NOTES ON BOUNDED RATIONALITY

Mariko Yasugi * and Sobei H. Oda †

Received January 16, 2002; revised April 17, 2002

Abstract. Rubinstein has defined the *information structure* in order to study models of *bounded rationality*. He shows that three properties of information structures denoted by P-1, P-2 and P-3 respectively imply satisfiability of the axioms A-1, A-2 and A-3, which are known as **T**, **4** and **5** in modal logic. We wish to add some more facts concerning them. It will be shown that the axiom A-i determines the property P-i, and that each axiom has a kind of a free information structure in terms of formulas of the propositional calculus. It is also shown that the so-called "two wise girls puzzle," which can be formulated in terms of A-2, is "solvable" by means of proof-theoretic arguments. We also propose to consider the "non-knowledge" operator.

1 Introduction Rubinstein in [8] proposed to study models of bounded rationality and presented some results he had obtained on this subject. His intention was to propose a counter approach to the "perfect rational man paradigm" (according to his expression).

In Chapter 3 of [8], he defines the information structure and discusses the three properties of information structures denoted by P-1,P-2 and P-3, and then shows that each of these properties implies satisfiablity of an axiom on a knowledge operator with respect to Kripke semantics. These axioms, denoted by A-1, A-2 and A-3, are known respectively as **T**, **4** and **5** in modal logic.

In this paper, we wish to add some facts with respect to these properties and axioms, namely to establish that the axiom A-i determines the property P-i for an information structure (Section 3) and to give a kind of a free information structure in terms of formulas of the propositional calculus and their provability/unprovability (Section 4).

In our previous works, we developed proof-theoretic treatments of the so-called "three wise men puzzle." (We modified it to a "two wise girls puzzle.") It is formulated in a system of modal logic known as **S4**, whose non-propositional axiom is A-2 in [8]. We will also add some facts about this treatment, that is, we will show that the framework of the proof-theory for **S4** is decidable, and hence we can claim that the puzzle in question is solvable (Section 5).

In most of treatment of a knowledge operator, the main subject has been how to interpret "one knows." However, in a puzzle as above, how to interpret "one does not know" is an important issue. In [11] and [12], we have discussed this matter. We would like to propose again to take up the "non-knowledge" operator (Section 6).

We will first review Rubinstein [8] concerning the information structure (Section 2).

We will not go into detail of proof-theory of the sequential calculi and various facts on modal logic. See, for example, [2] for the former and [1], [3], [4] and [7] for the latter. For our treatment of the wise men puzzle, see [5], [6], [10], [11] and [12]. See also [9].

2000 *Mathematics Subject Classification.* 91A26, 03B42.

Key words and phrases. Bounded rationality, information structure, knowledge operator, non-knowledge operator, formal systems of knowledge/non-knowledge, free structure, wise girls puzzle, proof-theory.

*The corresponding author. This work is partly supported by Scientific Grant-in-Aid (No.12878061).

†This work is partly supported by Scientific Grant-in-Aid (No.13480115).

2 Review: information structure We will briefly describe information structures associated with bounded rationality and related axioms according to Section 3.2 of [8].

Let Ω be a nonempty set (of states), and let P be a function from Ω to nonempty subsets of Ω. The pair (Ω, P) is called an *information structure* (of bounded rationality). Elements of Ω are denoted by ω, ω' etc.

We will quote three conditions of information structures associated with bounded rationality, from Section 3.2 of [8].

P-1 $\omega \in P(\omega)$

P-2 If $\omega' \in P(\omega)$, then $P(\omega') \subset P(\omega)$

P-3 If $\omega' \in P(\omega)$, then $P(\omega') \supset P(\omega)$

An information structure (Ω, P) is called *partitional* if there exists a partition of Ω such that $P(\omega)$ is a component and $\omega \in P(\omega)$ for each ω.

Note If an information structure satisfies P-1 and P-3, then it also satisfies P-2.

In Section 3.4 of [8], the language of the propositional calculus with a knowledge operator K is introduced. Let us call this language $\mathcal{L}_K$. KA represents "(an agent) knows that A holds."

The following are three axioms on the knowledge operator K of bounded rationality, corresponding to the three implications in Proposition 3.2 of [8].

