D)$. This means that u itself is inconsistent. Hence $u^{\backslash \mathrm{B}_i} \cup u^{\mathrm{B}_i}$ is consistent. Then we have a maximally consistent subset v of $\mathrm{Sub}^+(A)$ including $u^{\backslash \mathrm{B}_i} \cup u^{\mathrm{B}_i}$. Then $(u, v) \in R_i$.

Consider transitivity. Let $(u, v) \in R_i$ and $(v, w) \in R_i$. Take $\mathrm{B}_i(C)$ from u. Then $\mathrm{B}_i(C) \in v$. This implies $\mathrm{B}_i(C) \in w$ and $C \in w$. □

Now we prove by induction on the length of a formula that for any $C \in \mathrm{Sub}^+(A)$ and any $v \in W$,

$$C \in v \text{ if and only if } (\mathscr{K}, \sigma, v) \vDash C. \tag{9.1}$$

Suppose that (9.1) is proved. Since A is consistent, there is a $w \in W$ with $A \in w$. Thus, $(\mathscr{K}, \sigma, w) \vDash A$ by (9.1).

The 0-th step for (9.1) is the basis of the induction proof.

(0): Let C be a propositional variable in $\mathrm{Sub}^+(A)$. By M3, $C \in v$ if and only if $\sigma(v, C) = \top$. This is further equivalent to $(\mathscr{F}, \sigma, v) \vDash C$ by E0. This is (9.1).

[24] A *subformula* of A is a formula appearing in the inductive construction of A.

Consider a formula F in $\mathrm{Sub}^+(A)$ which is not a propositional variable. Now we assume the induction hypothesis that for any immediate subformula D of F, $D \in v$ if and only if $(\mathscr{K}, \sigma, v) \vDash D$. Then we prove (9.1) for F. In each step of (1)–(3), we use Lemma 9.1.

(1): Let the formula F in question be expressed as $\neg C$. Suppose $\neg C \in v$. Then $C \notin v$. By the induction hypothesis, $(\mathscr{K}, \sigma, v) \nvDash C$. Hence $(\mathscr{K}, \sigma, v) \vDash \neg C$. The converse is similar.

(2): Let F be $B \supset C$. Suppose $B \supset C \in v$. Then $\neg B \in v$ or $C \in v$. By the induction hypothesis, $(\mathscr{K}, \sigma, v) \nvDash B$ or $(\mathscr{K}, \sigma, v) \vDash C$. Hence $(\mathscr{K}, \sigma, v) \vDash B \supset C$. The converse is similar.

(3): Similarly, we can prove (9.1) in the case where F is expressed as $\bigwedge \Phi$ or $\bigvee \Phi$.

(4): Let the formula F in question be expressed as $\mathrm{B}_i(C)$.

First, we prove that $(\mathscr{K}, \sigma, v) \vDash \mathrm{B}_i(C)$ implies $\mathrm{B}_i(C) \in v$. Suppose $(\mathscr{K}, \sigma, v) \vDash \mathrm{B}_i(C)$. We show the inconsistency of $v^{\backslash \mathrm{B}_i} \cup v^{\mathrm{B}_i} \cup \{\neg C\}$. Suppose that this is consistent. There is a maximally consistent set u in $\mathrm{Sub}^+(A)$ including this set. Thus $(v, u) \in R_i$. However, $(\mathscr{F}, \sigma, v) \vDash \mathrm{B}_i(C)$ implies $(\mathscr{K}, \sigma, u) \vDash C$. By the induction hypothesis, we have $C \in u$, a contradiction to $\neg C \in u$. By the inconsistency of $v^{\backslash \mathrm{B}_i} \cup v^{\mathrm{B}_i} \cup \{\neg C\}$, there is a finite subset $\{D_1, ..., D_\ell, \mathrm{B}_i(D_{\ell+1}), ..., \mathrm{B}_i(D_k)\}$ of $v^{\backslash \mathrm{B}_i} \cup v^{\mathrm{B}_i}$ such that $\vdash_{\mathscr{CS}} D_1 \wedge ... \wedge D_\ell \wedge \mathrm{B}_i(D_{\ell+1}) \wedge ... \wedge \mathrm{B}_i(D_k) \supset C$. Then $\vdash_{\mathscr{CS}} \mathrm{B}_i(D_1 \wedge ... \wedge D_\ell) \wedge \mathrm{B}_i(D_{\ell+1}) \wedge ... \wedge \mathrm{B}_i(D_k) \supset \mathrm{B}_i(C)$. Hence $\mathrm{B}_i(C) \in v$.

Conversely, suppose $\mathrm{B}_i(C) \in v$. Then $C \in w$ for all w with $(v, w) \in R_i$ by M2. Hence $(\mathscr{F}, \sigma, w) \vDash C$ for all w with $(v, w) \in R_i$ by the induction hypothesis. This implies $(\mathscr{K}, \sigma, v) \vDash \mathrm{B}_i(C)$.

