

## Hypersets and dynamics of knowledge

Dedicated to Professor Hideki Ozeki on his sixtieth birthday

Shunichi TANAKA and Toru TSUJISHITA

(Received February 17, 1994; Revised November 4, 1994)

**Abstract.** Using hyperset theory, the knowledge states of the cognitive agents are introduced, which makes it possible to describe the effect of knowledge publication. Based on this framework, the process of conversation in the muddy boys puzzle is described as a dynamical system on the knowledge state space and its orbits are determined explicitly.

*Key words:* non-well-founded set, knowledge state space, common knowledge, Kripke structure, publication operator.

### Introduction

Hyperset theory, or non-well-founded set theory, whose foundation was laid down by Aczel [1], provides us a concise clear-cut language to express and analyse logically complex phenomena involving various form of circularity and radically increased the universality of the set theory as a basic language of mathematical science (see [2,4] for introduction to the hyperset theory). Main purpose of this paper is to apply this new set theory to analyze the notion of knowledge and to describe the effect of publication of some valid information.

There have been many researches on logic of knowledge by mathematical logicians and computer scientists [3,8]. Recent interest in common knowledge among the computer scientists seems to come chiefly from the realization that programming of various protocols of distributed systems become transparent by taking agents' knowledge into account. In [5], Chandy and Misra suggests potential usefulness of treating knowledge as local states of the agents.

In this paper, we introduce the concept of *knowledge state* of agents: The global state  $S$  of the world consists of the state of the external environment  $e \in E$  and the knowledge states  $\alpha_i$  of the agents living in the world:

$$S = \langle \alpha_1, \dots, \alpha_n, e \rangle.$$

The knowledge state  $\alpha_i$  of the agent  $i$  is a subset of the state space of the world. Such an object as  $S$  can be formulated as a set only in the hyperset theory, since we demand the validity of knowledge, i.e.,  $S \in \alpha_i$ , which implies  $S \ni \cdots \ni S$ , incompatible with the regularity axiom of the usual set theory.

Fixing the number of agents and the state space  $E$  of the environment external to agents, we define *the state space*  $\mathbf{W}_{n,E}$  of the world. The change of knowledge can be described as a dynamical system on this class. As an illustration of our framework, we analyze the famous Muddy Boys Puzzle [6,7,8,9,10]. We describe the process of the conversation between the teacher and the boys as a dynamical system on a subset of  $\mathbf{W}_{n,E}$  and determine the orbits. The change of knowledge state is usually explained intuitively by the truncation of binary hypercube [8]. We can now incorporate this intuition in theoretical framework by virtue of the hyperset theory.

## 1. Anti-foundation axiom

We shall work in the set theory with the axiom system ZFA: Zermelo-Fraenkel axioms with the regularity axiom replaced with the Anti-Foundation Axiom of Aczel [1]. In ZFA, we can uniquely solve any system of set equations: Let  $X$  be a class of atoms and  $\{a_x | x \in X\}$  be a system of  $X$ -sets. Then there exist unique family of sets  $\{b_x | x \in X\}$  such that

$$b_x = a_x[\{b_x/x\}] \quad (\forall x \in X),$$

where  $a[\{b_x/x\}]$  denotes the set obtained from an  $X$ -set  $a$  by substituting  $b_x$  to  $x$ .

We denote the class of all the hypersets by  $\mathbf{V}$  and the class of all the subsets of a class  $\mathbf{X}$  by  $\mathbf{pow X}$ .

## 2. Knowledge propositions

Let  $E$  be a non-empty set. Define  $\mathbf{KPROP}(n, E)$  to be the smallest set of words on the alphabet set

$$\mathbf{pow}(E) \coprod \{K_1, \dots, K_n, C, \neg, \rightarrow, \perp\}.$$

satisfying the following conditions:

- $\perp \in \mathbf{KPROP}(n, E)$ ,
- $\mathbf{pow}(E) \subseteq \mathbf{KPROP}(n, E)$ ,

- If  $\varphi, \psi \in \text{KPROP}(n, E)$ , then  $K_i\varphi, C\varphi, \varphi \rightarrow \psi \in \text{KPROP}(n, E)$ .

We use the following standard abbreviations:

$$\begin{aligned}\neg\varphi &:= \varphi \rightarrow \perp \\ \varphi_1 \vee \varphi_2 &:= \neg\varphi_1 \rightarrow \varphi_2 \\ \varphi_1 \wedge \varphi_2 &:= \neg(\neg\varphi_1 \vee \neg\varphi_2).\end{aligned}$$

For a map  $f$  defined on  $E$  with finite range, we define for  $i \in \{1, \dots, n\}$

$$K_i f = \bigvee_{v \in \text{Range}(f)} K_i(f^{-1}v).$$

We will give interpretations of elements of  $\text{KPROP}(n, E)$  so that

- $K_i\varphi$  will mean that the  $i$ -th agent knows  $\varphi$ ,
- $C\varphi$  will mean that  $\varphi$  is a common knowledge among all the agents.