A-1 $KF \Rightarrow F$

A-2 $KF \Rightarrow K(KF)$

A-3 $\neg KF \Rightarrow K(\neg KF)$

for an arbitrary $\mathcal{L}_K$-formula F. These are known to be respectively $\mathbf{T}, \mathbf{4}$ and $\mathbf{5}$ in modal logic.

3 Formal systems and semantics We will define logical systems with a knowledge operator of bounded rationality corresponding to the three axioms A-1, A-2 and A-3.

In the following, Γ and Δ denote finite sequences of formulas, possibly empty, and Θ denotes either a single formula or the emptiness.

Definition 3.1 (Logical system) The base logic with a knowledge operator will be called BKL, abbreviating "basic knowledge logic." BKL is the *classical propositional logic* with the following basic knowledge inference, (BK), added. (We will employ the sequential formulation of the system: see [2], for example.)

$$\frac{\Gamma \to \Theta}{K\Gamma \to K\Theta}\ (BK)$$

where $K\Gamma$ denotes $KA_1, KA_2, \cdots, KA_n$ if Γ is $A_1, A_2, \cdots, A_n$.

The three additional inference rules $(K1)$, $(K2)$ and $(K3)$, each corresponding to one of A-1, A-2 and A-3, are the following.

$$\frac{\Gamma, \Delta \to \Theta}{K\Gamma, \Delta \to \Theta}\ (K1)$$

$$\frac{\Gamma, K\Delta \to \Theta}{K\Gamma, K\Delta \to K\Theta}\ (K2)$$

In order to formulate the third inference rule, it is convenient to introduce a symbol N defined as follows.

$$NA \equiv \neg KA$$

NA represents "(an agent) does not know that A holds."

$$\frac{\Gamma, N\Delta \rightarrow N\Theta}{K\Gamma, N\Delta \rightarrow KN\Theta}\ (K3)$$

With N, the axiom A-3 can be expressed as $NF \rightarrow KNF$.

The systems obtained from BKL with $(K1)$,$(K2)$ and $(K3)$ added will be respectively called $KL1$, $KL2$ and $KL3$.

The latter two are known respectively as **S4** and **S5**.

Proposition 3.1 (Formal equivalence of axiom and inference) The axiom A-i is equivalent to the inference (Ki) over the base logic BKL.

Proof The equivalences are generally known, but since $(K3)$ is not in the usual form, we will give a proof for A-3 and $(K3)$.

$NF \rightarrow KNF$ (A-3) is an immediate consequence of $(K3)$:

$$\frac{NF \rightarrow NF}{NF \rightarrow KNF}\ (K3)$$

$(K3)$ becomes a derived rule over BKL by virtue of A-3. For simplicity, let us assume $\Delta \equiv A_1, A_2$. Then $N\Delta$ represents NA_1, NA_2. Similarly with $KN\Delta$.

$$\frac{[N\Delta \rightarrow KN\Delta] \quad \dfrac{\Gamma, N\Delta \rightarrow N\Theta}{K\Gamma, KN\Delta \rightarrow KN\Theta}\ (BK)}{K\Gamma, N\Delta \rightarrow KN\Theta}\ (cuts)$$

The last part is an abbreviation of the cuts

$$\frac{NA_1 \rightarrow KNA_1 \quad K\Gamma, KN\Delta \rightarrow KN\Theta}{K\Gamma, NA_1, KNA_2 \rightarrow KN\Theta}\ (cut)$$

and

$$\frac{NA_2 \rightarrow KNA_2 \quad K\Gamma, NA_1, KNA_2 \rightarrow KN\Theta}{K\Gamma, N\Delta \rightarrow KN\Theta}\ (cut)$$

Definition 3.2 (Models: Section 3.4 of [8]) Let π be a map

$$\Omega \times \Phi \rightarrow \{True, False\},$$

where Φ denotes the set of formulas of $\mathcal{L}_K$. Write M_π for the triple (Ω, P, π), which is called a Kripke structure.

The relation $(M_\pi, \omega) \models F$ is defined as follows.

For an atomic F, this relation holds when $\pi(\omega, F) = True$. For the proositional connectives, the classical truth value assignment is at work. If F is of the form KG, then $(M_\pi, \omega) \models F$ holds when, for every $\omega' \in P(\omega)$, $(M_\pi, \omega') \models G$.

For a sequent, $(M_\pi, \omega) \models A_1, A_2, \cdots, A_m \rightarrow B_1, B_2, \cdots, B_n$ can be interpreted to be $(M_\pi, \omega) \models A_1 \wedge A_2 \wedge \cdots \wedge A_m \Rightarrow B_1 \vee B_2 \vee \cdots \vee B_n$.