(5): Let F be expressed as $\mathrm{C}(D)$. Now we prove that $\mathrm{C}(D) \in v$ if and only if $(\mathscr{K}, \sigma, v) \vDash \mathrm{C}(D)$.

Suppose $\mathrm{C}(D) \in v$. We show by induction on k that if w is reachable from v in k steps, then D and $\mathrm{C}(D)$ are in w. Let $k = 1$. Observe that CA implies $\mathrm{B}_i(D) \in v$ and $\mathrm{B}_i\mathrm{C}(D) \in v$. If w is reachable from v in one step, i.e., $(v, w) \in R_i$ for some i, we have $D \in w$ and $\mathrm{C}(D) \in w$. Now we assume the claim for k. Suppose that w is reachable from v in $k + 1$ steps. Then there is a u such that is reachable from v in k steps and $(u, w) \in R_i$. By the induction hypothesis, D and $\mathrm{C}(D)$ are in u. By CA, $\mathrm{B}_i(D) \in u$ and $\mathrm{B}_i\mathrm{C}(D) \in u$. Since $(u, w) \in R_i$, we have $D \in w$ and $\mathrm{C}(D) \in w$. In sum, $D \in w$ for all w reachable from v. By our main induction hypothesis, $(\mathscr{K}, \sigma, w) \vDash D$ for all w reachable from v. Thus $(\mathscr{K}, \sigma, v) \vDash \mathrm{C}(D)$ by K6.

Conversely, suppose $(\mathscr{K}, \sigma, v) \vDash \mathrm{C}(D)$. We define $W_D := \{w : (\mathscr{F}, \sigma, w) \vDash \mathrm{C}(D)\}$. Since W_D is a set of subsets of $\mathrm{Sub}^+(A)$, this is a finite set. Let φ_w be $\bigwedge w$, i.e., the conjunction of w, and let $\varphi_{W_D} := \bigvee_{w \in W_D} \varphi_w$. We are going to prove

$$\vdash_{\mathscr{CS}} \varphi_{W_D} \supset D \wedge \mathrm{B}_1(\varphi_{W_D}) \wedge ... \wedge \mathrm{B}_n(\varphi_{W_D}). \tag{9.2}$$

Once this is done, we have $\vdash_{\mathscr{CS}} \varphi_{W_D} \supset \mathrm{C}(D)$ by CI. Since $v \in W_D$, we have $\vdash_{\mathscr{CS}} \varphi_v \supset \varphi_{W_D}$. Hence $\vdash_{\mathscr{CS}} \varphi_v \supset \mathrm{C}(D)$. Thus $\mathrm{C}(D) \in v$.

In the remaining, we prove (9.2). First, $\vdash_{\mathscr{CS}} \varphi_{W_D} \supset D$ is proved as follows. Let w be an arbitrary world in W_D. Since $(\mathscr{K}, \sigma, w) \models \mathrm{C}(D)$ implies $(\mathscr{K}, \sigma, w) \models D$ by K6. This implies $D \in w$ by the induction hypothesis. Thus $\vdash_{\mathscr{CS}} \varphi_{W_D} \supset D$. It remains to prove that $\vdash_{\mathscr{CS}} \varphi_{W_D} \supset \mathrm{B}_i(\varphi_{W_D})$. Let w be arbitrary in W_D and $i = 1, ..., n$. It suffices to prove $\vdash_{\mathscr{CS}} \varphi_w \supset \mathrm{B}_i(\varphi_{W_D})$. Since $\vdash_{\mathscr{CS}} \varphi_{W_D} \equiv \neg\left(\bigvee_{w' \in W - W_D} \varphi_{w'}\right)$, $\vdash_{\mathscr{CS}} \varphi_w \supset \mathrm{B}_i(\varphi_{W_D})$ is equivalent to $\vdash_{\mathscr{CS}} \varphi_w \supset \mathrm{B}_i(\bigwedge_{w' \in W - W_D} \neg\varphi_{w'})$. The latter follows if we prove that for each $w' \in W - W_D$,

$$\vdash_{\mathscr{CS}} \varphi_w \supset \mathrm{B}_i(\neg\varphi_{w'}). \tag{9.3}$$

Suppose that (9.3) does not hold for some $w' \in W - W_D$. Then $\varphi_w \wedge \neg\mathrm{B}_i(\neg\varphi_{w'})$ is consistent. We will show that this implies $w^{\backslash \mathrm{B}_i} \subseteq w'$ and $w^{\backslash \mathrm{B}_i} \subseteq w'^{\backslash \mathrm{B}_i}$. Suppose that this is proved. Then we have $(w, w') \in R_i$. Since $w \in W_D$ and $w' \in W - W_D$, we have $(\mathscr{K}, \sigma, w) \models \mathrm{C}(D)$ but $(\mathscr{K}, \sigma, w') \not\models \mathrm{C}(D)$. The latter implies that some $e \in N^{<\omega>}$, $(\mathscr{K}, \sigma, w') \not\models \mathrm{B}_e(D)$. However, the former implies $(\mathscr{F}, \sigma, w) \models \mathrm{B}_i\mathrm{B}_e(D)$, which together with $(w, w') \in R_i$ implies $(\mathscr{K}, \sigma, w') \models \mathrm{B}_e(D)$, a contradiction. Overall, we have (9.3).