Then  $K_i f$  will assert that the  $i$ -th agent knows the value of the function  $f$  is  $v$  for some  $v$  in the range of  $f$  without specifying the value  $v$ .

### 3. Knowledge states

Let  $\mathcal{U} \subseteq \mathbf{V}^n \times E$  be a subclass defined by

$$\begin{aligned}\mathcal{U} = \{S \mid \exists \alpha_1 \cdots \exists \alpha_n \in \mathbf{V} \\ \exists e \in E [S = \langle \alpha_1, \dots, \alpha_n, e \rangle \wedge S \in \alpha_1 \cap \cdots \cap \alpha_n]\}.\end{aligned}$$

Define a class operator  $\mathbf{X} \mapsto \mathbf{K}_{n,E}\mathbf{X}$  by

$$\mathbf{K}_{n,E}\mathbf{X} := ((\mathbf{pow}\mathbf{X})^n \times E) \cap \mathcal{U}.$$

This operator is obviously set continuous and have the largest fixed point  $\mathbf{W}_{n,E}$  defined by

$$\mathbf{W}_{n,E} = \bigcup_{\alpha \subseteq \mathbf{K}_{n,E}\alpha} \alpha,$$

(see [1]).

Let  $\pi_i : \mathbf{W}_{n,E} \rightarrow \mathbf{pow}\mathbf{W}_{n,E}$  and  $\pi_E : \mathbf{W}_{n,E} \rightarrow E$  be the restrictions to  $\mathbf{W}_{n,E}$ , respectively, of the projection to the  $i$ -factor:

$$(\mathbf{pow}\mathbf{W}_{n,E})^n \times E \rightarrow \mathbf{pow}\mathbf{W}_{n,E}$$

and to  $E$ :

$$(\mathbf{pow}\mathbf{K}_{n,E})^n \times E \rightarrow E.$$

In the universe  $\mathbf{V}$  of hypersets, the class  $\mathbf{W}_{n,E}$  is not empty. In fact, for each  $\gamma \in E$ , consider the set equation

$$x = \langle \{x\}, \dots, \{x\}, \gamma \rangle. \quad (3.1)$$

Let  $x = S^\gamma$  be the unique solution. Since  $\{S^\gamma\} \subseteq \mathbf{K}_{n,E} \{S^\gamma\}$  and  $S^\gamma \in \bigcap_{i=1}^n \pi_i S^\gamma$  hold, we have  $\{S^\gamma\} \subseteq \mathbf{W}_{n,E}$ , whence  $\mathbf{W}_{n,E}$  is not empty.

$a \subseteq \mathbf{W}_{n,E}$  is called *closed* if  $a \subseteq \mathbf{K}_{n,E}(a)$  holds.

#### 4. Interpretation of knowledge propositions

We associate a subclass  $[\varphi] \subseteq \mathbf{W}_{n,E}$  to  $\varphi \in \text{KPROP}(n, E)$  as follows:

- $[\perp] := \emptyset$ ,
- for  $\varphi \in \mathbf{pow}(E)$ ,

$$[\varphi] := \pi_E^{-1} \varphi,$$

- for  $\varphi, \varphi_i \in \text{KPROP}(n, E) (i = 1, 2)$ ,

$$[\neg\varphi] := [\varphi]^c,$$

$$[\varphi_1 \rightarrow \varphi_2] := [\varphi_1]^c \cup [\varphi_2],$$

$$[K_i\varphi] := \pi_i^{-1} \mathbf{pow}[\varphi],$$

- for  $\varphi \in \text{KPROP}(n, E)$ , the subclass  $[C\varphi] \subseteq \mathbf{W}_{n,E}$  is the largest fixed point of the set continuous operator

$$\mathbf{X} \mapsto [\varphi] \cap \mathbf{K}_{n,E}\mathbf{X}. \quad (4.1)$$

Note that

$$[K_i f] = \bigcup_{v \in \text{Range}(f)} \pi_i^{-1} \mathbf{pow}(\pi_E^{-1} f^{-1} v).$$

For  $S \in \mathbf{W}_{n,E}$  and  $\varphi \in \text{KPROP}(n, E)$ , we write  $S \models \varphi$  if  $S \in [\varphi]$ . When  $[\varphi] = \mathbf{W}_{n,E}$ , we call  $\varphi$  *valid* and write  $\models \varphi$ .