We will also denote the relation that $(M_\pi, \omega) \models F$ holds for every π and for every ω by $(\Omega, P) \models F$, and will say that F is *valid in* (Ω, P), or (Ω, P) is a *model* of F. F is called *valid* if F is valid in every (Ω, P).

It is a straightforward practice to prove the following.

Proposition 3.2 (Semantic consistency of BKL) Every theorem of BKL is valid, and hence BKL is consistent.

In Proposition 3.2, Section 3.4 of [8], it is claimed that the model condition P-i satisfies the axiom A-i, $i = 1, 2, 3$.

We wish to claim the converse.

Proposition 3.3 (Axioms determine model properties) If an information structure (Ω, P) is a model of the axiom A-i, then (Ω, P) satisfies the condition P_i.

Proof The proof technique is the same for all three axioms, that is, the axiom applied to an atomic formula (propositional variable), say X, determines the model property.

A-1: Suppose $(M_\pi, \omega) \models KX \Rightarrow X$ for every X, π and ω.

Fix any ω, and define, for this ω and for every propositional variable X, $\pi_\omega(\omega, X) = False$ and $\pi_\omega(\omega', X) = True$ if $\omega \neq \omega'$.

For any ω, π, X, (*) $(M_\pi, \omega) \models KX \Rightarrow X$ means

$$\forall \omega' \in P(\omega)(M_\pi, \omega') \models X \quad \text{implies} \quad (M_\pi, \omega) \models X,$$

and $(M_\pi, \omega) \models X$ means $\pi(\omega, X) = True$.

Letting π be the π_ω above, $\pi_\omega(\omega, X) = False$. So, for this assignment π_ω, (*) holds only if, for some $\omega' \in P(\omega)$, $\pi_\omega(\omega', X) = False$. According to the definition of π_ω, this is possible only if $\omega' = \omega$. This forces that $\omega \in P(\omega)$, or P-1.

A-2: Suppose $(M_\pi, \omega) \models KX \Rightarrow KKX$ for every π and ω.

This means that

$$\forall \omega' \in P(\omega)(M_\pi, \omega') \models X \quad \text{implies} \quad \forall \omega' \in P(\omega) \forall \omega'' \in P(\omega')(M_\omega, \omega'') \models X.$$

Take any ω, and define

$$\pi_\omega(\omega', X) = True \quad \text{if} \quad \omega' \in P(\omega); = False \quad \text{otherwise} \quad .$$

For this π_ω, $\forall \omega' \in P(\omega)(M_{\pi_\omega}, \omega') \models X$ holds, and so it must be the case that

$$\forall \omega' \in P(\omega) \forall \omega'' \in P(\omega')(M_\omega, \omega'') \models X.$$

For any $\omega' \in P(\omega)$ and $\omega'' \in P(\omega')$, $(M_{\pi_\omega}, \omega'') \models X$, or $\pi_\omega(\omega'', X) = True$ can hold only if $\omega'' \in P(\omega)$, and so $P(\omega') \subset P(\omega)$, or P-2.

A-3: Suppose, for every π and for every ω, $(M_\pi, \omega) \models \neg KX \Rightarrow K\neg KX$. This means that,

$$\exists \omega' \in P(\omega) \pi(\omega', X) = False.$$

$$\text{implies} \quad \forall \omega^* \in P(\omega) \exists \omega'' \in P(\omega^*) \pi(\omega'', X) = False$$

Now, for any ω and for any $\omega^* \in P(\omega)$, fixed, define

$$\pi_{\omega,\omega^*}(\omega', X) = True \quad \text{if} \quad \omega' \in P(\omega^*); = False \quad \text{otherwise}$$

For this π_{ω,ω^*}, there is an $\omega_0 = \omega^*$ such that

$$\forall \omega'' \in P(\omega^*), \pi(\omega'', X) = True,$$

and so the conclusion of the supposition is refuted. So, in order that the supposition hold, it is forced that, for all $\omega' \in P(\omega)$, $\pi_{\omega,\omega^*}(\omega', X) = True$. The value can be $True$ only if $\omega' \in P(\omega^*)$, that is, $\omega' \in P(\omega)$ implies $\omega' \in P(\omega^*)$, and hence $P(\omega) \subset P(\omega^*)$, or P-3.