It remains to show that $w^{\backslash \mathrm{B}_i} \subseteq w'$ and $w^{\backslash \mathrm{B}_i} \subseteq w'^{\backslash \mathrm{B}_i}$ follow from the consistency of $\varphi_w \wedge \neg\mathrm{B}_i(\neg\varphi_{w'})$. There are two cases to be considered. Consider case (a): $E \notin w'$ for some $\mathrm{B}_i(E) \in w$. Then $\neg E \in w'$. Thus, $\vdash_{\mathscr{CS}} E \supset \neg\varphi_{w'}$. Then $\vdash_{\mathscr{CS}} \mathrm{B}_i(E) \supset \mathrm{B}_i(\neg\varphi_{w'})$, which contradicts that $\varphi_w \wedge \neg\mathrm{B}_i(\neg\varphi_{w'})$ is consistent, since $\mathrm{B}_i(E) \in w$. Next, consider case (b): $\mathrm{B}_i(E) \notin w'$ for some $\mathrm{B}_i(E) \in w$. Then $\neg\mathrm{B}_i(E) \in w'$. Then $\vdash_{\mathscr{CS}} \mathrm{B}_i(E) \supset \neg\varphi_{w'}$. This implies $\vdash_{\mathscr{CS}} \mathrm{B}_i\mathrm{B}_i(E) \supset \mathrm{B}_i(\neg\varphi_{w'})$, which implies $\vdash_{\mathscr{CS}} \mathrm{B}_i(E) \supset \mathrm{B}_i(\neg\varphi_{w'})$. Again, we have a contradiction to the consistency of $\varphi_w \wedge \neg\mathrm{B}_i(\neg\varphi_{w'})$. After all, neither (a) nor (b) holds. Thus, $w^{\backslash \mathrm{B}_i} \subseteq w'$ and $w^{\backslash \mathrm{B}_i} \subseteq w'^{\backslash \mathrm{B}_i}$. $\square$

10 Conclusion

We have discussed both proof-theoretic and model-theoretic developments of epistemic logics and their applications to game theory. The author intended to show that the paper gives basic ideas of the logical approach and its scope, and hopes that this is successful.

The paper itself covers a lot of basic concepts, but does not talk about many relevant areas related to the logical approach. For example, extensions, such as predicate logics, of epistemic logics and their applications to economics are natural problems. Another related problem is the emergence of true or false beliefs from different sources such as individual experiences. The introduction of bounds to intrapersonal and interpersonal introspections for decision making is another problem. Although these must be targets of the logical approach, the present state of the logical approach is to wait for further research on these problems.

The reader who want to study those areas or to do research in the logical approach to economics and game theory may consult the papers in this issue as well as their references.

Over all, the author hopes that the reader finds some interests in this new field.

11 Appendix: Proof of completeness for classical logic CL

Our formulation of CL is quite efficient as an axiomatization, but the cost for an efficient axiomatization is practically difficult to prove some steps for the completeness of CL. For example, the following three claims are needed, but they need a lot of tedious steps, which is given in the end of this appendix.

Lemma 11.1.(a): $\vdash_0 (\neg A \supset \neg B) \supset (B \supset A)$; **(b)**: $A \supset (B \supset C) \vdash_0 A \wedge B \supset C$; **(c)**: $A \wedge B \supset C \vdash_0 A \supset (B \supset C)$.

Lemma 11.2 (Deduction theorem). $\Gamma \cup \{A\} \vdash_0 B$ implies $\Gamma \vdash_0 A \supset B$.

Proof. Let P be a proof of B from $\Gamma \cup \{A\}$. We prove by induction on the tree structure of P from its leaves that $\Gamma \vdash_0 A \supset C$ for any C in P. Let C be a formula associated with a leaf of P. Then C is an instance of L1–L5 or a formula in Γ. In either case, $\Gamma \vdash_0 C$. Since $\vdash_0 C \supset (A \supset C)$ by L1, we have $\Gamma \vdash_0 A \supset C$.

Now, let C be a formula associated with a non-leaf node in P. We assume the induction hypothesis that the induction assertion holds for the upper formulae of C in P. We should consider three cases: MP, $\bigwedge$-Rule and $\bigvee$-Rule. We consider MP and $\bigwedge$-Rule.

