Theoretical Computer Science 58 (1988) 155-174
North-Holland

# KNOWLEDGE, BELIEF AND TIME

Sarit KRAUS and Daniel LEHMANN

*Department of Computer Science, Hebrew University, Jerusalem 91904, Israel*

**Abstract.** In the conclusion of [7] Halpern and Moses expressed their interest in a logical system in which one could talk about knowledge *and* belief (and belief about knowledge, knowledge about belief and so on). We investigate such systems. In the first part of the paper knowledge and belief, without time, are considered. Common knowledge and common belief are defined and compared. A logical system and a family of models are proposed, a completeness result is proved and a decision procedure described. In the second part of the paper, time is considered. Different notions of beliefs are distinguished, obeying different properties of persistence. One interpretation of belief which obeys a very strong persistence axiom is put forward and used in the analysis of the "wise men" puzzle.

## 1. Introduction

Reasoning about knowledge and belief has been recently proposed as a tool for describing distributed systems, as well as many real-life situations. Examples include synchronization and cooperation protocols, cryptographic systems, games, economics, knowledge bases and intelligent programs. One of the outstanding questions is what is the best concept for analysing such situations: knowledge or belief. Following previous authors [5] it may be considered that the main difference between knowledge and belief is that when one knows $p$, then $p$ is true, but when one believes $p$, then $p$ must not be true. Some recent works examine the concept of knowledge [6, 7, 9] and others the concept of belief [10, 2, 3]. We think as suggested by Halpern and Moses in [7], that for some applications a good system has to be able to talk about belief *and* knowledge. We want to express statements such as "person $i$ believes $p$ and knows $q$", or "person $i$ believes that if he does not know $p$, then $q$ is true", and so on. In this paper we propose a logical system for many people that includes belief, knowledge, common knowledge and common belief. Following Lewis [11] the notion of common belief of some state of affairs $A$ in a population $P$ holds if and only if:

"(1) everyone in $P$ has reason to *believe* that $A$ holds;

(2) $A$ indicates to everyone in $P$ that everyone in $P$ has reason to *believe* that $A$ holds;

(3) $A$ indicates to everyone in $P$ that everyone in $P$..."

Lewis [11] calls this concept common knowledge, but we think the term common belief is more appropriate. The notion of common belief is much weaker than the

0304-3975/88/$3.50 © 1988, Elsevier Science Publishers B.V. (North-Holland)

notion of common knowledge. Even when $a$ is true and is common belief, $a$ must not be common knowledge. We think that in some cases, in particular those envisaged by Lewis, common belief is the proper concept and not common knowledge.

Another issue is how belief and knowledge are changing in time. In some circumstances, one cannot say anything nontrivial about the effect of the passage of time on knowledge or belief, but in others one would like to state that knowledge is never lost (see [9]) and that beliefs do not change without reason. When discussing the relation of belief and time one cannot escape the conclusion that there is more than one notion of belief. So, we add "time" to our system and for two different notions of belief we suggest different axioms to express the persistence of beliefs and common beliefs.

Finally we use one of our systems to partly analyse the "wise men" puzzle, (see [1]) in a completely formal way.

## 2. Logic of knowledge and belief

We now consider knowledge and belief alone (without time).

### 2.1. The language

Suppose a set Pvar of propositional variables and a finite set People $=^{\text{def}} \{1, 2, \ldots, n\}$ of participants are given. The following rules define the set of formulas $\Gamma$.

(1) A propositional variable $p \in$ Pvar is a formula

If $a$ and $b$ are formulas of $\Gamma$ and if $i \in$ People, then the following are formulas of $\Gamma$:

(2) $\neg a$ (not $a$) and $a \vee b$ ($a$ or $b$),

(3) $\mathrm{K}_i a$ ($i$ knows $a$) and $\mathrm{B}_i a$ ($i$ believes $a$).

(4) $\mathcal{K} a$ ($a$ is common knowledge) and $\mathcal{B} a$ ($a$ is common belief).

In this language one can express pure knowledge formulas, pure belief formulas, but one can also express formulas like $\mathrm{K}_i a \to \mathrm{B}_i a$ that means that person $i$ believes what he knows, or $\neg \mathrm{B}_i a \leftrightarrow \mathrm{K}_i \neg \mathrm{B}_i a$ that means that person $i$ does not believe $a$ iff he knows that he does not believe $a$. Let us define the connectives $\mathcal{E}$ and $\mathcal{F}$ by

$$\mathcal{E} a \overset{\text{def}}{=} \bigwedge_{i \in \text{People}} \mathrm{K}_i a, \qquad \mathcal{F} a \overset{\text{def}}{=} \bigwedge_{i \in \text{People}} \mathrm{B}_i a.$$

The formula $\mathcal{E} a$ means that everybody knows $a$ and $\mathcal{F} a$ means that everybody believes $a$.

### 2.2. The models

We use possible-worlds semantics for knowledge and belief. Person $i$ knows $a$ if $a$ is true in all the worlds that according to his knowledge could be possible, and person $i$ believes $a$ if $a$ is true in all the worlds that could be possible according to his beliefs. We shall now define models in the style of Kripke with two binary relations for each person: one relation corresponding to worlds possible according to knowledge, and the other relation corresponding to the worlds believed possible.

**Definition 2.1.** A model $\mathcal{M}$ is a structure $\langle S, s, l, \equiv_1, \equiv_2, \ldots, \equiv_n, \sim_1, \sim_2, \ldots, \sim_n \rangle$ where

(1) $S$ is a set, elements of $S$ will be called states;

(2) $s \in S$ is the real state of the world;

(3) $l: S \to 2^{\mathrm{Pvar}}$ says which of the propositional variables are true in each state;

(4) for every $i \in$ People, the relations $\equiv_i$ and $\sim_i$ satisfy the following rules:

(a) $\equiv_i$ is an equivalence relation (reflexive, transitive and symmetric),

(b) $\sim_i$ is serial (for all $s \in S$ there is some $t \in S$ such that $s \sim_i t$),

(c) $\sim_i$ is contained in $\equiv_i$ $(\sim_i \subseteq \equiv_i)$,

(d) for any $s, t, u \in S$ if $s \equiv_i t$ and $t \sim_i u$, then $s \sim_i u$.

The intuitive meanings of the relations $\equiv_i$ and $\sim_i$ are the following. Two states $s, u \in S$ are in relation $\equiv_i$ if the knowledge of person $i$ cannot enable him to distinguish between $s$ and $u$. Two states $s$, $u$ are in relation $\sim_i$ if in states $s$ person $i$ believes that state $u$ is a state the world could be in. The relation $\equiv_i$ is an equivalence relation and its intuitive meaning is that all the states in $S$ are divided into equivalence classes, and if person $i$ is in state $u \in S$ and if $u$ is member of the equivalence class $E \subseteq S$, then all the states that are member of $E$ (including $u$) are possible according to his knowledge.

In every state that person $i$ could be in, there is at least one state that he believes is possible, and therefore $\sim_i$ is serial.

It is easier to believe something than to know it, because one knows only true things. So one's beliefs can enable him to distinguish between more states than one's knowledge, and therefore there could be some states $s, u \in S$ such that $s \equiv_i u$ but $s \not\sim_i u$. On the other hand, one believes in what one knows and therefore $\sim_i$ is contained in $\equiv_i$. The interesting condition of the relations $\equiv_i$ and $\sim_i$ is condition *d*. This condition means that if $s, t, u \in S$, if $s$, $t$ are both possible according to the knowledge of person $i$, and if, when he is in state $t$, he believes that $u$ is possible, he also believes that $u$ is possible when he is in state $s$, since he cannot distinguish between $t$ and $s$.

We may prove some basic properties of $\sim_i$.

**Lemma 2.2.** *$\sim_i$ is euclidean and transitive.*

**Proof.** Suppose $s, t, u \in S$ and $s \sim_i t$ and $s \sim_i u$. In order to prove that $\sim_i$ is euclidean we have to prove that $t \sim_i u$. If $s \sim_i t$, then $t \equiv_i s$ because $\sim_i$ is contained in $\equiv_i$ and $\equiv_i$ is symmetric. However $s \sim_i u$, therefore $t \sim_i u$ by condition (4)(d) in Definition 2.1.

Suppose $s, t, u \in S$ and $s \sim_i t$ and $t \sim_i u$. In order to prove that $\sim_i$ is transitive we have to prove that $s \sim_i u$. If $s \sim_i t$, then $s \equiv_t t$ because $\sim_i$ is contained in $\equiv_i$. However $t \sim_i u$ and we may conclude that $s \sim_i u$ by condition (4)(d) in Definition 2.1. □

Let us now define two other binary relations corresponding to the connectives $\equiv_i$ and $\sim_i$:

$$\equiv^* \overset{\text{def}}{=} \left( \bigcup_{i \in \text{People}} \equiv_i \right)^+ \quad \text{and} \quad \sim^* \overset{\text{def}}{=} \left( \bigcup_{i \in \text{People}} \sim_i \right)^+,$$

where $^+$ represents transitive closure. It is easy to see that $\equiv^*$ is also reflexive and symmetric because every $\equiv_i$ is reflexive and symmetric. Therefore $\equiv^*$ is an equivalence relation. On the contrary, $\sim^*$ is transitive but not necessarily reflexive or symmetric. Suppose now that a model $\mathcal{M} = \langle S, s, l, \equiv_1, \equiv_2, \ldots, \equiv_n, \sim_1, \sim_2, \ldots, \sim_n \rangle$ is given. For any state $u$ we shall define the truth value of any formula $a \in \Gamma$.