The following proposition gives some fundamental facts concerning our interpretation of knowledge propositions.

**Proposition 4.1** *For  $\varphi \in \text{KPROP}(n, E)$ ,  $S \in \mathbf{W}_{n,E}$ , and a map  $f$  defined on  $E$  with finite range,*

- (i)  $[K_i\varphi] \subseteq [\varphi]$ ,
- (ii)  $[C\varphi] = [\varphi] \cap [K_1 C\varphi] \cap \dots \cap [K_n C\varphi]$ ,

- (iii) If  $a \subseteq [\varphi] \cap \mathbf{K}_{n,E}(a)$  then  $a \subseteq [C\varphi]$ ,  
 (iv)  $S \models \neg K_i f \iff |f\pi_E\pi_i S| \geq 2$ .

*Proof.* (i) Suppose  $S \models K_i \varphi$ . Then  $T \models \varphi$  for all  $T \in \pi_i S$ . Since  $S \in \pi_i S$ , we conclude  $S \models \varphi$ .

(ii) Obvious since

$$\begin{aligned} \mathbf{K}_{n,E}\mathbf{X} &= \pi_1^{-1}\mathbf{pow}\mathbf{X} \cap \cdots \cap \pi_n^{-1}\mathbf{pow}\mathbf{X} \\ &= [K_1\mathbf{X}] \cap \cdots \cap [K_n\mathbf{X}]. \end{aligned}$$

(iii) The largest fixed point of the set continuous operator defined by (4.1) is precisely the union of those  $a$  satisfying the condition in the statement.

(iv)  $S \models K_i f$  is equivalent to  $S \in \pi_i^{-1}\mathbf{pow}\pi_E^{-1}f^{-1}v$  for some  $v \in \text{Range}(f)$ . This means  $f(\pi_E\pi_i S) = \{v\}$ .  $\square$

By (ii), (iii),  $[C\varphi]$  is characterized as the largest closed subset of  $[\varphi]$ .

For  $S \in \mathbf{W}_{n,E}$ , we define its *knowledge closure*  $\mathbf{W}_{n,E}(S)$  as the smallest closed subset of  $\mathbf{W}_{n,E}$  that contains  $S$ . For  $\varphi \in \text{KPROP}(n, E)$  and  $S \models C\varphi$ , we have  $\mathbf{W}_{n,E}(S) \subseteq [C\varphi]$ .

**Proposition 4.2** For  $\varphi \in \text{KPROP}(n, E)$  and  $S \in \mathbf{W}_{n,E}$ ,

$$S \models C\varphi \iff \mathbf{W}_{n,E}(S) \subseteq [\varphi].$$

*Proof.* Suppose  $S \models C\varphi$ . Then  $\mathbf{W}_{n,E}(S) \subseteq [C\varphi] \subseteq [\varphi]$ . Conversely, suppose  $\mathbf{W}_{n,E}(S) \subseteq [\varphi]$ . Then  $\mathbf{W}_{n,E}(S)$ , being a closed subset of  $[\varphi]$ , is contained in  $[C\varphi]$ . In particular  $S \models C\varphi$ .  $\square$

## 5. Publication operator

For  $\varphi \in \text{KPROP}(n, E)$ , we define the *publication operator*

$$\kappa_\varphi : [\varphi] \rightarrow \mathbf{W}_{n,E}$$

by

$$\kappa_\varphi S = \langle \kappa_\varphi([\varphi] \cap \pi_1 S), \dots, \kappa_\varphi([\varphi] \cap \pi_n S), \pi_E S \rangle$$

where  $\kappa_\varphi \alpha = \{ \kappa_\varphi S \mid S \in \alpha \}$  for  $\alpha \subseteq \mathbf{W}_{n,E}$ .

Justification of this recursive definition based on AFA is as follows:

Consider the following system of set equations:

$$x_S = \langle \{x_{S'} \mid S' \in \pi_1 S \cap [\varphi]\}, \dots, \{x_{S'} \mid S' \in \pi_n S \cap [\varphi]\}, \pi_E S \rangle$$

on the unknown sets  $\{x_S \mid S \in [\varphi]\}$ . Let  $\{x_S = b_S \mid S \in [\varphi]\}$  be its unique solution. We define then

$$\kappa_\varphi S := b_S, \quad \text{for } S \in [\varphi].$$

The operator  $\kappa_\varphi$  describes the change of the knowledge status of the world after some true proposition  $\varphi$  is made public.