Suppose that C is inferred from D and $D \supset C$ by MP. The induction hypothesis is that $\Gamma \vdash_0 A \supset D$ and $\Gamma \vdash_0 A \supset (D \supset C)$. By Lemma 3.1.(1) and $\bigwedge$-Rule, $\Gamma \vdash_0 A \supset A \wedge D$, and by Lemma 11.1.(b), $\Gamma \vdash_0 A \wedge D \supset C$. By Lemma 3.1.(2), we have $\Gamma \vdash_0 A \supset C$.

Suppose that $D \supset \bigwedge \Phi$ is inferred from $\{D \supset E : E \in \Phi\}$. The induction hypothesis is that $\Gamma \vdash_0 A \supset (D \supset E)$ for all $E \in \Phi$. By Lemma 11.1.(b), $\Gamma \vdash_0 A \wedge D \supset E$ for all $E \in \Phi$. By $\bigwedge$-Rule, $\Gamma \vdash_0 A \wedge D \supset \bigwedge \Phi$. By Lemma 11.1.(c), $\Gamma \vdash_0 A \supset (D \supset \bigwedge \Phi)$. □

Then we have the following lemma.

Lemma 11.3. $\Gamma \nvdash_0 A$ if and only if $\Gamma \cup \{\neg A\}$ is consistent in CL.

Proof. We denote $\Gamma \cup \{\neg A\}$ by Γ'. Suppose $\Gamma \vdash_0 A$. Then $\Gamma' \vdash_0 A$ and $\Gamma' \vdash_0 \neg A$. Hence $\Gamma' \vdash_0 (A \supset A) \supset A$ and $\Gamma' \vdash_0 (A \supset A) \supset \neg A$ by L1. Thus, $\Gamma' \vdash_0 (A \supset A) \supset \neg A \wedge A$. By Lemma 3.1.(1), we have $\Gamma' \vdash_0 \neg A \wedge A$, i.e., Γ' is inconsistent.

Suppose that Γ' is inconsistent. Then $\Gamma' \vdash_0 \neg C \wedge C$. By Lemma 11.2 and L4, we have $\Gamma \vdash_0 \neg A \supset \neg C$ and $\Gamma \vdash_0 \neg A \supset C$. These together with L3 imply $\Gamma \vdash_0 A$. □

Proof of the Equivalence of the if Parts of Theorem 3.3. Suppose the *if* part of (1). We prove the negative form that of (2). Suppose that there is no model κ of Γ. Then $\Gamma \vDash \neg C \wedge C$ for some C. By (1), we have $\Gamma \vdash_0 \neg C \wedge C$.

Suppose the *if* part of (2). Let $\Gamma \nvdash_0 A$. Then $\Gamma \cup \{\neg A\}$ is consistent with respect to $\vdash_0$ by Lemma 11.3. By (2), there is a model κ of $\Gamma \cup \{\neg A\}$. This means $V_\kappa(A) = \bot$. Hence $\Gamma \nvDash A$. □

Proof of Completeness for CL. It suffices to prove that if Γ is consistent, there is an assignment κ such that $V_\kappa(B) = \top$ for all $B \in \Gamma$. In the following, we suppose that Γ is consistent in CL.

We order the set $\mathscr{P}$ as follows: $A_1, A_2, ...$ We construct a sequence $\Gamma_0, \Gamma_1, ...$ by induction on $A_1, A_2, ...$ as follows:

G-0: $\Gamma_0 = \Gamma$;

and for any $m > 0$,

G-1: $\Gamma_m = \begin{cases} \Gamma_{m-1} \cup \{A_m\} & \text{if } \Gamma_{m-1} \cup \{A_m\} \text{ is consistent} \\ \Gamma_{m-1} & \text{if } \Gamma_{m-1} \cup \{A_m\} \text{ is inconsistent.} \end{cases}$

We define $\tilde{\Gamma} = \bigcup_m \Gamma_m$.

Lemma 11.4.
(1): Each Γ_m is consistent and $\Gamma_0 \subseteq \Gamma_1 \subseteq ...$;

(2): $\tilde{\Gamma}$ is consistent.

Proof.
(1): This follows the definition of Γ_m.
(2): Suppose that $\tilde{\Gamma}$ is inconsistent. Then there is a proof of $\neg A \wedge A$ from a finite subset Γ' of $\tilde{\Gamma}$. This implies $\neg A \wedge A \in \Gamma_m$ for some m, a contradiction to (1). □

Lemma 11.5. The set $\tilde{\Gamma}$ defined above is maximally consistent, i.e., there is no other consistent set strictly including $\tilde{\Gamma}$, and satisfies the following properties:

(0): $\tilde{\Gamma} \vdash_0 A \Leftrightarrow A \in \tilde{\Gamma}$;

(1): either $A \in \tilde{\Gamma}$ or $\neg A \in \tilde{\Gamma}$;

(2): $A \supset B \in \hat{\Gamma} \Leftrightarrow A \notin \tilde{\Gamma}$ or $B \in \tilde{\Gamma}$;

(3): $\bigwedge \Phi \in \tilde{\Gamma} \Leftrightarrow A \in \tilde{\Gamma}$ for all $A \in \Phi$;

(4): $\bigvee \Phi \in \tilde{\Gamma} \Leftrightarrow A \in \tilde{\Gamma}$ for some $A \in \Phi$.