**Definition 2.3.** If $p \in \text{Pvar}$, then $p|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow p \in l(u)$. For formulas $a$ and $b$ of $\Gamma$,

$$\neg a|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow a|^u_{\mathcal{M}} = \textbf{false},$$

$$a \vee b|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow a|^u_{\mathcal{M}} = \textbf{true} \text{ or } b|^u_{\mathcal{M}} = \textbf{true},$$

$$\text{K}_i a|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow \text{for all states } t \text{ such that } u \equiv_i t, \text{ one has } a|^t_{\mathcal{M}} = \textbf{true},$$

$$\text{B}_i a|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow \text{for all states } t \text{ such that } u \sim_i t, \text{ one has } a|^t_{\mathcal{M}} = \textbf{true},$$

$$\mathcal{K} a|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow \text{for all states } t \text{ such that } u \equiv^* t, \text{ one has } a|^t_{\mathcal{M}} = \textbf{true},$$

$$\mathcal{B} a|^u_{\mathcal{M}} = \textbf{true} \Leftrightarrow \text{for all states } t \text{ such that } u \sim^* t, \text{ one has } a|^t_{\mathcal{M}} = \textbf{true}.$$

The definition above is clearly a correct definition by induction on the structure of the formula $a$. Satisfaction and validity will now be defined in the usual way.

**Definition 2.4.** Let $\mathcal{M}$ be a model as above and $a$ a formula of $\Gamma$. We say that $\mathcal{M}$ satisfies $a$ and write $\mathcal{M} \models a$ iff $a|^s_{\mathcal{M}} = \textbf{true}$.

**Definition 2.5.** A formula $a$ is valid ($\models a$) iff it is satisfied by all models.

### 2.3. *The logics*

Our logics are best viewed as composed of a number of levels. The *first* level concerns classical propositional calculus:

(A0) A suitable axiomatization of the propositional calculus;

(R0) (Modus Ponens) If $\vdash a$ and $\vdash a \rightarrow b$ then $\vdash b$.

The *second* level concerns general truths about knowledge and common knowledge as found in [9] ($\rightarrow$ associates to the right):

(A1) $\text{K}_i(a \rightarrow b) \rightarrow \text{K}_i a \rightarrow \text{K}_i b$ for any $i \in \text{People}$;

(A2) $K_i a \rightarrow a$ for any $i \in$ People;

(A3) $\neg K_i a \rightarrow K_i \neg K_i a$ for any $i \in$ People;

(A4) $\mathcal{K}(a \rightarrow b) \rightarrow \mathcal{K}a \rightarrow \mathcal{K}b$;

(A5) $\mathcal{K}a \rightarrow K_i a$ for any $i \in$ People;

(A6) $\mathcal{K}a \rightarrow K_i \mathcal{K}a$ for any $i \in$ People;

(A7) $\mathcal{K}(a \rightarrow \mathcal{E}a) \rightarrow a \rightarrow \mathcal{K}a$;

(R1) common knowledge generalization: if $\vdash a$, then $\vdash \mathcal{K}a$.

The *third* level concerns general truths about belief and common belief:

(A8) $B_i(a \rightarrow b) \rightarrow B_i a \rightarrow B_i b$ for any $i \in$ People (what person $i$ believes is closed under Modus Ponens);

(A9) $\neg B_i$**false** for any $i \in$ People (person $i$ does not believe contradictory things);

(A10) $\mathcal{B}(a \rightarrow b) \rightarrow \mathcal{B}a \rightarrow \mathcal{B}b$ (common belief is closed under Modus Ponens);

(A11) $\mathcal{B}a \rightarrow \mathcal{F}a$ (if something is common belief, then everybody believes it);

(A12) $\mathcal{B}a \rightarrow \mathcal{F}\mathcal{B}a$ (if something is common belief, then everybody believes that it is common belief);

(A13) $\mathcal{B}(a \rightarrow \mathcal{F}a) \rightarrow \mathcal{F}a \rightarrow \mathcal{B}a$ (induction rule).

The *last* and most interesting level describes the interrelation between knowledge and belief, and betweeen common knowledge and common belief.

(A14) $K_i a \rightarrow B_i a$ for any $i \in$ People (one believes everything one knows);

(A15) $B_i a \rightarrow K_i B_i a$ for any $i \in$ People (one knows about his own beliefs: beliefs are conscious);

(A16) $\mathcal{K}a \rightarrow \mathcal{B}a$ (anything that is common knowledge is common belief).

Axiom (A16) is needed to ensure that theorems are common beliefs. We want to emphasize that the interesting formula $B_i a \rightarrow B_i K_i a$ is not included in our system. It would imply $\vdash B_i a \leftrightarrow K_i a$.

We now list some interesting theorems that can be proved from the axioms above and will be used later:

$$K_i \neg a \rightarrow \neg B_i a \qquad \text{for any } i \in \text{People}, \tag{T1}$$

$$B_i a \leftrightarrow K_i B_i a \qquad \text{for any } i \in \text{People}, \tag{T2}$$

$$\neg B_i a \leftrightarrow K_i \neg B_i a \qquad \text{for any } i \in \text{People}, \tag{T3}$$

$$K_i a \leftrightarrow B_i K_i a \qquad \text{for any } i \in \text{People}, \tag{T4}$$

$$\neg K_i a \leftrightarrow B_i \neg K_i a \qquad \text{for any } i \in \text{People}, \tag{T5}$$

$$B_i b \leftrightarrow B_i B_i b, \qquad K_i b \leftrightarrow K_i K_i b \qquad \text{for any } i \in \text{People}, \tag{T6}$$

$$\neg B_i a \leftrightarrow B_i \neg B_i a \qquad \text{for any } i \in \text{People}, \tag{T7}$$

$$B_i(B_i a \rightarrow a) \qquad \text{for any } i \in \text{People}, \tag{T8}$$

$$\mathcal{B}a \leftrightarrow \mathcal{B}\mathcal{F}a, \qquad \mathcal{B}a \leftrightarrow \mathcal{F}\mathcal{B}a, \qquad \mathcal{F}\mathcal{B}a \rightarrow \mathcal{B}\mathcal{F}\mathcal{B}a, \qquad \mathcal{B}a \leftrightarrow \mathcal{B}\mathcal{B}a, \tag{T9}$$

$$\mathcal{K}(a \wedge b) \leftrightarrow \mathcal{K}a \wedge \mathcal{K}b, \qquad \mathcal{B}(a \wedge b) \leftrightarrow \mathcal{B}a \wedge \mathcal{B}b. \tag{T10}$$

**Theorem 2.6.** *Axioms* (A0)–(A16) *and the rules of inference* (R0)–(R1) *are sound and complete for the notion of validity defined above and validity may be decided in deterministic exponential time.*

**Proof.** The proof of the soundness is obvious. The completeness result proceeds by building a "universal" model, i.e., a model in which states are labelled by traces of suitable size in such a way that formulas true at state $s$ are exactly those formulas that appear in the label of $s$. This proof is similar to the corresponding proof in [12], but is different in some interesting details. Some standard notions, such as complete and consistent theory are supposed to be known: see, for example, [12] for definitions. First we shall define a set of formulas to be considered.

*Remark*: In all the following definitions $i \in$ People and $i$ is implicitly universally quantified.

**Definition 2.7.** If $a \in \Gamma$, we define $\Delta(a)$ to be the smallest subset of $\Gamma$ satisfying:

(a) $a \in \Delta(a)$;
(b) $\Delta(a)$ is closed under subformulas;
(c) if $b \in \Delta(a)$ and $b$ does not begin by a negation, then $\neg b \in \Delta(a)$;
(d) if $\mathrm{K}_i b \in \Delta(a)$, then $\mathrm{B}_i b \in \Delta(a)$ and $\mathrm{B}_i \mathrm{K}_i b \in \Delta(a)$;
(e) if $\mathcal{K} b \in \Delta(a)$, then $\mathrm{K}_i \mathcal{K} b \in \Delta(a)$, $\mathrm{K}_i b \in \Delta(a)$ and $\mathcal{B} b \in \Delta(a)$;
(f) if $\mathcal{B} b \in \Delta(a)$, then $\mathrm{B}_i \mathcal{B} b \in \Delta(a)$ and $\mathrm{B}_i b \in \Delta(a)$.

It is easy to see that the size of $\Delta(a)$ is linear in the length of $a$.