We also define  $\lambda_\varphi : \mathbf{W}_{n,E} \rightarrow \mathbf{W}_{n,E}$  by

$$\lambda_\varphi(S) := \begin{cases} \kappa_\varphi & \text{if } S \models \varphi \\ \kappa_{\neg\varphi} & \text{if } S \models \neg\varphi. \end{cases}$$

This operator describes the change caused by informing the truth value of the proposition  $\varphi$ . Note that  $\lambda_\varphi[\varphi] \not\subseteq [\varphi]$  in general (cf. Theorem 7.3 and 7.4). A fact concerning the knowledge status of agents may become false by being made public, which conforms with our daily common sense.

## 6. Muddy Boy's Puzzle

As an application of our formalism, we formulate and analyze the following famous puzzle:

There are  $n(\geq 2)$  boys in a room, each boy's face is either muddy or clean. Each can see whether each boy other than himself has clean face or not. Now their teacher comes in and tell them that at least one of them has dirty face. He orders the boys to close their eyes and asks them

(\*) "Raise you hand if you know whether your face is clean or not."

Now the teacher repeats the following question:

- If a boy raises his hand, then the teacher says to all the boys "Some of you know that his face is dirty or not," and then repeats (\*).
- If none of the boys raise their hands, then the teacher says to them "Nobody in this room knows whether his face is dirty or not," and then repeats (\*).

What will happen when the teacher continues this question?

The answer is well-known and easy to find: Let  $k$  denote the number of boys with muddy faces. Then, up to the  $(k - 1)$ -th question, no boy raises his hand, but at the  $k$ -th question, the boys with dirty faces raise their hands and at the  $(k + 1)$ -th question, every boy raises his hand.

In the following we interpret this puzzle as a dynamical system on the space of knowledge states and figure out the orbit.

## 7. Analysis of Muddy Boy's Puzzle

**7.1 Formulation of the puzzle.** Put  $E := \{0, 1\}^n$  and  $h_i : E \rightarrow \{0, 1\}$  ( $i = 1, \dots, n$ ) be the projection on the  $i$ -th factor. Let  $h : E \rightarrow \{0, 1\}^n$  be the identity map. For  $\gamma = \langle \gamma_1, \dots, \gamma_n \rangle \in E$ , we write

$$|\gamma| := \sum_{i=1}^n \gamma_i.$$

Define

$$\begin{aligned} \varphi_0 &:= \neg \{\mathbf{0}\} \quad (\mathbf{0} := \langle 0, \dots, 0 \rangle), \\ \psi &:= K_1 h_1 \vee \dots \vee K_n h_n, \\ \varphi_1 &:= (\wedge_{i \neq j} K_i h_j) \wedge \neg \psi. \end{aligned}$$

The initial knowledge status in the classroom is specified by the knowledge proposition  $\varphi_0 \wedge C\varphi_1$ . The first statement of the teacher is  $\varphi_0$ . A boy raising his hand as the answer to the teacher's question (\*) means the validity of  $\psi$ . Each time the teacher makes public whether or not  $\psi$  is true.

More formally, we can depict the process occurring in the room as the following dynamical system on  $\mathbf{W}_{n,E}$ :

- The initial state  $S_0$  is in  $[\varphi_0 \wedge C\varphi_1]$ .
- The first statement of the teacher changes  $S_0$  to

$$S_1 := \kappa_{\varphi_0} S_0 \in \mathbf{W}_{n,E}.$$

- After each cycle of conversation, the state  $S_i$  changes to

$$S_{i+1} := \lambda_{\psi}(S_i).$$

The puzzle can be regarded as the question to describe, or find, the orbit of this dynamical system on  $\mathbf{W}_{n,E}$  starting from a state  $S_0$  satisfying  $\varphi_0 \wedge C\varphi_1$ .

We will completely determine the orbits by applying the solution lemma

of the hyperset theory.

**7.2 Basic knowledge states.** For each  $k = 0, 1, \dots, n$ , consider the following system of set equations on  $\{x_\gamma \mid |\gamma| \geq k\}$ :

$$x^\gamma = \langle \beta_1^k(x^\gamma), \dots, \beta_n^k(x^\gamma), \gamma \rangle \quad (7.1)_k$$

where

$$\beta_i^k(x^\gamma) = \begin{cases} \{x^\gamma, x^{\gamma_i}\} & \text{if } |\gamma_i| \geq k, \\ \{x^\gamma\} & \text{if } |\gamma_i| < k \end{cases}$$

and  $\gamma_i \in E$  is defined by

$$\gamma_i := \langle \gamma_1, \dots, 1 - \gamma_i, \dots, \gamma_n \rangle.$$

Let  $\{x^\gamma = S_k^\gamma \mid |\gamma| \geq k\}$  be the solution of the equation (7.1)<sub>k</sub>. Put

$$\mathbf{S}_k = \{S_k^\gamma \mid |\gamma| \geq k\}.$$

Note that  $\mathbf{S}_k$  is closed.

