Systems for Knowledge and Belief

WIEBE VAN DER HOEK, *Free University Amsterdam, Department of Mathematics and Computer Science, De Boelelaan 1081, NL-1081 HV Amsterdam, The Netherlands. E-mail: wiebe@cs.vu.nl*

Abstract

We investigate modal systems for knowledge and belief, taking as a starting point a logic that was originally introduced by Kraus and Lehmann. We derive several properties and discuss (their) consequences for the epistemic operators. Kraus and Lehmann observed that adding the axiom $B_i\varphi \rightarrow B_iK_i\varphi$ to the system gives a collapse of knowledge and belief: $(K_i\varphi \leftrightarrow B_i\varphi)$. We investigate the cause(s) of this problem and suggest a 'similar' system that does allow the same axiom without the mentioned collapse. We consider as the main benefit of this paper, however, the techniques that are developed to come to this solution. It appears that applying basic correspondence theory to a multi-modal system allows a systematic examination of possible combinations of epistemic operators.

Keywords: Combined epistemic and doxastic logic, positive and negative introspection and extraspection, multi-modal logic, correspondence theory.

1 Introduction

We discuss (multi modal) logics for both knowledge and belief which are to be interpreted on Kripke structures. The basic system for our discussion is introduced by Kraus and Lehmann [19]. Knowledge (K) and belief (B) are both interpreted (as necessity operators with respect to two binary relations) over Kripke structures. (For an introduction to modal logic, see [4] or [18].) We denote their basic system with KB_{CD}. What is interesting in KB_{CD} is, that it does not only give notions of knowledge and belief (which on their own are rather familiar ones—see [9] or [23, 24]), but also some interaction properties between the two (an alternative approach to have both notions in one system is to define one in terms of the other—cf. [21] or [27]).

In the literature of philosophical logic, systems for knowledge and belief were studied in the 1960s (cf. [10]). In the 1980s, these notions became one of the central themes in the field of AI [9] and are thus gaining their place in the field of computer science [23, 24]. It now seems conventional to take the system *S5* for knowledge and *weak S5* (or *KD45*) for belief (cf. [9, 10, 23, 24]). To be more precise, it is customary to ascribe the following properties to belief. One does *not believe false assertions* ($\neg B\bot$), believers have *positive-* ($B\varphi \rightarrow BB\varphi$) as well as *negative introspection* ($\neg B\varphi \rightarrow B\neg B\varphi$). Knowledge should moreover also be *veridical:* ($K\varphi \rightarrow \varphi$). We will give a rather systematic classification of properties like these in Section 6. To mould the notions into a logical system, one usually adds the inference rules modus ponens ($\vdash \varphi, \vdash \varphi \rightarrow \psi \Rightarrow \vdash \psi$) and necessitation for K as well as B ($\vdash \varphi \Rightarrow \vdash K\varphi, \vdash \varphi \Rightarrow \vdash B\varphi$).

Once we have given the basic system KB_{CD} for knowledge and belief, we investigate some of its properties. One important theme in this paper is a problem that is also mentioned in [19]. It appears that adding the axiom $(B\varphi \rightarrow BK\varphi)$ to KB_{CD} yields $(K\varphi \leftrightarrow B\varphi)$, which is clearly undesirable. We will develop some techniques (in Section 4) to study this problem systematically, and suggest some solutions. These techniques are presented in a slightly more general setting than needed for this problem only, but the generalizations are obtained in a very natural way.

In Section 2 we introduce KB_{CD} and show the system in action by deriving some theorems. We will briefly discuss the impact of some of the properties of KB_{CD} on the notions of knowledge and belief. In Section 3 we give a Kripke semantics for a 'finitary part' of the logic, and prove completeness: in particular, we will construct a *canonical* model; this construction will be used throughout the paper to obtain completeness results for modified systems as well. In Section 4 we systematically investigate the impact of particular axioms on the canonical models (for those axioms). It will appear, that these correspondences are not hard to prove, but are, at the same time, easily transferable to more specific cases. It gives us an alternative way to derive KB_{CD}-theorems, and also enables us to prove that some formulas are not theorems.