Now, we want to restrict our attention to those subsets of $\Gamma$ that are reasonable candidates for labels.

**Definition 2.8.** Let $E \subseteq \Gamma$ be closed under subformulas. $D \subseteq E$ is a standard set on $E$ if it satisfies:

(a) if $\neg b \in E$, $\neg b \in D \Leftrightarrow b \notin D$;
(b) if $b \vee c \in E$, then $b \vee c \in D \Leftrightarrow b \in D$ or $c \in D$;
(c) if $\mathrm{K}_i b \in D$, then $b \in D$, $\mathrm{B}_i b \in D$ and $\mathrm{B}_i \mathrm{K}_i b \in D$;
(d) if $\mathrm{K}_i \mathrm{B}_i b \in E$ and $\mathrm{B}_i b \in D$, then $\mathrm{K}_i \mathrm{B}_i b \in D$;
(e) if $\mathrm{B}_i \mathrm{K}_i b \in D$, then $\mathrm{K}_i b \in D$;
(f) if $\mathcal{K} b \in D$, then $\mathcal{B} b \in D$, $\mathrm{K}_i \mathcal{K} b \in D$ and $\mathrm{K}_i b \in D$;
(g) if $\mathcal{B} b \in D$, then $\mathrm{B}_i \mathcal{B} b \in D$ and $\mathrm{B}_i b \in D$.

The sets in which all beliefs are true are of special interest.

**Definition 2.9.** $F \subseteq \Gamma$ is a normal set if it satisfies: if $\mathrm{B}_i b \in F$, then $b \in F$.

We now define two binary relations between subsets of $\Gamma$.

**Definition 2.10.** Let $D_1$ and $D_2$ be subsets of $\Gamma$. We say that $D_1 \sim_i D_2$ iff

(a) $D_2$ is normal;
(b) for all $b \in \Gamma$, $\mathrm{B}_i b \in D_1 \Leftrightarrow \mathrm{B}_i b \in D_2$.

**Definition 2.11.** Let $D_1$ and $D_2$ be subsets of $\Gamma$. We say that $D_1 \equiv_i D_2$ iff

(a) for all $b \in \Gamma$, $\mathrm{B}_i b \in D_1 \Leftrightarrow \mathrm{B}_i b \in D_2$;

(b) for all $b \in \Gamma$, $\mathrm{K}_i b \in D_1 \Leftrightarrow \mathrm{K}_i b \in D_2$.

Using the theorems that were mentioned before, it is easy to prove that if $T_1 \subseteq \Gamma$ and $T_2 \subseteq \Gamma$ are consistent and complete theories, then

(1) $T_1$ and $T_2$ are standard sets on $\Gamma$;

(2) $T_1 \equiv_i T_2$ iff for all $b \in \Gamma$, $\mathrm{K}_i b \in T_1 \Rightarrow b \in T_2$;

(3) $T_1 \sim_i T_2$ iff for all $b \in \Gamma$, $\mathrm{B}_i \mathrm{b} \in T_1 \Rightarrow b \in T_2$.

We may prove general properties about the relations $\sim_i$ and $\equiv_i$.

**Lemma 2.12.** *The relations* $\equiv_i$ *and* $\sim_i$ *defined in Definitions* 2.10 *and* 2.11 *have the following properties*:

(a) $\equiv_i$ *is an equivalence relation* (*reflexive, transitive and symmetric*).

(b) *If* $E \subseteq \Gamma$ *is closed under subformulas, and* $D_1$ *and* $D_2$ *are standard sets on* $E$, *then if* $D_1 \sim_i D_2$, *then* $D_1 \equiv_i D_2$;

(c) *For any* $D_1, D_2, D_3$ *subsets of* $\Gamma$, *if* $D_1 \equiv_i D_2$ *and* $D_2 \sim_i D_3$, *then* $D_1 \sim_i D_3$.

**Proof.** (a): Obvious.

(b): Suppose there exist $D_1$, $D_2$ subsets of $E$ such that $D_1 \sim_i D_2$. We have to prove that $D_1$ and $D_2$ satisfy the conditions of Definition 2.11. Condition (a) is obvious by condition (b) of the definition of $\sim_i$. For condition (b) $\mathrm{K}_i b \in D_1$ iff $\mathrm{B}_i \mathrm{K}_i b \in D_1$ since $D_1$ is a standard set, $\mathrm{B}_i \mathrm{K}_i b \in D_1$ iff $\mathrm{B}_i \mathrm{K}_i b \in D_2$ since $D_1 \sim_i D_2$, and $\mathrm{B}_i \mathrm{K}_i b \in D_2$ iff $\mathrm{K}_i b \in D_2$ since $D_2$ is a standard set.

(c): We have to prove that for any $D_1$, $D_2$, $D_3$ subsets of $\Gamma$ if $D_1 \equiv_i D_2$ and $D_2 \sim_i D_3$, then $D_1 \sim_i D_3$. $D_3$ is normal because $D_2 \sim_i D_3$. $\mathrm{B}_i b \in D_1 \Leftrightarrow \mathrm{B}_i b \in D_2$ since $D_1 \equiv_i D_2$. But if $\mathrm{B}_i b \in D_2$, then $\mathrm{B}_i b \in D_3$ since $D_2 \sim_i D_3$. □

We want to remark that the restrictions of $\equiv_i$ and $\sim_i$ on a subset of $\Gamma$ keep the properties of Lemma 2.12.

**Lemma 2.13.** *Let* $D_1$ *and* $D_2$ *be subsets of* $\Gamma$. *If* $E \subseteq \Gamma$ *is closed under subformulas, then*

(a) $D_1 \equiv_i D_2$ *iff* $D_1 \cap E \equiv_i D_2 \cap E$;

(b) $D_1 \sim_i D_2$ *iff* $D_1 \cap E \sim_i D_2 \cap E$.

**Proof.** Obvious. □

We shall now define an iteration process of elimination on the standard sets on $\Delta(a)$. The main goal of this process is to reach a situation in which we are left with a group of traces which can be used in constructing a universal model.

**Definition 2.14.** For all $l \geqslant 0$ we define a set $\mathcal{W}_l$ of standard sets on $\Delta(\alpha)$: $\mathcal{W}_0$ is the set of the standard sets on $\Delta(a)$. For $l \geqslant 0$, let $\mathcal{W}_{l+1}$ consists of all those $D$'s $\in \mathcal{W}_l$ that satisfy

(1) there exists a $D_1 \in \mathcal{W}_l$ such that $D \sim_i D_1$;

(2) for $\mathrm{K}_ib \in \Delta(a)$ if $\mathrm{K}_ib \notin D$, then there exists a $D_1 \in \mathcal{W}_l$ such that $D \equiv_i D_1$ and $b \notin D_1$;

(3) for $\mathrm{B}_ib \in \Delta(a)$ if $\mathrm{B}_ib \notin D$, then there exists a $D_1 \in \mathcal{W}_l$ such that $D \sim_i D_1$ and $b \notin D_1$;

(4) for $\mathcal{K}b \in \Delta(a)$ if $\mathcal{K}b \notin D$, then there exists a $k \geqslant 0$ and for all $0 \leqslant j \leqslant k$ there exists a $D_j \in \mathcal{W}_l$ such that $D = D_0$, $D_j \equiv_{g_j} D_{j+1}$ for $j = 0, \ldots, k-1$, $g_j \in$ People and $b \notin D_k$;

(5) for $\mathcal{B}b \in \Delta(a)$ if $\mathcal{B}b \notin D$, then there exists a $k \geqslant 0$ and for all $0 \leqslant j \leqslant k$ there exists a $D_j \in \mathcal{W}_l$ such that $D = D_0$, $D_j \sim_{g_j} D_{j+1}$ for $j = 0, \ldots, k-1$, $g_j \in$ People, and $b \notin D_k$.

Clearly, from the finiteness of $\mathcal{W}_0$ there is some $i_0$ where this construction closes up, i.e., for every $j > i_0$, $\mathcal{W}_j = \mathcal{W}_{i_0}$. Accordingly, we set $\mathcal{W} = \mathcal{W}_{i_0}$. The main results concerning $\mathcal{W}$ are the following:

**Lemma 2.15.** *For any $D \subseteq \Delta(a)$ if there exists some consistent and complete theory $T$ such that $D = T \cap \Delta(a)$, then $D \in \mathcal{W}$.*

**Proof.** We show, by induction on $k$, that such a set is in $\mathcal{W}_k$ for every $k$.

*Basis step* ($k = 0$): It is obvious from the axiomatic system and the theorems above and by [12, Lemma 1] that any such $D \subseteq \Delta(a)$ is a standard set.

*Induction step* ($k > 0$): We have to prove that any $D \in \mathcal{W}_{k-1}$ such that there exists some consistent and complete theory $T$ and $D = T \cap \Delta(a)$ satisfies the conditions of Definition 2.14 with $l = k - 1$.