**7.3 Solution of the puzzle.** The initial state space in the room is described by the following theorem.

**Theorem 7.1**

$$[\varphi_0 \wedge C\varphi_1] = \{S_0^\gamma \mid \gamma \neq \mathbf{0}\}.$$

The first statement of the teacher, which announces the information  $\varphi_0$  has the following effect:

**Theorem 7.2**

$$\kappa_{\varphi_0}(S_0^\gamma) = S_1^\gamma \quad \text{for } \gamma \neq \mathbf{0}.$$

The subsequent cycle of conversation changes the status of the room as follows:

**Theorem 7.3**

$$\lambda_\psi S_k^\gamma = \begin{cases} S_{k+1}^\gamma & \text{if } |\gamma| > k, \\ S^\gamma & \text{if } |\gamma| = k, \end{cases}$$



where  $S^\gamma$  is the set defined by the equation (3.1).

Moreover, at each stage, the knowledge status has the following feature:

**Theorem 7.4** For  $k \geq 1$

$$S_k^\gamma \models \begin{cases} \neg\psi & \text{if } |\gamma| > k, \\ (\bigwedge_{\gamma_i=1} K_i h_i) \wedge (\bigwedge_{\gamma_j=0} \neg K_j h_j) & \text{if } |\gamma| = k. \end{cases}$$

In particular  $S_k^\gamma \models \psi$  if  $|\gamma| = k$ .

The proof of these theorems will be given in the appendix.

## 8. KPROP( $n, E$ ) as a modal logic

In [10], modal systems  $KT3, KT4, KT5$  are introduced and applied to analyse knowledge puzzles. When the temporal factor is ignored, these systems are reduced to  $K3, K4$  and  $K5$  systems, which have axiom schemes (A1–6), (A1–7) and (A1–8) respectively, where, in our notation,

- (A1)  $\neg\neg\varphi \rightarrow \varphi$
- (A2)  $\varphi \rightarrow (\psi \rightarrow \varphi)$
- (A3)  $(\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_3)) \rightarrow ((\varphi_1 \rightarrow \varphi_2) \rightarrow (\varphi_1 \rightarrow \varphi_3))$
- (A4)  $K_i\varphi \rightarrow \varphi$
- (A5)  $C\varphi \rightarrow CK_i\varphi$
- (A6)  $K_i(\varphi_1 \rightarrow \varphi_2) \rightarrow (K_i(\varphi_1) \rightarrow K_i(\varphi_2))$
- (A7)  $K_i\varphi \rightarrow K_iK_i\varphi$
- (A8)  $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ .

Here  $0 \leq i \leq n$  and  $K_0$  denotes  $C$ .

The elements of  $KPROP(n, E)$  can be regarded as formulas in a modal propositional logic with modal operators  $K_i (1 \leq i \leq n)$  and  $C$ .

**Proposition 8.1** (i) For any  $\varphi, \varphi_i$ , the formulas (A1–6) are valid in  $\mathbf{W}_{n,E}$ .

(ii) Neither (A7) nor (A8) is valid in  $\mathbf{W}_{n,E}$ .

(iii) For any  $\varphi$  and  $0 \leq i \leq n$ , the following is a valid formula.

$$C\varphi \rightarrow K_i C\varphi. \tag{C}$$

To prove this proposition, we make some definitions. Let  $\mathbf{Y}$  be a subclass of  $\mathbf{W}_{n,E}$ . Denote by  $C\mathbf{Y}$  the largest closed subclass of  $\mathbf{Y}$ . Define also

$$K_i\mathbf{Y} := \{S \mid \pi_i S \subset \mathbf{Y}\},$$

which is included in  $\mathbf{Y}$ .

**Lemma 8.2** (i)  $[K_i\varphi] = K_i[\varphi]$ .  
(ii)  $[C\varphi] = C[\varphi]$ .  
(iii)  $CK_i\mathbf{Y} = C\mathbf{Y}$  ( $1 \leq i \leq n$ ).