According to Proposition 3.1 in [8], an information structure (Ω, P) is partitional if and only if it satisfies all P-1, P-2 and P-3. With this fact, Proposition 3.2 in [8] and Proposition 3.3, we can claim that an information structure (Ω, P) is partitional if and only if it satisfies all the axioms A-1, A-2 and A-3.

4 Free structure We will construct a kind of free information structure (Ω, P_i) for each axiom A-i (or condition P-i), consisting of propositional formulas (without knowledge operator) and determined by provability/unprovability in the classical propositional calculus.

We regard each structure below as "free" for the reason that it is constructed purely syntactically using provability/unprovability of formulas.

Let Ω be the set of all formulas of the propositional calculus, with propositional variables prepared. We define a function $P_i : \Omega \to \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ denotes the power set of Ω, as follows..

P_i will satisfy P-i but not P-j for $j \neq i$.

P-1: $P_1(A) = \{B | \vdash A \to B \quad \text{or} \quad \vdash B \to A\}$, where $\vdash$ expresses the provability in the propositional calculus.

It is obvious that $A \in P_1(A)$, and hence P-1 holds.

For any distinct propositional variables X, Y and Z, $X \vee Y \in P_1(X)$ and $Y \in P_1(X \vee Y)$, but $Y \notin P_1(X)$. So, $P_1(X \vee Y) \not\subset P_1(X)$. That is, P-2 does not hold.

$X \vee Z \in P_1(X)$ but $X \vee Z \notin P_1(X \vee Y)$, and so $P_1(X) \not\subset P_1(X \vee Y)$, hence P-3 does not hold.

P-2: $P_2(A) =$the set of all theorems of the propositional calculus if A is a theorem.

$P_2(A) = \{B | \vdash A \to B \quad \text{but not} \quad \vdash B \to A\}$ if A is not a theorem.

In either case, every theorem belongs to $P_2(A)$.

Since $\vdash X \to X$, $X \notin P_1(X)$, and so P-1 does not hold.

Suppose A is not a theorem and $B \in P_2(A)$. That is, $\vdash A \to B$ and not $\vdash B \to A$. If $C \in P_2(B)$, then $\vdash B \to C$, and so $\vdash A \to C$. If $\vdash C \to A$, then $\vdash B \to A$, yielding a contradiction. So, $C \in P_2(A)$. This proves $P_2(B) \subset P_2(A)$. If A is a theorem, then $B \in P_2(A)$ if and only if B is a theorem, and so $P_2(A) = P_2(B)$. So P-2 holds.

For any distinct X and Y, $X \vee Y \in P_2(X)$, and $X \vee Y \notin P_2(X \vee Y)$, and so P-3 does not hold.

P-3: If A is a contradiction ($\vdash \neg A$), then

$$P_3(A) = \text{the set of all theorems.}$$

If A is not a contradiction (called consistent), then

$$P_3(A) = \{B | B \text{ is consistent }\}.$$

According to this definition, if A and B are each consistent, then $P_3(A) = P_3(B)$. If A is consistent, then $P_3(A)$ strictly contains all theorems. Note that a theorem is consistent.

Suppose A is a contradiction and $B \in P_3(A)$. Then B is a theorem, and hence

$$P_3(B) \supset P_3(A).$$

If A is consistent and $B \in P_3(A)$, then B is consistent, and hence $P_3(B) = P_3(A)$. So, P-3 holds.

For a contradiction A, $A \notin P_3(A)$, hence P-1 does not hold. For a contradiction A, if $B \in P_3(A)$, then B is consistent, and so $P_3(B)$ strictly contains $P_3(A)$, hence $P_3(B) \not\subset P_3(A)$, and so P-2 does not hold.

At the end, we will define a partition:

$$P_0(A) = \{B | \vdash A \leftrightarrow B\}$$

This P_0 defines an equivalence class of formulas with respect to $\vdash A \leftrightarrow B$. $A \in P_0(A)$, and $B \in P_0(A)$ if and only if $P_0(B) = P_0(A)$. Obviously P is partitional, since it satisfies all of P-1, P-2 and P-3 (cf. Proposition 3.1 in [8]).

Let us call the system BKL augmented by all the three axioms A-1, A-2 and A-3 (or, equivalently, all inferences (K1), (K2) and (K3)) KL_0.