(1): Let

$$F \overset{\text{def}}{=} \{b \mid \mathrm{B}_ib \in D\} \cup \{\mathrm{B}_ib \mid \mathrm{B}_ib \in D\} \cup \{\neg\mathrm{B}_ic \mid \neg\mathrm{B}_ic \in D\}.$$

We shall prove that the theory $F$ is consistent. Suppose that $F$ is inconsistent. Then there are formulas $\mathrm{B}_ib_j \in D$, $j = 1, \ldots, n$; $\neg\mathrm{B}_ic_l \in D$, $l = 1, \ldots, m$ and $d_k$, $k = 1, \ldots, p$ such that $\mathrm{B}_id_k \in D$ for $k = 1, \ldots, p$ and

$$\vdash \mathrm{B}_ib_1 \wedge \mathrm{B}_ib_2 \wedge \cdots \wedge \mathrm{B}_ib_n \wedge \neg\mathrm{B}_ic_1 \wedge \neg\mathrm{B}_ic_2 \wedge \cdots \wedge \neg\mathrm{B}_ic_m \\ \rightarrow \neg(d_1 \wedge d_2 \wedge \cdots \wedge d_p).$$

It follows that

$$\vdash \mathrm{B}_i[\mathrm{B}_ib_1 \wedge B_ib_2 \wedge \cdots \wedge \mathrm{B}_ib_n \wedge \neg\mathrm{B}_ic_1 \wedge \neg\mathrm{B}_ic_2 \wedge \cdots \wedge \neg\mathrm{B}_ic_m \\ \rightarrow (d_1 \wedge d_2 \wedge \cdots \wedge d_p)]$$

by (R1), (A16) and (A11). Then

$$\vdash \mathrm{B}_i\mathrm{B}_ib_1 \wedge \mathrm{B}_i\mathrm{B}_ib_2 \wedge \cdots \wedge \mathrm{B}_i\mathrm{B}_ib_n \wedge \mathrm{B}_i\neg\mathrm{B}_ic_1 \wedge \mathrm{B}_i\neg\mathrm{B}_ic_2 \wedge \cdots \wedge \mathrm{B}_i\neg\mathrm{B}_ic_m$$
$$\rightarrow \mathrm{B}_i\neg(d_1 \wedge d_2 \wedge \cdots \wedge d_p)$$

by (A8), (T10). $\mathrm{B}_ib_j \in D$ and $\neg\mathrm{B}_ic_l \in D$, therefore

$$\mathrm{B}_i\mathrm{B}_ib_1 \wedge \mathrm{B}_i\mathrm{B}_ib_2 \wedge \cdots \wedge \mathrm{B}_i\mathrm{B}_ib_n \wedge \mathrm{B}_i\neg\mathrm{B}_ic_1 \wedge \cdots \wedge \mathrm{B}_i\neg\mathrm{B}_ic_m \in T$$

by (T6), (T7) and by [12, Lemma 1]. We can conclude that $\mathrm{B}_i\neg(d_1 \wedge c_2 \wedge \cdots \wedge d_p) \in T$. But $\mathrm{B}_id_j \in D$, $j = 1, \ldots, p$ and therefore $\mathrm{B}_id_j \in T$, $j = 1, \ldots, p$, and it follows that $\mathrm{B}_i(d_1 \wedge d_2 \wedge \cdots \wedge d_p) \in T$, contradiction. So $F$ is a consistent theory and it may be completed by [12, Lemma 1(d)] to a consistent and complete theory $T'$. Let $D_1 =^{\mathrm{def}} T' \cap \Delta(a)$. By the induction hypothesis, $D_1 \in \mathcal{W}_{k-1}$. We shall now show that $D \sim_i D_1$, i.e., satisfies the conditions of Definition 2.10.

First if $\mathrm{B}_ib \in D$, then, by the definition of $F$, $\mathrm{B}_ib \in F$ and $\mathrm{B}_ib \in T'$ and $\mathrm{B}_ib \in D_1$ because $\mathrm{B}_ib \in \Delta(a)$. Suppose, on the contrary, that $\mathrm{B}_ib \notin D$; we have to distinguish between two different cases: If $\mathrm{B}_ib \notin \Delta(a)$, then $\mathrm{B}_ib \notin D_1$. But if $\mathrm{B}_ib \in \Delta(a) - D$, then $\neg\mathrm{B}_ib \in D$ ($\neg\mathrm{B}_ib \in \Delta(a)$ and $D$ is standard set) and $\neg\mathrm{B}_ib \in F$ (by the definition of $F$) and $\neg\mathrm{B}_ib \in D_1$.

It remains to be shown that $D_1$ is normal, i.e., satisfies the condition of Definition 2.9. Suppose $\mathrm{B}_ib \in D_1$, then $\mathrm{B}_ib \in D$ (as proved just above) and $b \in F$ by the definition of $F$; therefore, $b \in D_1$.

(2): Let $\mathrm{K}_ib \in \Delta(a)$ such that $\mathrm{K}_ib \notin D$. If the theory

$$R \overset{\mathrm{def}}{=} \{\mathrm{K}_ic \mid \mathrm{K}_ic \in D\} \cup \{\neg\mathrm{K}_id \mid \neg\mathrm{K}_id \in D\} \cup \{\mathrm{B}_ie \mid \mathrm{B}_ie \in D\}$$
$$\cup \{\neg\mathrm{B}_if \mid \neg\mathrm{B}_id \in D\} \cup \{\neg b\}$$

is consistent, then it may be completed to a satisfactory $T_1$ by [12, Lemma 1(d)] and we may define $D_1 =^{\mathrm{def}} T_1 \cap \Delta(a)$ as in the previous case. Suppose $R$ is inconsistent. Then there are formulas $\mathrm{K}_ic_j \in D$, $j = 1, \ldots, n$; $\neg\mathrm{K}_id_l \in D$, $l = 1, \ldots, m$; $\mathrm{B}_ie_k \in D$, $k = 1, \ldots, p$ and $\neg\mathrm{B}_if_g \in D$, $g = 1, \ldots, q$ and

$$\vdash \mathrm{K}_ic_1 \wedge \cdots \wedge \mathrm{K}_ic_n \wedge \neg\mathrm{K}_id_1 \wedge \cdots \wedge \neg\mathrm{K}_id_m \wedge \mathrm{B}_ie_1 \wedge \cdots \wedge \mathrm{B}_ie_p$$
$$\wedge \neg\mathrm{B}_if_1 \wedge \cdots \wedge \neg\mathrm{B}_if_q \rightarrow b.$$

It follows that

$$\vdash \mathrm{K}_i[\mathrm{K}_ic_1 \wedge \cdots \wedge \mathrm{K}_ic_n \wedge \neg\mathrm{K}_id_1 \wedge \cdots \wedge \neg\mathrm{K}_id_m \wedge \mathrm{B}_ie_1 \wedge \cdots \mathrm{B}_ie_p$$
$$\wedge \neg\mathrm{B}_if_1 \wedge \cdots \wedge \neg\mathrm{B}_if_q \rightarrow b]$$

by (R1) and (A5). Then

$$\vdash \mathrm{K}_i\mathrm{K}_ic_1 \wedge \cdots \wedge \mathrm{K}_i\mathrm{K}_ic_n \wedge \mathrm{K}_i\neg\mathrm{K}_id_1 \wedge \cdots \wedge \mathrm{K}_i\neg\mathrm{K}_id_m \wedge \mathrm{K}_i\mathrm{B}_ie_1 \wedge \cdots \wedge \mathrm{K}_i\mathrm{B}_ie_p$$
$$\wedge \mathrm{K}_i\neg\mathrm{B}_if_1 \wedge \cdots \wedge \mathrm{K}_i\neg\mathrm{B}_if_q \rightarrow \mathrm{K}_ib$$

by (A1) and (T10).

Since $D = T \cap \Delta(a)$ and $\mathrm{K}_i c_j \in D$, $j = 1, \ldots, n$; $\neg\mathrm{K}_i d_l \in D$, $l = 1, \ldots, m$; $\mathrm{B}_i e_k \in D$, $k = 1, \ldots, p$ and $\neg\mathrm{B}_i f_g \in D$, $g = 1, \ldots, q$, we can conclude, using [12, Lemma 1], that $\mathrm{K}_i\mathrm{K}_i c_j \in T$, $j = 1, \ldots, n$; $\mathrm{K}_i\neg\mathrm{K}_i d_l \in T$, $l = 1, \ldots, m$ (by (A3) and (T6)), $\mathrm{K}_i\mathrm{B}_i e_k \in T$, $k = 1, \ldots, p$ (by (T2)); $\mathrm{K}_i\neg\mathrm{B}_i f_k \in T$, $g = 1, \ldots, q$ (by (T3)). Therefore, $\mathrm{K}_i b \in T$, a contradiction.

(3): Similar to the proof of (2) and simpler.