*Proof.* The assertions (i) and (ii) are obvious. To prove (iii), it suffices to show that  $C\mathbf{Y} \subset K_i\mathbf{Y}$ , since this implies that  $C\mathbf{Y}$  is also the largest closed subclass of  $K_i\mathbf{Y}$  and hence must be  $CK_i\mathbf{Y}$ . Suppose  $S \in C\mathbf{Y}$ . Then  $S \in \mathbf{K}_{n,E}C\mathbf{Y}$ , which means  $\pi_i S \subset C\mathbf{Y} \subset \mathbf{Y}$  for  $1 \leq i \leq n$ . Hence  $S \in K_i\mathbf{Y}$ .  $\square$

*Proof of Proposition 8.1* (i) (A1–3) and (A6) are obvious and (A4) is proved in Proposition 4.1. The validity of (C) is obvious from the definition of  $[C\varphi]$ . (A5) is obvious from the (iii) of the above lemma.

To see that neither (A7) nor (A8) is valid, we construct an element  $S \in \mathbf{W}_{n,E}$  and  $\varphi$  such that neither  $S \models (A7)$  nor  $S \models (A8)$ . We assume  $n = 2$ . Let  $S_1, S_2, S_3$  be the hypersets characterized by the following equations.

$$\begin{aligned} S_1 &= \langle \{S_1, S_2\}, \{S_1, S_2\}, e_1 \rangle \\ S_2 &= \langle \{S_2, S_3\}, \{S_2, S_3\}, e_2 \rangle \\ S_3 &= \langle \{S_3\}, \{S_3\}, e_3 \rangle. \end{aligned}$$

These hypersets belongs to  $\mathbf{W}_{n,E}$ , since

$$\{S_1, S_2, S_3\} \subset \mathbf{K}_{n,E} \{S_1, S_2, S_3\}.$$

Obviously

$$\begin{aligned} S_1 &\models K_1 \{e_1, e_2\} \\ S_2 &\models K_1 \{e_2, e_3\}. \end{aligned}$$

Suppose  $e_3 \neq e_1, e_2$ . Then  $S_3 \not\models \{e_1, e_2\}$ , which implies  $S_2 \not\models K_1 \{e_1, e_2\}$  and  $S_1 \not\models K_1 K_1 \{e_1, e_2\}$ . Hence

$$S_1 \not\models K_1 \{e_1, e_2\} \rightarrow K_1 K_1 \{e_1, e_2\},$$

which means (A7) is not valid generally.

Similarly, we can show

$$S_1 \not\models \neg K_1 \{e_2, e_3\} \rightarrow K_1 \neg K_1 \{e_2, e_3\}.$$

In fact,  $S_1 \models \neg K_1 \{e_2, e_3\}$  but  $S_1 \not\models K_1 \neg K_1 \{e_2, e_3\}$ , since  $S_2 \models K_1 \{e_2, e_3\}$ . Hence (A8) is not valid generally.  $\square$

Theorem 7.1 implies immediately

**Corollary 8.3**

$$\models \varphi_0 \wedge C\varphi_1 \rightarrow (A7) \wedge (A8).$$

Thus in our formulation, the formulas (A7–8) are consequences of the specification of the puzzle.

We do not know whether all the valid formulas are consequences of (A1–6) and (C) with the modus ponens as the only deduction rule.

Finally we note that  $\mathbf{W}_{n,E}$  can be considered as the universal Kripke structure for the modal logic  $\text{KPROP}(n, E)$  with the axiom schemes (A1–6) and (C) [11].

## Appendix

### A. Proof of Theorems

#### A.1 Proof of Theorem 7.1

**Lemma A.1** For  $S \in \mathbf{W}_{n,E} \cap \pi_E^{-1}\gamma$ ,  $S \models \varphi_1 \iff \pi_E \pi_i S = \{\gamma, \gamma i\}$ .

*Proof.* Let  $S \in \mathbf{W}_{n,E} \cap \pi_E^{-1}\gamma$ .

$$\begin{aligned} S \models \varphi_1 &\iff \begin{cases} S \models K_i h_j & \text{if } j \neq i, \\ S \not\models K_i h_j & \text{if } j = i, \end{cases} \\ &\iff \begin{cases} |h_j \pi_E \pi_i S| = 1 & \text{if } j \neq i, \\ |h_j \pi_E \pi_i S| \geq 2 & \text{if } j = i, \end{cases} \\ &\iff h_j \pi_E \pi_i S = \begin{cases} \{\gamma_j\} & \text{if } j \neq i, \\ \{0, 1\} & \text{if } j = i, \end{cases} \\ &\iff \pi_E \pi_i S = \{\gamma, \gamma i\}. \end{aligned}$$

□

**Corollary A.2**  $S_0^\gamma \models \varphi_1$  for all  $\gamma$ .

*Proof.* From (7.1)<sub>0</sub>, we have  $\pi_i S_0^\gamma = \{\gamma, \gamma i\}$  for all  $\gamma \in E$  and  $i \in \{1, \dots, n\}$ , whence the conclusion by the above lemma. □