Proposition 4.1 (Consistency of KL_0) The system KL_0 is semantically consistent, and hence each system KLi is semantically consistent.

Proof The basic logic BKL is consistent with respect to any information structure by Proposition 3.2. Any partition satisfies all three axioms by virtue of Proposition 3.2 in Section 3.4 of [8], and hence KL_0 is consistent with respect to (Ω, P_0).

5 Two wise girls puzzle We will first briefly explain a puzzle (which we call "two wise girls puzzle").

Two girls are put on white hats on their heads; the first girl can see the second girl's hat but not her own, and the second girl cannot see either hat. The first girl is asked if her hat is white. She answers "I do not know if my hat is white." Then the second girl is asked the same question, and she answers "I know that my hat is white."

We assume that the logical ability of each girl is equivalent to the system $KL2$, with knowledge operators K_1 and K_2 for respectively the first girl and the second girl.

In order that the first girl can conclude that she does not know if her hat is white, she has to "jump out" of her logical system and study it from outside. This jumping out is formulated in terms of proof-theory. Details are seen in [5], [6], [10], [11] and [12]. In particular, the cut elimination theorem is known to hold for $KL2$ (cf. also [4] and [7]). We will not repeat the argument here, but will add one lemma as below.

Proposition 5.1 (Mono-conclusion lemma) Suppose $S : K_i\Gamma \to K_i\Delta$ is provable and Γ and Δ are K_i-free. Then there is a Ξ, which is either empty or a single formula in Δ, so that $K_i\Gamma \to K_i\Xi$ is provable.

Proof Consider a cut-free proof-figure of the sequent S. We prove the proposition by induction on the total number ν of formulas in Γ and Δ.

If ν is 1, then the proposition is obvious. When $\nu > 1$, there are three possibilities of obtaining S.

1. S is an initial sequent. Then, the proposition is obvious.

2. S is obtained by a thin, and the upper sequent is $K_i\Gamma' \to K_i\Delta'$. Then the number of formulas is $\nu - 1$, and hence the induction hypothesis applies. Thus, there is a Ξ as required so that $K_i\Gamma' \to K_i\Xi$ is provable. By applying "thin", if necessary, one obtains $K_i\Gamma \to K_i\Xi$.

3. S is obtained by (K_i2). Then, by the condition of the inference, the cardinality of Δ must be at most 1.

Using some lemmas, we showed, for example, that a claim such as "I do not know if my hat is white" can be interpreted as the unprovability of "I know my hat is white" in a logical system. Then we gave a formal proof of the second player's solution within her logical system.

Question of Solvability There arise two questions here. First, if the proof-theory we employ is transcendental, then it is unlikely that a player can claim unprovability of a statement. Second, not everybody can make correct inferences, and so the second player may not be able to construct a correct proof of her conclusion. Nonetheless, we claim that the puzzle is *solvable*. This claim can be interpreted as that judgements of both provability and unprovability are recursive; more preciesly, decidability of unprovability and automated theorem proving of a theorem.

This can be attained by means of the *resolution method*. The propositional part of the theorem proving (cheking) can be executed just as usual, say by Wang's algorithm. In resolving a sequent, only the new inference $(K2)$ has to be taken care of.

Outline of resolution method

The resolution of a formula in a propositional sequent can be regarded as the reverse operation of an inference. For example, the formula $A \Rightarrow B$ in a sequent $\Gamma \to \Theta, A \Rightarrow B$ can be resolved into $A, \Gamma \to \Theta, B$, which is just the reverse of the rule $(\to\Rightarrow)$. The sequent $A, \Gamma \to \Theta, B$ is called a resolution of $\Gamma \to \Theta, A \Rightarrow B$.

The resolution of $A \vee B$ in $\Gamma \to \Theta, A \vee B$ is $\Gamma \to \Theta, A, B$, and this is essentially the rule $(\to \vee)$.

The figure which starts with a given sequent S and consists of continuation of resolutions as described above will be called a resolution tree of S.

At each resolution, the number of logical connectives in a sequent decreases, and so eventually one obtains a sequent (called a top sequent) without connectives. Such a sequent is said to be valid if it contains a same propositional variable (or constant) in both sides of $\to$. It is known that a given sequent of the propositional calculus is provable if and only if all the top sequents in its resolution tree are valid. (We will then say that the tree *is valid.*) Furthermore, a proof-figure can be automatically created from the resolution tree.