In Section 5 we discuss the problem that we mentioned above: adding $(B\varphi \rightarrow BK\varphi)$ to KB_{CD} yields $(K\varphi \leftrightarrow B\varphi)$. In Section 6, properties like positive and negative introspection (and 'extraspection') are introduced. From Section 4 we know how these properties are related with the Kripke structure, so that we can investigate which properties KB_{CD} does not have. It will turn out that KB_{CD} is 'saturated' with respect to introspection and extraspection properties: adding any of them to KB_{CD} yields $(K\varphi \leftrightarrow B\varphi)$. We show how one can define systems for knowledge and belief with various degrees of introspection, without having $(K\varphi \leftrightarrow B\varphi)$. In Section 7 we give some conclusions.

2 KB_{CD} as a basis for knowledge and belief

Kraus and Lehmann [19] introduced a system (which we will denote with KB_{CD}) that can deal with knowledge and belief simultaneously. They used $2n$ operators $K_1, \ldots, K_n, B_1, \ldots, B_n$, modelling the knowledge and belief of n agents from an index set $People = \{1, \ldots, n\}$. In general, given a set P of propositional atoms and O of operators, a language $\mathsf{L}(\mathsf{P}, \mathsf{O})$ is the smallest set $S \supseteq \mathsf{P}$ which is closed both under infix attachment of $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$, and prefix placing of $\neg$ and operators $O \in \mathsf{O}$. For the moment, our language $\mathsf{KB_{CD}}$ for KB_{CD} is $\mathsf{L}(\mathsf{P}, \mathsf{O})$, where P is a set of atoms and $\mathsf{O} = \{C, D, E, F, K_i, B_i | i \leq n\}$. If $|\mathsf{O}| > 1$, we say to have a multi modal logic. In the sequel, if we write K_i or B_i, i is a member of $People$. The system KB_{CD} has four levels, the first of which is a propositional one:

(A0) Any axiomatization of the propositional calculus
(R0) $\vdash \varphi, \vdash \varphi \rightarrow \psi \Rightarrow \vdash \psi$.

Next, there is a level concerning properties of knowledge (K_i) and common knowledge (C). $E\varphi$ (everybody knows that φ) is defined as follows: $E\varphi \equiv K_1\varphi \wedge K_2\varphi \wedge \ldots \wedge K_n\varphi$. $C\varphi$ is supposed to mean $(E\varphi \wedge EE\varphi \wedge \ldots)$. Somewhat surprisingly, this infinite conjunction can be axiomatized.

(A1) $K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$
(A2) $K_i\varphi \rightarrow \varphi$
(A3) $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$
(A4) $C(\varphi \rightarrow \psi) \rightarrow (C\varphi \rightarrow C\psi)$
(A5) $C\varphi \rightarrow E\varphi$
(A6) $C\varphi \rightarrow EC\varphi$
(A7) $C(\varphi \rightarrow E\varphi) \rightarrow (\varphi \rightarrow C\varphi)$
(R1) $\vdash \varphi \Rightarrow \vdash C\varphi$.

Then, a level concerning general properties about belief (B_i), and common belief (D). $F\varphi$

(everybody believes that φ) is defined as follows: $F\varphi \equiv B_1\varphi \wedge B_2\varphi \wedge \ldots \wedge B_n\varphi$. $D\varphi$ is supposed to be the infinite conjunction $(F\varphi \wedge FF\varphi \wedge \ldots)$.

(A8) $B_i(\varphi \to \psi) \to (B_i\varphi \to B_i\psi)$
(A9) $\neg B_i \mathit{false}$
(A10) $D(\varphi \to \psi) \to (D\varphi \to D\psi)$
(A11) $D\varphi \to F\varphi$
(A12) $D\varphi \to FD\varphi$
(A13) $D(\varphi \to F\varphi) \to (F\varphi \to D\varphi)$.

Finally, there is a level combining (common) knowledge and (common) belief:

(A14) $K_i\varphi \to B_i\varphi$
(A15) $B_i\varphi \to K_iB_i\varphi$
(A16) $C\varphi \to D\varphi$.