(4): Suppose that $\mathscr{K}b \in \Delta(a)$ and $\mathscr{K}b \notin D$. We may define

$$\mathscr{D} \overset{\text{def}}{=} \left\{F \mid F \in \mathscr{W}_{k-1}, D\left(\bigcup_{i \in \text{People}} \equiv_i\right)^+ F\right\}$$

and we have to prove that there exists an $F' \in \mathscr{D}$ such that $b \notin F$. If $G$ is a subset of $\Gamma$, we define the characteristic formula of $G$ by $\chi_G =^{\text{def}} \bigwedge_{a \in G} a$. The characteristic formula $\chi_G$ is not in general in $G$, but it is easy to prove that for any consistent and complete theory $T$, $G = T \cap \Delta(a) \Leftrightarrow \chi_G \in T$. We may define $\psi =^{\text{def}} \bigvee_{F \in \mathscr{D}} \chi_F$. Notice that for any consistent and complete theory $T$, $T \cap \Delta(a) \in \mathscr{D} \Leftrightarrow \psi \in T$. We have to prove that $\vdash \psi \rightarrow \bigwedge_i \mathrm{K}_i\psi$. Suppose not. Then there is a consistent and complete theory $S$ such that $\psi \in S$ and $\bigwedge_i \mathrm{K}_i\psi \notin S$. It follows that there exists a $j \in$ People such that $\mathrm{K}_j\psi \notin S$; therefore there exists a consistent and complete theory $S'$ and $S \equiv_j S'$ and $\psi \notin S'$. We may conclude that $S' \cap \Delta(a) \notin \mathscr{D}$. But $S \equiv_j S'$ and therefore $S \cap \Delta(a) \equiv_j S' \cap \Delta(a)$ by Lemma 2.13, and $S \cap \Delta(a) \in \mathscr{D}$ and we may conclude that $S' \cap \Delta(a) \in \mathscr{D}$. A contradiction.(*)

So, from $\vdash \psi \rightarrow \bigwedge_i \mathrm{K}_i\psi$ it follows that $\vdash \mathscr{K}[\psi \rightarrow \bigwedge_i \mathrm{K}_i\psi]$ by (R1), and by using the induction rule (A7) we may conclude that $\vdash \psi \rightarrow \mathscr{K}\psi$. But $\vdash \chi_D \rightarrow \psi$ because $D \in \mathscr{D}$; therefore $\vdash \chi_D \rightarrow \mathscr{K}\psi$.

Now we may finish the proof that there exists an $F' \in \mathscr{D}$ such that $b \notin F$. Suppose that $\forall F \in \mathscr{D}$, $b \in F$. It follows that $\vdash \psi \rightarrow b$; then $\vdash \mathscr{K}\psi \rightarrow \mathscr{K}b$ and $\vdash \chi_D \rightarrow \mathscr{K}b$; but $\chi_D \in T$ and then $\mathscr{K}b \in T$ and we may conclude that $\mathscr{K}b \in D$, contradiction.

(5): Suppose that $\mathscr{B}b \in \Delta(a)$ and $\mathscr{B}b \notin D$. The beginning of the proof is similar to the proof of the previous case till (*) by replacing $\equiv_i$ by $\sim_i$, $\mathscr{K}$ by $\mathscr{B}$ and $\mathrm{K}_i$ by $\mathrm{B}_i$. So, from $\vdash \psi \rightarrow \bigwedge_i \mathrm{B}_i\psi$ it follows that $\vdash \mathscr{B}[\psi \rightarrow \bigwedge_i \mathrm{B}_i\psi]$ by (R1) and (A16), and by using the induction rule (A13) we may conclude that $\vdash \bigwedge_i \mathrm{B}_i\psi \rightarrow \mathscr{B}\psi$.

We have to prove that $\vdash \chi_D \rightarrow \mathrm{B}_i\psi$. Suppose not. Then $\mathrm{B}_i\psi \notin T$ since $\chi_D \in T$. Therefore there exists a consistent and complete theory $T'$ and $T \sim_i T'$ and $\psi \notin T'$. We may conclude that $T' \cap \Delta(a) \notin \mathscr{D}$. But $T \sim_i T'$ and therefore $T \cap \Delta(a) \sim_i T' \cap \Delta(a)$ by Lemma 2.13, and $T \cap \Delta(a) \in \mathscr{D}$ and we may conclude that $T' \cap \Delta(a) \in \mathscr{D}$, a contradiction.

So, $\vdash \chi_D \rightarrow \bigwedge_i \mathrm{B}_i\psi$ and $\vdash \bigwedge_i \mathrm{B}_i\psi \rightarrow \mathscr{B}\psi$; therefore, $\vdash \chi_D \rightarrow \mathscr{B}\psi$, and we may finish the proof as in the previous case. □

**Lemma 2.16.** *The restrictions of the relations* $\equiv_i$ *and* $\sim_i$ *on* $\mathscr{W}$ *have the following properties*:

(1) *all the properties of Definition* 2.1:

(a) $\equiv_i$ *is an equivalent relation* (*reflexive, transitive and symmetric*),

(b) $\sim_i$ *is serial* (*for all* $s \in \mathcal{W}$ *there is some* $t \in \mathcal{W}$ *such that* $s \sim_i t$).

(c) $\sim_i$ *is contained in* $\equiv_i$ $(\sim_i \subseteq \equiv_i)$,

(d) *for any* $s, t, u \in \mathcal{W}$ *if* $s \equiv_i t$ *and* $t \sim_i u$, *then* $s \sim_i u$;

(2) *for* $\mathrm{K}_i b \in \Delta(a)$ *if* $\mathrm{K}_i b \notin D$, *then there exists a* $D_1 \in \mathcal{W}$ *such that* $D \equiv_i D_1$ *and* $b \notin D_1$;

(3) *for* $\mathrm{B}_i b \in \Delta(a)$ *if* $\mathrm{B}_i b \notin D$, *then there exists a* $D_1 \in \mathcal{W}$ *such that* $D \sim_i D_1$ *and* $b \notin D_1$;

(4) *for* $\mathcal{K}b \in \Delta(a)$ *if* $\mathcal{K}b \notin D$, *then there exists* $k \geq 0$ *and for all* $0 \leq j \leq k$ *there exists a* $D_j \in \mathcal{W}$ *such that* $D = D_0$, $D_j \equiv_{g_j} D_{j+1}$ *for* $j = 0, \ldots, k-1$, $g_j \in \mathrm{People}$, *and* $b \notin D_k$;

(5) *for* $\mathcal{B}b \in \Delta(a)$ *if* $\mathcal{B}b \notin D$, *then there exists a* $k \geq 0$ *and for all* $0 \leq j \leq k$ *there exists a* $D_j \in \mathcal{W}$ *such that* $D = D_0$, $D_j \sim_{g_j} D_{j+1}$ *for* $j = 0, \ldots, k-1$, $g_j \in \mathrm{People}$, *and* $b \notin D_k$.

**Proof.** Properties (a), (c), and (d) were proved in Lemma 2.12. Properties (b) and (2)-(5) are straightforward from the construction of $\mathcal{W}$ in Definition 2.14. □

**Proof of Theorem 2.6.** (*continued*). The completeness proof now proceeds in the following way: suppose that $\nvdash a$; we shall build a model that does not satisfy $a$. First, by [12, Lemma 1(e)] there is a consistent and complete theory $T_a$ that contains $\neg a$. From Lemma 2.15, $D_a = T_a \cap \Delta(a)$ is in $\mathcal{W}$. The model $\mathcal{M} = \langle S, s, l, \equiv_1, \equiv_2, \ldots, \equiv_n, \sim_1, \sim_2, \ldots, \sim_n \rangle$ that does not satisfy $a$ is defined in the following way:

(1) $S = \mathcal{W}$,

(2) $s = D_a$,

(3) $l(D) = \{p \mid p \in D\}$,

(4) $\equiv_i$ and $\sim_i$ are defined as in Definitions 2.10 and 2.11.

Our main result concerning the model $\mathcal{M}$ is the following lemma.

**Lemma 2.17.** *Let* $b \in \Delta(a)$ *and* $s \in \mathcal{W}$; *then* $b|^s_{\mathcal{M}} = \mathbf{true} \Leftrightarrow b \in S$.

**Proof.** The proof is by induction on the structure of $b$.

*Case* $b = p$:

$$p|^s_{\mathcal{M}} = \mathbf{true} \Leftrightarrow p \in l(s) \text{ by the definition of the model.}$$

*Case* $b = \neg c$: $\neg c \in s \Leftrightarrow c \notin s$ (since $s$ is a standard set) $\Leftrightarrow$ $c|^s_{\mathcal{M}} = \mathbf{false}$ by the induction hypothesis.

*Case* $b = c \vee d$: $c \vee d \in s \Leftrightarrow c \in s$ or $d \in s$ (since $s$ is a standard set)$\Leftrightarrow$c$|^s_{\mathcal{M}} = \mathbf{true}$ or $d|^s_{\mathcal{M}} = \mathbf{true}$ by the induction hypothesis.