Proof. The maximality of $\tilde{\Gamma}$ follows from the definition of each Γ_m.

(0): $\Rightarrow$: immediate. The other direction follows from the definition of $\tilde{\Gamma}$.

(1): It is impossible that both A and $\neg A$ are in $\tilde{\Gamma}$ by Lemma 11.4.(2). If $A = A_m \notin \tilde{\Gamma}$, then $\Gamma_{m-1} \cup \{A_m\}$ is inconsistent. By Lemma 11.3, we have $\Gamma_{m-1} \vdash_0 \neg A_m$. By (0), we have $\neg A \in \tilde{\Gamma}$. The other case is symmetric.

(2): Let $A \supset B \in \hat{\Gamma}$. If $A \in \tilde{\Gamma}$, then $B \in \tilde{\Gamma}$ by MP. Conversely, let $A \notin \tilde{\Gamma}$ or $B \in \tilde{\Gamma}$. First, consider the case: $B \in \tilde{\Gamma}$. Then by L1, we have $A \supset B \in \hat{\Gamma}$. Second, let $A \notin \tilde{\Gamma}$. Then $\neg A \in \tilde{\Gamma}$ by (1). Since $\neg A \wedge A \notin \tilde{\Gamma}$, we have $\neg(\neg A \wedge A) \in \tilde{\Gamma}$. Hence $\neg B \supset \neg(\neg A \wedge A) \in \tilde{\Gamma}$ by L1. By Lemma 11.1.(a), we have $\neg A \wedge A \supset B \in \tilde{\Gamma}$. This is equivalent to $\neg A \supset (A \supset B) \in \tilde{\Gamma}$. Hence $(A \supset B) \in \tilde{\Gamma}$.

(3): Suppose $\bigwedge \Phi \in \tilde{\Gamma}$. Since $\vdash_0 \bigwedge \Phi \supset A$ for all $A \in \Phi$, we have $\tilde{\Gamma} \vdash_0 A$ for all $A \in \Phi$.

Conversely, suppose that $A \in \tilde{\Gamma}$ for all $A \in \Phi$. By (0), $\tilde{\Gamma} \vdash_0 A$ for all $A \in \Phi$. By L1, we have $\tilde{\Gamma} \vdash_0 (C \supset C) \supset A$ for all $A \in \Phi$. Hence $\tilde{\Gamma} \vdash_0 (C \supset C) \supset \bigwedge \Phi$. **(4)**: Suppose $A \notin \tilde{\Gamma}$ for all $A \in \Phi$. Let C be any formula. By (2), $A \supset C \in \tilde{\Gamma}$ and $A \supset \neg C \in \tilde{\Gamma}$ for all $A \in \Phi$. By (0) and $\bigvee$-Rule, we have $\bigvee \Phi \supset C \in \tilde{\Gamma}$ and $\bigvee \Phi \supset \neg C \in \tilde{\Gamma}$. By L3, we have $\neg \bigvee \Phi \in \Gamma$. Conversely, suppose that $A \in \tilde{\Gamma}$ for some $A \in \Phi$. By L5, we have $\bigvee \Phi \in \tilde{\Gamma}$. □

Now we can define an assignment κ by $\kappa(p) = \top$ iff $p \in \tilde{\Gamma}$. Using this κ, we have V_κ by E0–E4 of Section 4. It remains to prove $V_\kappa(A) = \top$ for all $A \in \Gamma$. For this purpose, it suffices to show that $V_\kappa(A) = \top$ if and only if $A \in \tilde{\Gamma}$. This is proved by induction on the structure of a formula: Lemma 11.5 is used for the following steps.