LEMMA 2.1
Let $[^{\alpha}/_{\beta}]\varphi$ be any formula, which arises from φ by substituting any occurrence(s) of β in φ by α. Then the following rule of substitution Sub is derivable in KB_{CD}

$$\text{Sub} \quad \vdash \alpha \leftrightarrow \beta \Rightarrow \vdash \varphi \leftrightarrow [^{\alpha}/_{\beta}]\varphi.$$

PROOF. Here, we omit the simple, but tedious proof by induction on the complexity of φ, which should be preceded by an inductive definition of substitution. ∎

The following theorem shows that the notions of knowledge and belief, as defined in KB_{CD} (and considered separately) have at least the properties of those in $S5$ and *weak* $S5$, respectively (cf. the introduction, or [9, 23, 24]).

LEMMA 2.2
In the system KB_{CD}, knowledge (K_i) has all the properties of $S5$ whereas belief (B_i) has those of *weak* $S5$.

PROOF. Modus ponens is immediate from R0. Also, R1, $(\vdash \varphi \Rightarrow \vdash C\varphi)$, together with A5 $(\vdash C\varphi \to E\varphi)$ and the definition of E $(\equiv K_1\varphi \wedge \ldots \wedge K_n\varphi)$ gives necessitation for K_i $(\vdash \varphi \Rightarrow \vdash K_i\varphi)$. Axiom A14 $(K_i\varphi \to B_i\varphi)$ then yields necessitation for B_i as well. Veridicality and negative introspection are explicitly added for K_i to KB_{CD} (A2 and A3, respectively). Positive introspection for K_i follows from A2 and A3 ($K_i\varphi \Rightarrow_{A2} \neg K_i\neg K_i\varphi \Rightarrow_{A3} K_i\neg K_i\neg K_i\varphi \Rightarrow_{A3} K_iK_i\varphi$). Thus we have that knowledge in KB_{CD} is '$S5$-like'. Concerning belief, to show that this is '*weak* $S5$-like', since we have A9, we only have to derive the two introspection properties for B_i. Positive introspection follows immediately from A15 $(B_i\varphi \to K_iB_i\varphi)$ and A14 $(K_i\psi \to B_i\psi)$. Finally we prove negative introspection for B_i:

1.	$B_i\varphi \leftrightarrow K_iB_i\varphi$	'$\leftarrow$': A2, '$\rightarrow$': A15
2.	$\neg B_i\varphi \leftrightarrow \neg K_iB_i\varphi$	A0, 1
3.	$\neg K_iB_i\varphi \to K_i\neg K_iB_i\varphi$	A3
4.	$\neg B_i\varphi \to K_i\neg K_iB_i\varphi$	A0, 2, 3
5.	$\neg B_i\varphi \to K_i\neg B_i\varphi$	Sub (subst of $B_i\varphi$ for $K_iB_i\varphi$ (1) in 4)
6.	$K_i\neg B_i\varphi \to B_i\neg B_i\varphi$	A14
7.	$\neg B_i\varphi \to B_i\neg B_i\varphi$	A0, 5, 6

∎

DEFINITION 2.3
We say that an operator $\Box$ is a *(normal) modal operator (in L)* if it satisfies:

(i) $\vdash \varphi \Rightarrow \vdash \Box\varphi$ necessitation
(ii) $\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$ distribution

Moreover, we call a modal logic L *normal* if it contains A0, R0, necessitation and distribution.

LEMMA 2.4
The operators K_i, B_i, C, D, E and F are all normal modal operators in KB_{CD}.

REMARK 2.5
The observation above immediately follows from the definition of E and F and the axioms of KB_{CD}. This implies that we may apply our modal intuitions to derive several properties of our operators. To mention some, we have $\vdash \varphi \to \psi \Rightarrow \vdash \Box\varphi \to \Box\psi$ (i), $\vdash \Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$ (ii) and $\vdash (\Box\varphi \vee \Box\psi) \to \Box(\varphi \vee \psi)$ (iii). When we want to use such properties for $\Box$ (e.g. when deriving some KB_{CD}-theorems (2.8)), we refer to them as 2.5. These properties naturally provide some attributes for the epistemic operators they are supposed to model; for a discussion we refer to [13].