In resolving a sequent in the system $KL2$, only the the case where no propositional resolution applies, that is, the case where the sequent is of the form $\Delta, K(\Gamma) \to K(\Theta), \Lambda$, where Δ and Λ consist of atomic formulas and they are mutually disjoint. (For simplicity, we will explain the resolution for a system with a single knowledge operator K.) A resolution of such a sequent looks like this:

$$\frac{\Gamma_1, K(\Gamma_2) \to \Theta^*}{\Delta, K(\Gamma) \to K(\Theta), \Lambda}\ (K2),$$

where Γ_1 and Γ_2 form a decomposition of Γ, Θ^* is a subset of Θ of cardinality at most 1 and either Γ_1 or Θ is non-empty.

There can be as many possibilities of upper sequents $\Gamma_1, K(\Gamma_2) \to \Theta^*$ as the number of decompositions of Γ and the choice of Θ^* from Θ, but in any case the number of the operator K decreases.

Once a $(K2)$-resolution applies, a tree forks into several trees according to decompositions of Γ and choices of Θ^*, so that there will be a forest of trees.

It is a lengthy but straightforward procedure to show the following. (For an interested reader, details of construction of resolution forests can be seen in Section 13 of [10].)

Proposition 5.2 (Resolution forest) If one of the trees in a forest of a sequent is valid, then it induces a proof-figure of the given sequent. Otherwise, the given sequent is unprovable. Construction of a resolution forest and checking algorithm of its validity are *decidable* procedures. (We will call this procedure a theorem checking.)

Note (1) In case a sequent is of the form

$$\Delta, \{K_i(\Gamma_i)\}_{1\leq i\leq n} \to \{K_i(\Theta_i)\}_{1\leq i\leq n}, \Lambda$$

a resolution will assume the form

$$\Gamma_i^1, K_i(\Gamma_i^2) \to \Theta_i^*$$

for some i.

(2) The $(K2)$ resolution is a reverse expression of Proposition 5.1, the mono-conclusion lemma.

As a way of an example, we will give a resolution-forest of $\Gamma \to K_2 2W$, where iW is a propositional constant representing "Player i knows her hat is white," and Γ denotes the knowledge set of Player 2. In order to make expressions simple, we will demonstrate a tree (in a forest) which turns out to be valid. (In a similar manner, one can construct a counter-example of a resolution tree of $\Gamma \to K_2\neg 2W$, which shows that this sequent cannot be provable.)

Let Π denote $1W \vee 2W$. Then the target sequent S_0 is

$$K_2K_1\Pi, K_2(\neg K_1 1W), K_2(\neg 2W \Rightarrow K_1\neg 2W) \to K_2 2W$$

The following is a desired tree of S_0.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Pi \to 1W, 2W}{\Pi, \neg 2W \to 1W}\,(\to \neg)}{K_1\Pi, K_1\neg 2W \to 2W, K_1 1W}\,(K_1 2)}{K_1\Pi, \neg K_1 1W, K_1\neg 2W \to 2W}\,(\neg \to) \qquad \dfrac{2W, K_1\Pi, \neg K_1 1W \to 2W}{K_1\Pi, \neg K_1 1W \to 2W, \neg 2W}\,(\to \neg)}{K_1\Pi, \neg K_1 1W, \neg 2W \Rightarrow K_1\neg 2W \to 2W}\,(\Rightarrow\to)}{K_2K_1\Pi, K_2\neg K_1 1W, K_2(\neg 2W \Rightarrow K_1\neg 2W) \to K_2 2W}\,(K_2 2)$$

It is obvious that $\Pi \to 1W, 2W$ can be resolved into a vaild sequent.

By inserting some applications of *thin*, one can immediately obtain a (cut-free) proof-figure of S_0.

6 On non-knowledge operator: a proposal

The main motif of this article has been the knowledge operator. Taking the negative of a knowledge operator, we can express "non-knowledge." The notion of "does not know" is, however, not unique. For example, in Section 2 in this article and in our previous works, $\neg KA$ has been employed for "does not know A." Another candidate is $K\neg KA$, and this works for the two wise girls puzzle just as well. It can easily be shown that the latter implies the former in $KL2$, and "the former implis the latter" is exactly A-3. (See also a remark in the final section of [12].)

There can be other versions of non-knowledge. We will here propose to employ the non-knowledge operator N as primitive, and see what sort of properties are requied fot it.