(1): for any $p \in PV$, $V_\kappa(p) = \top \Leftrightarrow \kappa(p) = \top \Leftrightarrow A \in \tilde{\Gamma}$;
(2): $V_\kappa(\neg A) = \top \Leftrightarrow V_\kappa(A) = \bot \Leftrightarrow A \notin \tilde{\Gamma} \Leftrightarrow \neg A \in \hat{\Gamma}$;
(3): $V_\kappa(A \supset B) = \top \Leftrightarrow V_\kappa(A) = \bot$ or $V_\kappa(B) = \top \Leftrightarrow A \notin \tilde{\Gamma}$ or $B \in \hat{\Gamma} \Leftrightarrow A \supset B \in \hat{\Gamma}$;
(4): $V_\kappa(\bigwedge \Phi) = \top \Leftrightarrow V_\kappa(A) = \top$ for all $A \in \Phi \Leftrightarrow A \notin \tilde{\Gamma}$ for all $A \in \Phi \Leftrightarrow \bigwedge \Phi \in \hat{\Gamma}$;
(5): $V_\kappa(\bigvee \Phi) = \top \Leftrightarrow V_\kappa(A) = \top$ for some $A \in \Phi \Leftrightarrow A \in \tilde{\Gamma}$ for some $A \in \Phi \Leftrightarrow \bigvee \Phi \in \hat{\Gamma}$. □

Proof of Lemma 11.1.[25] (a),(b) and (c) are proved as (7), (18) and (19). In the following, we use Lemma 3.1.(2) without mentioning.

(1): $\vdash_0 (B \supset C) \supset ((A \supset B) \supset (A \supset C))$.
*): Since $(B \supset C) \supset (A \supset (B \supset C))$ and $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$ are instances of L1 and L2, we have, by Lemma 3.1.(2), $\vdash_0 (B \supset C) \supset ((A \supset B) \supset (A \supset B))$.

(2): $A \supset (B \supset C) \vdash_0 B \supset (A \supset C)$.
*): By L2, we have $A \supset (B \supset C) \vdash_0 (A \supset B) \supset (A \supset C)$. This together with $B \supset (A \supset B)(-\text{L1})$ implies $A \supset (B \supset C) \vdash_0 B \supset (A \supset C)$ by Lemma 3.1.(2).

(3): $\vdash_0 (A \supset B) \supset ((B \supset C) \supset (A \supset C))$.
*): Regarding $B \supset C$, $A \supset B$ and $A \supset C$ as A, B and C of (2), we have, by (1), $\vdash_0 (A \supset B) \supset ((B \supset C) \supset (A \supset C))$.

(4): $\vdash_0 A \supset ((A \supset B) \supset B)$.
*): Regarding $A \supset B$ and A as A and B of (2), we have, using Lemma 3.1.(1), $\vdash_0 A \supset ((A \supset B) \supset B))$.

(5): $\vdash_0 (A \supset (A \supset B)) \supset (A \supset B)$.
*): Since $(A \supset ((A \supset B) \supset B)) \supset ((A \supset (A \supset B)) \supset (A \supset B))$ is an instance of L2 and $\vdash_0 A \supset ((A \supset B) \supset B)$ by (4), we have $\vdash_0 (A \supset (A \supset B)) \supset (A \supset B)$.

(6): $\vdash_0 \neg\neg A \supset A$ (the law of double negation).
*): Since $\neg\neg A \supset (\neg A \supset \neg\neg A)$ and $(\neg A \supset \neg\neg A) \supset ((\neg A \supset \neg A) \supset A)$ are instance of L1 and L3, respectively, we have, by Lemma 3.1.(2), $\vdash_0 \neg\neg A \supset$

[25] The following proof is due to T. Nagashima.

$((\neg A \supset \neg A) \supset A)$. Hence we have, by (2), $\vdash_0 (\neg A \supset \neg A) \supset (\neg\neg A \supset A)$. By Lemma 3.1.(1), we have $\vdash_0 \neg\neg\, A \supset A$.

(7): $\vdash_0 (\neg A \supset \neg B) \supset (B \supset A)$.
*) Since $\vdash_0 B \supset (\neg A \supset B)$ by L1 and $\vdash_0 [B \supset (\neg A \supset B)] \supset [((\neg A \supset B) \supset A) \supset (B \supset A)]$ by (3), we have $\vdash_0 ((\neg A \supset B) \supset A) \supset (B \supset A)$. This and L3 imply $\vdash_0 (\neg A \supset \neg B) \supset (B \supset A)$.

(8): $\vdash_0 A \supset \neg\neg A$ (the converse of (6)).
*): Since $\vdash_0 \neg\neg\neg A \supset \neg A$ by (6) and $\vdash_0 (\neg\neg\neg A \supset \neg A) \supset (A \supset \neg\neg A)$ by (7), we have $\vdash_0 A \supset \neg\neg A$.

(9): $\vdash_0 (\neg A \supset B) \supset (\neg B \supset A)$.
*) Since $\vdash_0 (B \supset \neg\neg B) \supset ((\neg A \supset B) \supset (\neg A \supset \neg\neg B))$ by (1) and $\vdash_0 B \supset \neg\neg B$ by (8), we have $\vdash_0 (\neg A \supset B) \supset (\neg A \supset \neg\neg B)$. This together with $\vdash_0 (\neg A \supset \neg\neg B) \supset (\neg B \supset A)$ by (7) implies $\vdash_0 (\neg A \supset B) \supset (\neg B \supset A)$.