THEOREM 2.6
In [19], it is claimed (not proven) that KB_{CD} has the following theorems:

(T1)	$K_i\neg\varphi \to \neg B_i\varphi$	(T8)	$B_i(B_i\varphi \to \varphi)$
(T2)	$B_i\varphi \leftrightarrow K_iB_i\varphi$	(T9)	$D\varphi \leftrightarrow DF\varphi$
(T3)	$\neg B_i\varphi \leftrightarrow K_i\neg B_i\varphi$	(T10)	$D\varphi \leftrightarrow FD\varphi$
(T4)	$K_i\varphi \leftrightarrow B_iK_i\varphi$	(T11)	$FD\varphi \leftrightarrow DFD\varphi$
(T5)	$\neg K_i\varphi \leftrightarrow B_i\neg K_i\varphi$	(T12)	$D\varphi \leftrightarrow DD\varphi$
(T6)	$B_i\varphi \leftrightarrow B_iB_i\varphi$	(T13)	$C(\varphi \wedge \psi) \leftrightarrow C\varphi \wedge C\psi$
(T7)	$\neg B_i\varphi \leftrightarrow B_i\neg B_i\varphi$	(T14)	$D(\varphi \wedge \psi) \leftrightarrow D\varphi \wedge D\psi$

REMARK 2.7
Where in this logic, knowledge and belief are defined as separate entities with some interaction (A14-A16) axioms, an alternative approach is to take one of the two as basic, and connect the two in one fundamental definition. A popular direction follows the slogan 'knowledge = justified, true belief' (already advocated in the 1960s by e.g. [21]), but an opposite view is taken in [27], where belief (or rather $B(\varphi, \varphi_{ass})$, the belief in φ relative to some 'unusuality assertion' φ_{ass}) is defined in terms of knowledge. In [27] it is shown that, when $S5$ is taken for knowledge, the $KD45$-properties for belief follows from their fundamental definition! The same even holds for the interaction axioms A14 and A15 of KB_{CD} and the theorems T1–T7 (T8 can be shown to be also valid in their approach). However, from their proofs it follows that when the B-operator occurs more then once in a theorem, it is assumed that all the unusuallity assertions are the same. For example, one can derive in their system $B(B(\varphi, \varphi_{ass}), \varphi_{ass}) \leftrightarrow B(\varphi, \varphi_{ass})$ (cf. T8), but not $(B(B\varphi, \varphi_{ass}), (B\varphi)_{ass})) \leftrightarrow B(\varphi, \varphi_{ass})$. More generally, as is also stated in [27], it is not always clear which choice should be made for φ_{ass}.

To see our system KB_{CD} in action, we provide a derivation for T8. For a proof of the other theorems of 2.6, we refer to [13].

PROPOSITION 2.8
The following proves T8:

PROOF.

1.	$B_i\varphi \leftrightarrow K_iB_i\varphi$	A2, A15
2.	$\neg K_iB_i\varphi \rightarrow K_i\neg K_iB_i\varphi$	A3
3.	$\neg B_i\varphi \rightarrow K_i\neg K_iB_i\varphi$	A0, 1, 2
4.	$\neg B_i\varphi \rightarrow K_i\neg B_i\varphi$	Sub $(\neg B_i\varphi)$ for $\neg K_iB_i\varphi$ (1) in 3)
5.	$\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$	A14, 4
6.	$(\neg B_i\varphi \vee B_i\varphi) \rightarrow (B_i\neg B_i\varphi \vee B_i\varphi)$	A0, 5
7.	$(B_i\neg B_i\varphi \vee B_i\varphi) \rightarrow B_i(\neg B_i\varphi \vee \varphi)$	2.5
8.	$B_i(\neg B_i\varphi \vee \varphi) \rightarrow B_i(B_i\varphi \rightarrow \varphi)$	A0, 2.5
9.	$(\neg B_i\varphi \vee B_i\varphi) \rightarrow B_i(B_i\varphi \rightarrow \varphi)$	A0, 6, 7, 8
10.	$B_i(B_i\varphi \rightarrow \varphi)$	9, A0, R0

∎

Note how first negative introspection for B_i is derived (5), which then immediately (using only propositional logic and modal observations for B_i) yields the result. We will later also argue semantically (as a consequence of 4.5), that 10 follows directly from 5.