The first candidate for N is $N \equiv \neg K$ and the second one is $N \equiv K\neg K$ as mentioned above. In $KL3$, or equivalently in BKL with the axiom A-3, these two are equivalent. By taking $N \equiv \neg K$, we can derive $NNX \to KX$ in $KL3$ as follow.

$$\dfrac{\dfrac{\dfrac{\neg KX \to K\neg KX \quad K\neg KX, \neg K\neg KX \to}{\neg KX, \neg K\neg KX \to}\ (cut)}{\neg K\neg KX \to \neg\neg KX}\ (\to \neg) \qquad \neg\neg KX \to KX}{\neg K\neg KX \to KX}\ (cut)$$

This can hold in a situation that, if an agent does not know of her non-knowledge of a fact, it indicates that she does know of the fact.

On the other hand, we may consider an axiom

$$NNF \to NF.$$

For example, if an agent does not know that she does not know there is a language called Dragon, it indicates that she does not know there is a language called Dragon.

Related to this sequent, we can list some of properties which we wish to require for N.

A formula is called propositional if its outermost operator is not N.
Desirable properties.

1. $NNF \to NF$ is admitted.
2. If $F \to$ is provable for a propositional formula F, then $\to NF$ is provable.
3. For propositional A and B, $NA \vee NB \to N(A \wedge B)$ is provable.
4. For propositional A and B, $N(A \vee B) \to NA \wedge NB$ is provable.
5. If, for a propositional F, $\to F$ is provable, then $NF \to$ is provable.
6. $NF, N\neg F \to$ is not necessarily provable. In general, $\to A, B$ does not necessarily imply $NA, NB \to$.

We wish to make more investigations on interpretations of non-knowledge operators in future.

References

[1] B. Chellas, *Modal Logic*, Cambridge University Press, Cambridge (1980).

[2] G. Gentzen, "Investigations into logical deduction", (English translation), in *The Collected Papers of Gerhard Gentzen*, North-Holland, 1969, 68-131.

[3] J.H. Halpern and Y. Moses, *A guide to completeness and complexity for modal logics of knowledge and beliefs*, Artificial Intelligence 54, 319-379 (1992).

[4] M. Kaneko and T. Nagashima, *Game Logic and its Applications II*, in Studia Logica, vol.58(1997),273-303.

[5] S.H. Oda and M. Yasugi, *Inference within knowledge*, Discussion Paper Series No.27, The Society of Economics and Business Administration, Kyoto Sangyo University, 1998. (See also http://www.kyoto-su.ac.jp/~yasugi/recent.html)

[6] S.H.Oda and M.Yasugi, *Jumping out from the system*, The Proceedings of International Workshop on Emergent Synthesis, pp257-262; available at http://www.kyoto-su.ac.jp/~oda/projectsE (1999).

[7] M. Ohnishi and K. Matsumoto, *Gentzen Method in Modal Calculi. I*, Osaka Math. J., **9**, (1957), pp113-130.

[8] A. Rubinstein, *Modeling Bounded Rationality*, The MIT Press, Cambridge, Massachusetts(1998).

[9] M. Sato, *A Study of Kripke-type Models for Some Modal Logics by Gentzen's Sequential Method*, Publications of RIMS,vol.13(1977), 381-468, Research Institute of Mathematical Sciences, Kyoto University.

[10] M. Yasugi and S. H. Oda, *A proof-theoretic approach to knowledge*, Discussion Paper Series 29, The Society of Economics and Business Administration, Kyoto Sangyo University; available at http://www.kyoto-su.ac.jp/ỹasugi/Recent.

[11] M. Yasugi and S.H. Oda, *Jumping out of the system: an analysis in terms of proof-theory* (in Japanese), Journal of the Japan Association for Philosophy of Science, vol.28,no.2(2001), 33-38.

[12] M. Yasugi and S.H. Oda, *A note on the wise girls puzzle*, Economic Theory,to appear in vol.19,issue 1(2002),145-156,Springer-Verlag.

Faculty of Science, Kyoto Sangyo University, Kita-ku, Kyoto, Japan 603-8555
Tel: (+81)75-705-1780
Fax: (+81)75-705-1640
e-mail: yasugi@cc.kyoto-su.ac.jp

Faculty of Economics, Kyoto Sangyo University, Kita-ku, Kyoto, Japan 603-8555
email: oda@cc.kyoto-su.ac.jp