(10): $\vdash_0 (A \supset B) \supset (\neg B \supset \neg A)$.
*): Since $\vdash_0 (B \supset \neg\neg B) \supset ((A \supset B) \supset (A \supset \neg\neg B))$ by (1) and $\vdash_0 B \supset \neg\neg B$ by (8), we have $\vdash_0 (A \supset B) \supset (A \supset \neg\neg B)$. This together with $\vdash_0 (A \supset \neg\neg B) \supset (\neg B \supset \neg A)$ by (7) implies $\vdash_0 (A \supset B) \supset (\neg B \supset \neg A)$.

(11): $\vdash_0 (A \supset \neg B) \supset (B \supset \neg A)$.
*) Since $\vdash_0 (\neg\neg A \supset A) \supset ((A \supset \neg B) \supset (\neg\neg A \supset \neg B))$ by (3) and $\vdash_0 \neg\neg A \supset A$ by (6), we have $\vdash_0 (A \supset \neg B) \supset (\neg\neg A \supset \neg B)$. This and (7) imply $\vdash_0 (A \supset \neg B) \supset (B \supset \neg A)$.

(12) $\vdash_0 A \supset (B \supset \neg(A \supset \neg B))$.
*) Since $\vdash_0 A \supset ((A \supset \neg B) \supset \neg B)$ by (4) and $\vdash_0 ((A \supset \neg B) \supset \neg B) \supset (B \supset \neg(A \supset \neg B))$ by (11), we have $\vdash_0 A \supset (B \supset \neg(A \supset \neg B))$.

(13): $\vdash_0 \neg A \supset (A \supset B)$.
*) Since $\neg A \supset (\neg B \supset \neg A)$ is an instance of L1, we have, by (7), $\vdash_0 \neg A \supset (A \supset B)$.

(14): $\vdash_0 \neg(A \supset \neg B) \supset A$.
*): Since $\vdash_0 \neg A \supset (A \supset \neg B)$ by (13) and $\vdash_0 (\neg A \supset (A \supset \neg B)) \supset (\neg(A \supset \neg B) \supset A)$ by (9), we have $\vdash_0 \neg(A \supset \neg B) \supset A$.

(15): $\vdash_0 \neg(A \supset \neg B) \supset B$.
*): Since $\vdash_0 \neg B \supset (A \supset \neg B)$ by L1 and $\vdash_0 (\neg B \supset (A \supset \neg B)) \supset (\neg(A \supset \neg B) \supset B)$ by (9), we have $\vdash_0 \neg(A \supset \neg B) \supset B$.

(16): $\vdash_0 \neg(A \supset \neg B) \supset A \wedge B$.
*): Using $\bigwedge$-Rule, it follows from (14) and (15) that $\vdash_0 \neg(A \supset \neg B) \supset A \wedge B$.

(17): $\vdash_0 A \supset (B \supset A \wedge B)$.
*) Since $\vdash_0 (\neg(A \supset \neg B) \supset A \wedge B) \supset [(B \supset \neg(A \supset \neg B)) \supset (B \supset A \wedge B)]$ by (1), we have $\vdash_0 (B \supset \neg(A \supset \neg B)) \supset (B \supset A \wedge B)$ using (16). This and (12) imply $\vdash_0 A \supset (B \supset A \wedge B)$.

(18): $A \supset (B \supset C) \vdash_0 A \wedge B \supset C$.
*): Since $\vdash_0 A \wedge B \supset A$ by L4, we have $A \supset (B \supset C) \vdash_0 A \wedge B \supset (B \supset C)$. By (2), $A \supset (B \supset C) \vdash_0 B \supset (A \wedge B \supset C)$. This together with $\vdash_0 A \wedge B \supset B$ by L4

implies $A \supset (B \supset C) \vdash_0 A \wedge B \supset (A \wedge B \supset C)$, so $A \supset (B \supset C) \vdash_0 A \wedge B \supset C$ by (5).

(19): $A \wedge B \supset C \vdash_0 A \supset (B \supset C)$.
*) Since $\vdash_0 (A \wedge B \supset C) \supset ((B \supset A \wedge B)) \supset (B \supset C))$ by (1) and $A \wedge B \supset C$ is an assumption, we have $A \wedge B \supset C \vdash_0 (B \supset A \wedge B) \supset (B \supset C)$. This and (17) imply $A \wedge B \supset C \vdash_0 A \supset (B \supset C)$. □