*Case* $b = \mathrm{K}_i c$: $\mathrm{K}_i c \in s \Rightarrow \mathrm{K}_i c \in t$ for all $t \in \mathcal{W}$ such that $s \equiv_i t$ by the definition of $\equiv_i$ $\Rightarrow c \in t$ since $t$ is standard $\Rightarrow c|^t_{\mathcal{M}} = \mathbf{true}$ by the induction hypothesis. Suppose now that $\mathrm{K}_i c \notin s$. Since $\mathrm{K}_i c \in \Delta(a)$ it follows by condition 2 of Lemma 2.16 that there exists a $t \in \mathcal{W}$ such that $s \equiv_i t$ and $b \notin t$ and we may conclude by the induction hypothesis that $b|^s_{\mathcal{M}} = \mathbf{false}$.

*Case* $b = \mathrm{B}_i c$: $\mathrm{B}_i c \in s \Rightarrow \mathrm{B}_i c \in t$ for all $t \in \mathcal{W}$ such that $s \sim_i t$ by the definition of $\sim_i \Rightarrow c \in t$ since $t$ is a normal set $\Rightarrow c|^t_{\mathcal{M}} = \mathbf{true}$ by the induction hypothesis. The other direction is like the previous case by condition (3) of Lemma 2.16.

*Case* $b = \mathcal{K}c$: Suppose $\mathcal{K}c \in s$. We have to prove that, for any $g \geqslant 0$ and for any $s_0, s_1, \ldots, s_j, \ldots, s_g$, if $s_j \in \mathcal{W}$, $j = 0, \ldots, g$; $s = s_0$, $s_j \equiv_{k_j} s_{j+1}$ for $j = 0, \ldots, g-1$, $k_j \in$ People, then $c \in s_g$. $\mathcal{K}c \in s_j$; therefore, $\mathrm{K}_{k_j}\mathcal{K}c \in s_j$ since $s_j$ is standard; then $\mathcal{K}c \in s_{j+1}$. But $\mathcal{K}c \in s_0$; therefore $\mathcal{K}c \in s_g$ and $c \in s_g$ since $s_g$ is standard. By the induction hypothesis we may conclude that $c|^{s_g}_{\mathcal{M}} = \mathbf{true}$. The opposite direction is similar to the previous case by condition (5) of Lemma 2.16.

*Case* $b = \mathcal{B}c$: Similar to the previous case. □

**Proof of Theorem 2.6** (*conclusion*). So we can finish the proof of the completeness by concluding that $\mathcal{M} \nvDash a$. □

## 3. Time

We may now extend our logic to capture time by adding new modal operators to the language. The new operators are ○ (next), □ (always) and **Until** (until).

We think that when talking about how beliefs are changing in time, one must distinguish between at least two different notions of belief. First, belief can mean readiness to bet. "Person $i$ believes that this afternoon it will rain" means, operationally, that he will take his raincoat with him in the morning. If it does not rain, no problem, he will be slightly inconvenienced by having to take his raincoat back and forth. But a second acceptation is possible, in which belief is a much more serious matter. One cannot allow reality to contradict one's beliefs because that would be too traumatic an experience. Therefore one may believe only things that may not, ever, under any circumstances, be shown to be false. One may not believe that it will rain because he could come to know that his belief was erroneous and he does not want to take that chance. One may believe things that one knows to be true or things that cannot be proven false. For example, one may believe the Continuum Hypothesis (or its negation) since its negation is not a theorem of set theory. This meaning of belief is perhaps close to the meaning of religious belief. So, talking about this last meaning of belief (the "serious" meaning), if person $i$ believes that something will be true tomorrow, it must be that he knows that he will not discover tomorrow that it is wrong. Therefore the following axiom seems reasonable for this second interpretation:

(A17) $\mathrm{B}_i \bigcirc a \rightarrow \mathrm{K}_i \bigcirc \neg \mathrm{K}_i \neg a$ for any $i \in$ People.

Now, if person $i$ believes that tomorrow $a$ will be true, since he cannot find tomorrow that he is wrong, he has all the reasons to persist in his beliefs and we may accept

(A18) $B_i \bigcirc a \rightarrow \bigcirc B_i a$ for any $i \in$ People.

If (A18) is accepted, one can also accept

(A19) $\mathcal{B} \bigcirc a \rightarrow \bigcirc \mathcal{B} a.$

Now one may notice that (A17) is provable from (A18) and forget (A17). Fagin and Halpern in [2] argue that (A18) "is moving us away from our goal of capturing realistic notions of belief". Indeed, in everyday life, the verb "believe" is probably used with a meaning of the type "readiness to bet". A logical, nonnumeric, convincing treatment of such a notion is problematic.

In such an interpretation of belief, (the "weak" meaning) axioms (A18) and (A19) do not seem reasonable. Somebody may believe (in the "weak" meaning) that it will rain tomorrow, but tomorrow he will not believe so. Nevertheless, if one believes that tomorrow $a$ is true, one believes (today) that tomorrow one will continue to believe $a$. Therefore we propose the axiom (for the first interpretation):

(A20) $B_i \bigcirc a \rightarrow B_i \bigcirc B_i a$ for any $i \in$ People.

This axiom says nothing about the future and we need some more expressive axiom to capture what happens to our beliefs in time. No one likes to change his beliefs, and one shall change them only if one is forced to do so. For example, somebody who drove to his office in the morning and left his car in the parking lot believes that his car is in the parking lot and in good mechanical condition. He will continue to believe it unless a friend of his calls him to tell him that his car has been damaged by another car or has been stolen. Then he stops believing that his car is in the parking lot, and he knows the opposite. We want to say that if person $i$ believes something, he will keep on believing it until he knows it is false. The way to say so is:

(A21) $B_i \bigcirc a \rightarrow \bigcirc B_i a \vee \bigcirc K_i \neg a$ for any $i \in$ People.

The following is provable from A21, but weaker:

(A22) $B_i \bigcirc a \rightarrow \bigcirc B_i a \vee \bigcirc B_i \neg a$ for any $i \in$ People.

The meaning of this axiom is that one stops believing in something when one *believes* it is false. For example, does the driver (from the previous example) *know* that his car was damaged?! He did not see his car, and maybe his friend was joking?!

An open problem is: find a natural family of models for which the systems considered above are complete.

The ways by which common knowledge and common belief may be achieved in a distributed environment are fundamentally different. Halpern and Moses [6] proved that, in order for common knowledge to be attainable in a system, the system must be capable of simultaneity. This limitation is a direct consequence of the following theorem: $\mathcal{K}a \leftrightarrow \mathrm{K}_i\mathcal{K}a$ for all $i \in \text{People}$. On the contrary, the formula: $\mathrm{B}_i\mathcal{B}a \rightarrow \mathcal{B}a$ is not valid. Note that $\mathcal{B}a \leftrightarrow \mathcal{F}\mathcal{B}a$ is valid. Therefore it could be that $\Box \mathrm{B}_1\mathcal{B}a$ and $\bigcirc\Box \mathrm{B}_2\mathcal{B}a$ and $\bigcirc\bigcirc\Box \mathrm{B}_3\mathcal{B}a \ldots$ and at instant $n$ ($n = |\text{People}|$) $\mathcal{F}\mathcal{B}a$ will be true and will be common belief. We just described a system in which common belief is gained and which is not capable of simultaneity. Similarly, in a system in which, at first, $a$ is common belief, but later on, agent $i$ stops believing $a$ is common belief, common belief is lost, without any simultaneity being necessary. We may conclude that common belief seems a more realistic notion than common knowledge, which can be attainable without simultaneity.

## 4. The puzzle of the wise men

As an example of the possible use of the logic of belief, common belief and time (without knowledge), we provide a new analysis of the puzzle that was analysed in [9]. We use the system that includes a suitable axiomatization of the temporal logic of linear time (see [4]), the axioms (A8)–(A13) and (A19), the rules of inference (R0) and a common belief generalization rule: if $\vdash a$, then $\vdash \mathcal{B}a$.

We think that axiom (A19) seems very reasonable since common belief is a strong notion. Therefore, if it is common belief that tomorrow $a$ will be true, it seems reasonable that tomorrow $a$ will be common belief.

The puzzle could be told the following way. Once upon a time, the happy chairman of the Computer Science Department at Utopia University was told by the President that the good financial situation of the university allowed for some pay increases in his department. The chairman decided to distribute the pay increases among his three professors in a way that would provide for both fun and justice. He asked all three professors to a room and showed them five hats: two white hats and three black hats. He told them: at exactly 12:00 noon I will turn off the light and place a hat on each of your three heads; you are not allowed to put those hats off your head; I will destroy the two remaining hats and will not know myself which of the hats have been destroyed; then I will leave the room and switch the light on; at 1:00 p.m. I will come back and ask whether somebody can tell the colour of his hat; a wrong answer means unemployment, a correct answer means a large pay increase; if one (at least) of you speaks at 1:00 p.m., then the game is finished; if nobody speaks at that time, then I will come back at 2:00 p.m. and ask the same question but offer only a smaller pay increase to anybody who can tell the colour of his hat; this will go on at each hour (for decreasing financial rewards) until one of you speaks up. The chairman did as he said, left the room and switched the light on. As soon as he sat down in his office a telephone call from his wife asked him in no uncertain terms to be home at 2:30 p.m. to greet her parents. He agreed. The question is: how did the chairman know he could leave the university in time, i.e., that he would not have to be back for the 3:00 p.m. visit to his staff?