We mentioned already in the introduction that one typical property that distinguishes knowledge from belief is that knowledge is *veridical*, i.e. known facts are true. Although this property does not hold for belief, T8 expresses that agent i *believes* that it *does* hold; $B_i(B_i\varphi \rightarrow \varphi)$. Note that T8 implies that, by definition of F, we also have $\vdash B_i(F\varphi \rightarrow \varphi)$. Since this is true for arbitrary $i \in People$, we have

$$\vdash F(F\varphi \rightarrow \varphi) \quad \text{(T)}$$

expressing that everybody believes that 'the belief of everybody' is also veridical. In the system KB_{CD}, knowledge is stronger than belief, which is expressed by A14, $K_i\varphi \rightarrow B_i\varphi$. A14 seems perfectly reasonable[1] (but cf. also [28]). Of course, one does not want knowledge and belief to collapse, so in particular, we do not want A14′: $B_i\varphi \rightarrow K_i\varphi$. For one class of formulas, however, belief and knowledge *are* the same.

DEFINITION 2.9
A formula with occurrences of K_i or B_i is called an *epistemic formula*. The *belief set (knowledge set)* of an agent i in a system L is defined as $\{\varphi | L \vdash B_i\varphi\}(\{\varphi | L \vdash K_i\varphi\})$. A formula φ is *i-doxastic sequenced* if there are ψ, operators $X_1, \ldots X_n \in \{K_i, B_i, \neg K_i, \neg B_i\}$ and $n > 0$ such that $\varphi = X_1X_2 \ldots X_n\psi$. We will not always mention reference to agent i.

THEOREM 2.10
For any i-doxastic sequenced φ, we have:
$KB_{CD} \vdash (K_i\varphi \leftrightarrow \varphi) \wedge (\varphi \leftrightarrow B_i\varphi)$.

PROOF. Immediate from A2 and A3, combined with 2.2 and T2-T7 of 2.6. ∎

COROLLARY 2.11
For all i-doxastic sequenced φ:
$KB_{CD} \vdash \varphi \Leftrightarrow KB_{CD} \vdash K_i\varphi \Leftrightarrow KB_{CD} \vdash B_i\varphi$

Theorem 2.10 implies that in KB_{CD} i-doxastic sequenced formulas are believed by agent i iff they are known by agent i. Thus, knowledge and belief do collapse for believed facts and for

[1] However, in natural language it is common that one expresses the strongest facts one knows. If a judge says that he *believes* that p committed a crime, he implicitly says that he does not *know* it yet. However, having $K\varphi \rightarrow B\varphi$ and $B\varphi \rightarrow \neg K\varphi$ in one system is not interesting.

facts φ for which $\neg K_i\varphi$ holds. In particular, $B_iK_i\varphi \rightarrow K_iK_i\varphi$ is valid. The following theorem shows that the formulas of 2.9 can be reduced to a formula with *at most one main epistemic operator*, provided that all epistemic operators have the same subscript. It implies that KB_{CD} is 'optimally manageable': all sequences of operators and $\neg$s can be rewritten to a sequence with at most one operator. So, if KB_{CD} models 'our' knowledge and belief, in every-day-life we never need to use complicated 'epistemic phrases' like 'I believe, that I know to believe . . . '.

THEOREM 2.12
Let i be given, $1 \leq i \leq n$. Let $X, Y \in \{K_i, B_i, \neg\}$ such that $Y \neq \neg$ and let $\overline{X}$ be a sequence of X's. Let φ be any KB_{CD}- formula. Then $KB_{CD} \vdash \overline{X}Y\varphi \leftrightarrow (\neg)Y\varphi$, where the '$\neg$' is present if the number of '$\neg$' in $\overline{X}$ is odd.