**Corollary A.3**  $\mathbf{S}_0 \subseteq [C\varphi_1]$ .

*Proof.* Since  $\mathbf{S}_0$  is closed, Corollary A.2 implies

$$\mathbf{W}_{n,E}(S_0^\gamma) \subseteq \mathbf{S}_0 \subseteq [C\varphi_1].$$

Hence, by Proposition 4.2,  $S_0^\gamma \models C\varphi_1$  for all  $\gamma \in E$ . □

*Proof of Theorem 7.1* It remains to show that

$$[C\varphi_1] \subseteq \{S_0^\gamma \mid \gamma \in E\}.$$

Let  $S \in [C\varphi_1]$ . We consider the following system of set equations:

$$x_T = \langle \{x_P \mid P \in \pi_1 T\}, \dots, \{x_P \mid P \in \pi_n T\}, \pi_E T \rangle \tag{A.1}$$

for the family of unknown sets

$$\{x_T \mid T \in \mathbf{W}_{n,E}(S)\}.$$

Obviously  $\{x_T = T \mid T \in \mathbf{W}_{n,E}(S)\}$  is the solution for this equation.

On the other hand

$$\{x_T = S_0^{\pi_E T} \mid T \in \mathbf{W}_{n,E}(S)\} \tag{A.2}$$

also satisfies the above equation. In fact, for  $T \in \mathbf{W}_{n,E}(S) \cap \pi_E^{-1}\gamma$ , we have

$$T \in [\varphi_1] \cap \pi_E^{-1}\gamma$$

(Proposition 4.2) and then  $\pi_E \pi_i T = \{\gamma, \gamma^i\}$  (Lemma A.1). Hence

$$\begin{aligned} \{S_0^{\pi_E P} \mid P \in \pi_i T\} &= \{S_0^\lambda \mid \lambda \in \pi_E \pi_i T\} \\ &= \{S_0^\gamma, S_0^{\gamma^i}\} \end{aligned}$$

for  $i = 1, \dots, n$ . Therefore under the substitution (A.2), the right hand side of (A.1) is

$$\langle \{S_0^\gamma, S_0^{\gamma^1}\}, \dots, \{S_0^\gamma, S_0^{\gamma^n}\}, \gamma \rangle$$

which is  $S_0^\gamma$ , the left hand side of (A.1) under the substitution (A.2).

The uniqueness of the solution implies  $S = S_0^{\pi_E S} \in \mathbf{S}_0$ . Hence  $[C\varphi_1] \subseteq \mathbf{S}_0$ . □

### A.2 Proof of Theorem 7.2

*Proof.* The defining equation of  $\kappa_{\varphi_0} S_0^\gamma$  ( $\gamma \neq 0$ ) is

$$x^\gamma = \langle \{x^\lambda \mid S_0^\lambda \in \pi_1 S_0^\gamma \cap [\varphi_0]\}, \dots, \{x^\lambda \mid S_0^\lambda \in \pi_n S_0^\gamma \cap [\varphi_0]\}, \gamma \rangle.$$

Since

$$\pi_i S_0^\gamma \cap [\varphi_0] = \begin{cases} \{S_0^\gamma, S_0^{\gamma^i}\} & \text{if } |\gamma^i| \neq 0, \\ \{S_0^\gamma\} & \text{if } |\gamma^i| = 0, \end{cases}$$

the above equation can be written as

$$x^\gamma = \langle \beta_1^1(x^\gamma), \dots, \beta_n^1(x^\gamma), \gamma \rangle,$$

which is precisely the defining equation of  $S_1^\gamma$  ( $|\gamma| > 0$ ). □

### A.3 Proof of Theorem 7.4

*Proof.* If  $|\gamma| > k$ ,

$$\pi_i S_k^\gamma = \{S_k^\gamma, S_k^{\gamma_i}\} \quad \text{for all } i \in \{1, \dots, n\},$$

which implies, by (iv) of Proposition 4.1,

$$S_k^\gamma \models \neg K_i h_i \quad \text{for all } i \in \{1, \dots, n\}.$$

Suppose now that  $|\gamma| = k$ . Since  $|\gamma_i| \geq k$  if and only if  $\gamma_i = 0$ , we have

$$\pi_i S_k^\gamma = \begin{cases} \{S_k^\gamma, S_k^{\gamma_i}\} & \text{if } \gamma_i = 0, \\ \{S_k^\gamma\} & \text{if } \gamma_i = 1, \end{cases}$$

whence

$$S_k^\gamma \models (\bigwedge_{\gamma_i=1} K_i h_i) \wedge (\bigwedge_{\gamma_j=0} \neg K_j h_j).$$