It should be pointed out that the chairman cannot indeed exclude a possible cooperation between his professors but he cannot either count on such a cooperation, things being as they are.

Of interest to us in this puzzle are both the kind of reasoning used by the professors and the chairman, and the exact list of all assumptions hidden in this puzzle. The notion of common belief will prove itself useful in both respects. We think that this analysis of the puzzle is sharper and more realistic than the analysis using common knowledge that was provided in [9], and we may compare them later. For the description of the, sometimes hidden, assumptions and the participants' reasoning we shall use the following basic propositions (propositional variables). To minimize propositional variables, we build some of the assumptions in the interpretation of those variables. It would not be difficult to state exactly those assumptions by using some more variables. The professors will be numbered 1, 2, 3. The instants of time considered are 12:00 noon, 1:00 p.m., 2:00 p.m. and so on. We use the propositional variable $b_i$ to mean "professor $i$ has, now, a black hat on his head", and $w_i$ to mean "professor $i$ has, now, a white hat on his head". For $i=1,\ldots,3$ we use the propositional variable $d_i$ to mean "professor $i$ declares, now, the colour of his hat". We shall also use the following notations:

$$\alpha_1 \overset{\text{def}}{=} \bigvee_{i=1,\ldots,3} b_i, \qquad \alpha_2 \overset{\text{def}}{=} (b_1 \vee b_2) \wedge (b_2 \vee b_3) \wedge (b_3 \vee b_1),$$

$$\alpha_3 \overset{\text{def}}{=} \bigwedge_{i=1,\ldots,3} b_i.$$

the formula $\alpha_k$ means "there are at least $k$ black hats".

We now proceed to describe the assumptions of the puzzle. *Assumption* 1 is that it is common belief that one has either a black hat or a white hat.

$$\text{(CBWB)} \quad \mathscr{B}\Box(\neg b_i \leftrightarrow w_i) \quad \text{for } i=1,\ldots,3.$$

This assumption is weaker than the corresponding hidden assumption in [9]. In [9] the assumptions that $\neg b_i$ is equivalent to $w_i$ is built in the description of the puzzle and therefore it is common knowledge that hats are either black or white; here however it is only common belief. Therefore one of the hats could be yellow.

*Assumption* 2 is that it is common belief that hats do not change colours and do not move from one head to another.

$$\text{(CBHI)} \quad \mathscr{B}\Box(b_i \leftrightarrow \bigcirc b_i) \wedge \mathscr{B}\Box(w_i \leftrightarrow \bigcirc w_i) \quad \text{for } i=1,\ldots,3.$$

The similar assumption (HI) from [9] is much stronger than this one. In (CBHI) the professors *believe* that the hats do not change colours and do not move from one head to another and that everyone believes that hats do not change colours and so on, but they do not know it and it could also be that it is not true.

*Assumption* 3 is that it is common belief that no professor is blind, i.e. every professor believes he sees the color of the hat of each of his colleagues. It does not mean that one of the professors could not be color blind and maybe he does not *know* the real color of the hats, but he believes that he sees the right color and this fact is common belief.

(CBNB) $\mathcal{B}\Box(b_j \rightarrow \mathrm{B}_i b_j) \wedge \mathcal{B}\Box(w_j \rightarrow \mathrm{B}_i w_j)$ for $i = 1, \ldots, 3, j = 1, \ldots, 3, i \neq j$.

*Assumption* 4 expresses that a professor declares the colour of his hat to be white only when he believes the colour of his hat is white (respectively black); on the other hand, he will declare as soon as he can after he has come to believe the colour of his hat to be white (respectively black). To simplify matters a little bit and to avoid introducing the **Until** connective in our formulas, we assume that once a professor declares the colour of his hat, he goes on declaring it at every subsequent instant (hour).

(CBSR) $\mathcal{B}\Box(\mathrm{B}_i b_i \vee \mathrm{B}_i w_i \leftrightarrow \bigcirc d_i)$ for $i = 1, \ldots, 3$.

This axiom is not weaker than (SR) from [9]. Here a professor does not wait until he *knows* the colour of his hat, as in (SR), but in our version this is only common belief and not common knowledge.

*Assumption* 5 expresses that the fact for a participant to speak up, or, more important, to stay quiet, is public enough to create common belief. In other words, whether a professor speaks up or not is immediately common belief.

(ND1)$\Box(\mathcal{B}d_i \leftrightarrow d_i) \wedge \Box(\mathcal{B}\neg d_i \leftrightarrow \neg d_i)$ for $i = 1, \ldots, 3$.

(ND1) is the only axiom that considers the reality. All the others assumed certain common beliefs.

Our last assumption (*assumption* 6) is that it is common belief that, at 12:00 noon, there is at least one professor donning a black hat:

(CBTWH)$\mathcal{B}\alpha_1$.

Our claim about the puzzle is that if nobody speaks up at 1:00 p.m. or at 2:00 p.m., then every single professor will speak up at 3:00 p.m. In particular, the chairman knows that a final decision will, at the latest, be made at 2:01 p.m. formally, we claim

$$\vdash (\mathrm{ND1}) \wedge (\mathrm{CBWB}) \wedge (\mathrm{CBHI}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBTWH}) \wedge (\mathrm{CBNB})$$

$$\bigvee_{i=1,\ldots,3} \bigcirc d_i \vee \bigvee_{i=1,\ldots,3} \bigcirc\bigcirc d_i \vee \bigwedge_{i=1,\ldots,3} \bigcirc\bigcirc\bigcirc d_i.$$

We want to remark that we do not prove anything about how the professors acquire their beliefs, and when the professors will not speak. Even in the case all the three professors have black hats ($\alpha_3$), we prove that at least one of the professors will speak till 3:00, but we are not able to prove, using our system, that no one will speak before this time. It could be that one of them will start believing the colour of his hat white (respectively black) before 2:00, by e.s.p. for example. Building a logical system which puts general sensible restriction on the way one acquires beliefs is an open problem.

The proof is done in three main parts. First we prove that it is common belief that at 1:00 p.m. if it is common belief that no one speaks up, then it is common belief that there are at least two black hats. In the second part we prove that it is common belief that at 1:00 p. m. if it is common belief that there are at least two black hats, then if at 2:00 p.m. it is common belief that no one speaks up, then at 2:00 it will be common belief that at 3:00 p.m everybody will speak up. In the parts of the proof described above we use only the assumptions that begin with common belief ((CBHI), (CBSR), (CBTWH), (CBNB), (CBWB)) and not assumption (ND1) that considers reality. We use this last axiom only in the third part when we prove our final claim by putting our results from the previous parts together.

The main stages of the proof will be described.

$$(\mathrm{CBNB}) \rightarrow \mathcal{B}(w_j \rightarrow \mathrm{B}_i w_j) \quad \text{for } i=1,\ldots,3, j=1,\ldots,3,\ i\neq j, \tag{1}$$

$$(\mathrm{CBTWH}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBWB}) \rightarrow \mathcal{B}(w_j \wedge w_k \rightarrow \mathrm{B}_i b_i) \tag{2}$$

for $i=1,\ldots,3$, $j=1,\ldots,3$, $k=1,\ldots,3$, $i\neq j$, $j\neq k$, $k\neq i$. This follows from (1). Now we have

$$(\mathrm{CBSR}) \wedge (\mathrm{CBTWH}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBWB}) \rightarrow \mathcal{B}\left(\neg\alpha_2 \rightarrow \bigvee_{i=1,\ldots,3} \bigcirc d_i\right), \tag{3}$$

by (2).

$$(\mathrm{CBHI}) \rightarrow \mathcal{B}(\alpha_2 \leftrightarrow \bigcirc\alpha_2), \tag{4}$$

$$(\mathrm{CBHI}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBTWH}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBWB})$$

$$\rightarrow \mathcal{B}\bigcirc\left(\bigwedge_{i=1,\ldots,3} \neg d_i \rightarrow \alpha_2\right), \tag{5}$$

by (3) and (4).

$$(\mathrm{CBHI}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBTWH}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBWB})$$

$$\rightarrow \bigcirc\mathcal{B}\left(\bigwedge_{i=1,\ldots,3} \neg d_i \rightarrow \alpha_2\right), \tag{6}$$

by (5) and axiom (A19).