References

1. Aumann, R. J.: Agreeing to disagree. Annals of Statistics **4**, 1236–1239 (1976)
2. Aumann, R. J.: Interactive epistemology I: knowledge. International Journal of Game Theory **28**, 261–300 (1999)
3. Bacharach, M., Mongin, P.: Epistemic logic and the foundations of game theory. Theory and Decision **37**, 1–6 (1994)
4. Bacharach, M., Gérard-Varet, L. A., Mongin, P., Shin, H. S.: Epistemic logic and the theory of games and decisions. Boston: Kluwer 1997
5. Boolos, G.: The unprovability of consistency. Cambridge: Cambridge University Press 1979
6. Chellas, B.: Modal logic. Cambridge: Cambridge University Press 1980
7. Fagin, R., Halpern, J. Y., Moses, Y., Verdi, M. Y.: Reasoning about knowledge. Cambridge: The MIT Press 1995
8. Halpern, J. H., Moses, Y.: A guide to completeness and complexity for modal logics of knowledge and beliefs. Artificial Intelligence **54**, 319–379 (1992)
9. Heifetz, A.: Iterative and fixed point common belief. Journal of Philosophical Logic **28**, 61–79 (1999)
10. Hintikka, J.: Knowledge and belief. Ithaca: Cornell University Press 1962
11. Gentzen, G.: Untersuchungen über das logische Schliessen. Mathematische Zeitschrift **39**, 176–210, 405–431 (1935). English translation: Investigations into logical deduction. In: Szabo, M. E. (ed.)The Collected Papers of Gerhard Gentzen. Amsterdam: North-Holland 1969
12. Hughes, G. E., Cresswell, M. J.: A companion to modal logic. London: Methuen 1984
13. Kaneko, M.: Epistemic considerations of decision making in games. Mathematical Social Sciences **38**, 105–137 (1999)
14. Kaneko, M.: Common knowledge logic and game logic. Journal of Symbolic Logic **64**, 685–700 (1999)
15. Kaneko, M., Matsui, A.: Inductive game theory: discrimination and prejudices. Journal of Public Economic Theory **1**, 101–137 (1999)
16. Kaneko, M., Nagashima, T.: Game logic and its applications. Part I: Studia Logica **57**, 325–354 (1996). Part II: Studia Logica **58**, 273–303 (1997)
17. Kaneko, M., Nagashima, T.: Axiomatic indefinability of common knowledge in finitary logics. In: Bacharach, M. et al. (eds.) Epistemic logic and the theory of games and decision, pp. 69–93. Boston: Kluwer 1997
18. Kaneko, M., Suzuki, N.-Y.: Epistemic models of shallow depths and decision making in games. IPPS. DP. No. 828, University of Tsukuba (1999)
19. Kaneko, M., Suzuki, N.-Y.: Bounded interpersonal inferences and decision making. Economic Theory (this issue)
20. Kaneko, M., Nagashima, T., Suzuki, N.-Y., Tanaka, Y.: A map of common knowledge logics. Studia Logica (to appear) (2001)
21. Kleene, S. C.: Mathematical logic. New York: Wiley 1967
22. Lismont, L. Mongin, P.: On the logic of common belief and common Knowledge. Theory and Decision **37**, 75–106 (1994)
23. Limont. L., Mongin, P.: On the logic of common belief and common knowledge. In: Bacharach, M. et al. (eds.) Epistemic logic and the theory of games and decision, pp. 3–34. Boston: Kluwer 1997
24. Mendelson, E.: Introduction to mathematical logic. Monterey: Wadsworth 1987

25. Meyer, J.-J. Ch., van der Hoek, W.: Epistemic logic for AI and computer science. Cambridge: Cambridge University Press 1995
26. Mithen, S.: The prehistory of the mind. London: Thames and Hudson 1996
27. Moser, P. K.: Empirical knowledge – readings in contemporary epistemology. Chicago: Rowman & Littlefield 1986
28. Moulin, H.: Game theory for the social sciences. New York: New York University Press 1982
29. Myerson, R. B.: Game theory. Cambridge: Harvard University Press 1991
30. Ohnishi, M., Matsumoto, K.: Gentzen method in modal calculi I. Osaka Mathematical Journal **9**, 113–130 (1957)
31. Nash, J. F.: Noncooperative game. Annals of Mathematics **54**, 286–295 (1951)
32. Plato: The Republic of Plato. Translated by: Cornford, F. M. London: Oxford University Press 1941
33. Schulte O.: Minimal belief change, Pareto-optimality and logical consequence. Economic Theory (this issue)
34. Takeuti, G.: Proof theory. Amsterdam: North-Holland 1987
35. Terutoshi, Y., Okitsu, K., Enomoto, S. (eds.): Collections of Meiji-Taisho comic stories (in Japanese), Vol. 3. Tokyo: Kodansha 1980
36. Wolter, F.: First order common knowledge Logics. Studia Logica **65**, 249–271 (2000)
37. van Dalen, D.: Intuitionistic logic. In: Gabbay, D., Guenthner, F. (eds.) Handbook of philosophical logic, Vol. III. Amsterdam: Reidel 1986