□

### A.4 Proof of Theorem 7.3

*Proof.* Suppose first  $|\gamma| > k$ . Then

$$\lambda_\psi = \kappa_{\neg\psi}$$

because  $S_k^\gamma \models \neg\psi$  by Theorem 7.4. Therefore  $\{x^\gamma = \lambda_\psi S_k^\gamma \mid |\gamma| > k\}$  is the solution of

$$x^\gamma = \left\langle \left\{ x^\delta \mid S_k^\delta \in \pi_1 S_k^\gamma \cap [\neg\psi] \right\}, \dots, \left\{ x^\delta \mid S_k^\delta \in \pi_n S_k^\gamma \cap [\neg\psi] \right\}, \gamma \right\rangle.$$

By Theorem 7.4 and the defining equation (7.1)<sub>k</sub>, this equation can be rewritten as

$$x^\gamma = \left\langle \left\{ x^\delta \mid \delta \in \{\gamma, \gamma 1\} \text{ and } |\delta| \geq k+1 \right\}, \dots, \left\{ x^\delta \mid \delta \in \{\gamma, \gamma n\} \text{ and } |\delta| \geq k+1 \right\}, \gamma \right\rangle,$$

namely

$$x^\gamma = \left\langle \beta_1^{k+1}(x^\gamma), \dots, \beta_n^{k+1}(x^\gamma), \gamma \right\rangle.$$

This shows  $\lambda_{\neg\psi}S_k^\gamma = S_{k+1}^\gamma$ .

Suppose now  $|\gamma| = k$ .  $x^\gamma = \lambda_\psi S_k^\gamma = \kappa_\psi S_k^\gamma$  is the solution of

$$x^\gamma = \langle \{x^\delta \mid S_k^\delta \in \pi_1 S_k^\gamma \cap [\psi]\}, \dots, \{x^\delta \mid S_k^\delta \in \pi_n S_k^\gamma \cap [\psi]\}, \gamma \rangle.$$

By Theorem 7.4,  $S_k^\delta \models \psi$  if and only if  $|\delta| = k$ , whence

$$\pi_i S_k^\gamma \cap [\psi] = \{S_k^\gamma\},$$

for all  $i$  and the equation reduces to

$$x^\gamma = \langle \{x^\gamma\}, \dots, \{x^\gamma\}, \gamma \rangle.$$

This proves  $\lambda_\psi S_k^\gamma = S^\gamma$ . □

## References

- [1] Aczel P., *Non-well-founded sets*. CSLI Lecture Notes No. 14, Center for the Study of Language and Information. Leland Stanford Junior University, 1988.
- [2] Barwise J. and Etchemendy J., *The liar, an essay on truth and circularity*. Oxford Univ. Press, 1987.
- [3] Barwise J., *On the model theory of common knowledge*. pp. 201–220 in Barwise J.: *The situation in logic*, CSLI Lecture Notes No. 17, Center for the Study of Language and Information, Leland Stanford Junior University, 1989.
- [4] Barwise J. and Moss L., *Hypersets*. The mathematical intelligencer, 13(4), pp. 31–41, 1991.
- [5] Chandy K.M. and Misra J., *How processes learn*. Distributed Computing 1. pp. 40–52, 1986.
- [6] Halpern J.Y. and Fagin R., *Modelling knowledge and action in distributed systems*. Distributed Computing 3, pp. 159–177, 1989.
- [7] Halpern J.Y. and Moses Y., *Knowledge and common knowledge in a distributed environment*. Journal of Association for Computing Machinery, 37(3), pp. 549–587, 1990.
- [8] Halpern J.Y. and Vardi M.Y., *Model checking vs. theorem proving: a manifesto*. pp. 151–176 in Artificial Intelligence and Mathematical Theory of Computation, Academic Press, 1991.
- [9] Moses Y., Dolev D. and Halpern J.Y., *Cheating husbands and other stories: a case study of knowledge, action, and communication*. Distributed Computing 1, pp. 167–176, 1986.
- [10] Sato M., *A study of kripke-type models for some modal logics by Gentzen's sequential method*. Publ. RIMS, Kyoto Univ. 13, pp. 381–469, 1977.

- [11] Tsujishita T., *Construction of universal modal world based on hyperset theory*. preprint.

Shunichi Tanaka  
Graduate school of Mathematics  
Kyushu University  
Fukuoka 812, Japan  
E-mail: tanaka@math.kyushu-u.ac.jp

Toru Tsujishita  
Department of Mathematics, Faculty of Science  
Hokkaido University  
Sapporo 060, Japan  
E-mail: tujisita@math.hokudai.ac.jp