$$(\mathrm{CBHI}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBTWH}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBWB})$$

$$\rightarrow \bigcirc\left(\bigwedge_{i=1,\ldots,3} \mathcal{B}\neg d_i \rightarrow \mathcal{B}\alpha_2\right), \tag{7}$$

by (6) and axiom (A10) and (T10)

$$\text{(CBHI)} \wedge \text{(CBSR)} \wedge \text{(CBTWH)} \wedge \text{(CBNB)} \wedge \text{(CBWB)}$$
$$\rightarrow \mathcal{B}\bigcirc\left(\bigwedge_{i=1,\ldots,3} \mathcal{B}\neg d_i \rightarrow \mathcal{B}\alpha_2\right) \tag{8}$$

by (7) and since all the hypotheses begin by $\mathcal{B}$. This is the end of the first part of our proof.

$$\text{(CBWB)} \rightarrow \mathcal{B}(\alpha_2 \wedge w_j \rightarrow b_i) \quad \text{for } i = 1, \ldots, 3, j = 1, \ldots\ldots, 3, \ i \neq j, \tag{9}$$

by propositional calculus.

$$\text{(CBWB)} \rightarrow \mathcal{B}(\mathrm{B}_i\alpha_2 \wedge \mathrm{B}_i w_j \rightarrow \mathrm{B}_i b_i) \quad \text{for } i = 1, \ldots, 3, j = 1, \ldots, 3, \ i \neq j, \tag{10}$$

by (9) and axioms (A8) and (A11) and theorems (T9)(T10).

$$\text{(CBWB)} \wedge \text{(CBNB)} \wedge \text{(CBSR)} \rightarrow \mathcal{B}(\mathcal{B}\alpha_2 \wedge w_j \rightarrow \bigcirc d_i)$$
$$\text{for } i = 1, \ldots, 3, j = 1, \ldots, 3, \ i \neq j, \tag{11}$$

by (10), (1) and axiom (A11).

$$\text{(CBWB)} \wedge \text{(CBNB)} \wedge \text{(CBSR)} \wedge \text{(CBHI)}$$
$$\rightarrow \mathcal{B}\left(\mathcal{B}\alpha_2 \rightarrow \bigcirc\left(\bigvee_{i=1,\ldots,3} w_i \rightarrow \bigvee_{i=1,\ldots,3} d_i\right)\right), \tag{12}$$

by (11).

$$\text{(CBWB)} \wedge \text{(CBNB)} \wedge \text{(CBSR)} \wedge \text{(CBHI)}$$
$$\rightarrow \mathcal{B}\mathcal{B}\alpha_2 \rightarrow \mathcal{B}\bigcirc\left(\bigvee_{i=1,\ldots,3} w_i \rightarrow \bigvee d_i\right), \tag{13}$$

by (12) and axiom (A10).

$$\text{(CBWB)} \wedge \text{(CBNB)} \wedge \text{(CBSR)} \wedge \text{(CBHI)}$$
$$\rightarrow \mathcal{B}\alpha_2 \rightarrow \mathcal{B}\bigcirc\left(\bigwedge_{i=1,\ldots,3} \neg d_i \rightarrow \alpha_3\right), \tag{14}$$

by (13) and (T10).

$$\text{(CBWB)} \wedge \text{(CBNB)} \wedge \text{(CBSR)} \wedge \text{(CBHI)}$$
$$\rightarrow \mathcal{B}\alpha_2 \rightarrow \bigcirc\left(\mathcal{B}\bigwedge_{i=1,\ldots,3} \neg d_i \rightarrow \mathcal{B}\alpha_3\right), \tag{15}$$

by (14) and axioms (A19) and (A10).

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \mathscr{B}\alpha_2 \rightarrow \mathscr{B}\bigcirc\left(\mathscr{B} \bigwedge_{i=1,\ldots,3} \neg d_i \rightarrow \mathscr{B}\alpha_3\right),\end{aligned} \tag{16}$$

by (15), (T9), (A10) and since all the hypotheses begin by $\mathscr{B}$.

$$(\mathrm{CBSR}) \rightarrow \mathscr{B}\bigcirc\left(\mathscr{B}\alpha_3 \rightarrow \bigwedge_{i,\ldots,3} \bigcirc d_i\right), \tag{17}$$

by axiom (A11) and since (CBSR) begins by $\mathscr{B}\Box$.

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \mathscr{B}\alpha_2 \rightarrow \mathscr{B}\bigcirc\left(\mathscr{B} \bigwedge_{i=1,\ldots,3} \neg d_i \rightarrow \bigwedge_{i=1,\ldots,3} \bigcirc d_i\right),\end{aligned} \tag{18}$$

by (17) and (16).

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \mathscr{B}\alpha_2 \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\bigcirc d_i,\end{aligned} \tag{19}$$

by (18) and axioms (A19), (A10) and (T10).

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \mathscr{B}\bigcirc\left(\mathscr{B}\alpha_2 \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\bigcirc d_i\right),\end{aligned} \tag{20}$$

by (19) and since all the hypotheses begin by $\mathscr{B}\Box$. This is the end of the second part of our proof.

Now putting together (8) and (20):

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \mathscr{B}\bigcirc\left(\bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\bigcirc d_i\right)\end{aligned} \tag{21}$$

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \bigcirc\mathscr{B} \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigcirc\mathscr{B}\bigcirc\left( \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigwedge_{i=1,\ldots,3} \mathscr{B}\bigcirc d_i\right),\end{aligned} \tag{22}$$

by (21) and axioms (A19) and (A10).

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigcirc\bigcirc\left( \bigwedge_{i=1,\ldots,3} \mathscr{B}\neg d_i \rightarrow \bigwedge_{i=1,\ldots,3} \bigcirc\mathscr{B} d_i\right),\end{aligned} \tag{23}$$

by (22) and axioms (A19), (A10) and (T9), (t10).

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI})\\ &\quad \rightarrow \bigcirc \bigwedge_{i=1,\ldots,3} \mathcal{B}\neg d_i \rightarrow \bigcirc\bigcirc\Bigg(\bigwedge_{i=1,\ldots,3} \mathcal{B}\neg d_i \leftarrow \bigwedge_{i=1,\ldots,3} \bigcirc\mathcal{B} d_i\Bigg),\end{aligned} \tag{23}$$

by (22) and axioms (A19), (A10) and theorems (T9), (T10).

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI}) \wedge (\mathrm{ND1})\\ &\quad \rightarrow \bigwedge_{i=1,\ldots,3} \bigcirc\neg d_i \rightarrow \bigwedge_{i=1,\ldots,3} \bigcirc\bigcirc\neg d_i \rightarrow \bigwedge_{i=1,\ldots,3} \bigcirc\bigcirc\bigcirc d_i,\end{aligned} \tag{24}$$

by (23). The proof may now be successfully completed; we get

$$\begin{aligned}&(\mathrm{CBWB}) \wedge (\mathrm{CBNB}) \wedge (\mathrm{CBSR}) \wedge (\mathrm{CBHI}) \wedge (\mathrm{ND1})\\ &\quad \rightarrow \bigvee_{i=1,\ldots,3} \bigcirc d_i \vee \bigvee_{i=1,\ldots,3} \bigcirc\bigcirc d_i \vee \bigwedge_{i=1,\ldots,3} \bigcirc\bigcirc\bigcirc d_i.\end{aligned} \tag{25}$$

## References

[1] J. Barwise, Scenes and other situations, *J. Philos.* **LXXVIII** (1981) 368-397.

[2] R. Fagin and J.Y. Halpern, Belief, awareness, and limited reasoning, in: *Proc. 9th Internat. Joint Conf. on Artificial Intelligence*, Los Angeles, CA (1985) 491-501.

[3] R. Fagin and M.Y. Vardi, An internal semantics for modal logic, in: *Proc. 17th ACM Symp. on Theory of Computing*, Providence, RI (1985) 305-315.

[4] D. Gabbay, A. Pnueli, S. Shelah and J. Stavi, On the temporal analysis of fairness, in: *Conf. Record 7th Ann. ACM Symp. on Principles of Programming Languages*, Las Vegas, NV (1980) 174-183.

[5] J. Hintikka, *Knowledge and Belief, an Introduction to the Logic of the Two Notions* (Cornell Univ. Press, Ithaca/London, 1962).

[6] J.Y. Halpern and Y. Moses, Knowledge and common knowledge in distributed environment, in: *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, Vancouver, BC (1984) 50-61.

[7] J.Y. Halpern and Y. Moses, Towards a theory of knowledge and ignorance, Techn. Rep. IBM, RJ 4448, 1984.

[8] J.Y. Halpern and Y. Moses, A guide to the modal logic of knowledge and belief in: *Proc. 9th Internat. Joint Conf. on Artificial Intelligence*, Los Angeles, CA (1985) 480-490.

[9] D.J. Lehmann, Knowledge, common knowledge, and related puzzles, in: *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, Vancouver, BC (1984) 62-67.

[10] H.J. Levesque, A logic of implicit and explicit belief in: *Proc. Nat. Conf. on Artificial Intelligence*, Austin, TX (1984) 198-202.

[11] D.K. Lewis, *Convention, a Philosophical Study* (Harvard University Press, Cambridge, MA, 1969).

[12] D. Lehman and S. Shelah, Reasoning with time and chance, *Inform. and Control* **53** (1982) 165-198.