PROOF. Immediate from 2.10. ∎

3 Kripke semantics for KB

In this section we introduce a semantics for KB_{CD}-like systems. Unlike the completeness proof of [19], which is based on the construction of a 'universal model' using labelled traces, we will construct a 'canonical model' for any consistent formula φ, thus applying ideas from classical modal logic (cf. [4, 18]). To emphasize that the structure of a model for a particular system heavily depends on the specific axioms of that system, we start out with a kind of 'barest' model. Moreover, since in this paper the notions of knowledge and belief (and their interactions) are our primary concern, we start out by simplifying KB_{CD}: in the sequel, we will not consider 'common knowledge' or 'common belief' any longer. This enlightens our considerations on completeness substantially (cf. 3.12, 3.15).

DEFINITION 3.1
The system KB is a logic in the language $\mathsf{KB} = \mathsf{L}(\mathsf{P}, \{K_i, B_i, E, F\})$. It consists of the axioms A0–A3, A8–A9, and A14–A15. As inference rules it has R0 and Necessitation for K. From now, we will use '$\vdash$' and '$\vdash_{KB}$' interchangeably.

LEMMA 3.2
For any $\varphi \in \mathsf{KB}, \vdash_{KB} \varphi$ iff $\vdash_{KB_{CD}} \varphi$.

As a consequence of 3.2 we know that the theorems T1–T8 are derivable in KB.

DEFINITION 3.3
A *Kripke model* $\mathcal{M}$ for a modal language L with one modal operator $\Box$ is a tuple $\langle W, R, \pi\rangle$, where W is a *set of worlds*, $R \subseteq W \times W$ a *binary relation*, and $\pi : W \rightarrow P \rightarrow \{\mathbf{true}, \mathbf{false}\}$ a *truth assignment* to the propositional atoms for each world $w \in W$. Truth definition for $\varphi \in L$ at w, written $(\mathcal{M}, w) \models \varphi$, is:

1. $(\mathcal{M}, w) \models p$ iff $\pi(w)(p) = \mathbf{true}$
2. $(\mathcal{M}, w) \models \psi \wedge \chi$ iff $(\mathcal{M}, w) \models \psi$ and $(\mathcal{M}, w) \models \chi$
3. $(\mathcal{M}, w) \models \neg\psi$ iff not $(\mathcal{M}, w) \models \psi$
4. $(\mathcal{M}, w) \models \Box\psi$ iff for all v for which R_iwv, $(\mathcal{M}, v) \models \psi$.

We say that an operator that is defined like $\Box$ for R, is a *necessity operator for R*. For any modal operator $\Box$, we define $\underline{\Box} = \neg\Box\neg$. $\Box$ is called the *possibility operator for R*. We then say that φ is *satisfiable* if φ is true at some world w in some model $\mathcal{M}$, φ is *true in model* $\mathcal{M}$ $(\mathcal{M} \models \varphi)$ if $(\mathcal{M}, w) \models \varphi$ for all worlds of $\mathcal{M}$, and, finally, φ is *valid* $(\models \varphi)$ if it is true in all models. For any class C of models, we write $\models_C \varphi$ if φ is true in all models in C.

As is easily verified, we have ($\models \varphi \Rightarrow \models \Box\varphi$) and ($\models \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$). Since also all propositional tautologies and modus ponens are valid, this explains why Kripke structures are so suitable for interpreting modal formulas: necessitation and distribution are valid. To summarize, we have the following (K is the 'minimal' normal modal logic).

LEMMA 3.4
For all $\varphi \in \mathsf{L}, \vdash_K \varphi \Rightarrow \models \varphi$.

The proof of the converse, (which is equivalent to saying that K-consistent formulas are satisfiable (in some Kripke model)) is also a fact from the modal logic folklore. However, for future reference, we will sketch the idea of the proof (and the construction of the model). This construction is known as the *Henkin construction*, and takes full benefit of the similarity between properties of maximal consistent sets on the syntactic side (3.5) and the truth definition of formulas in a world on the model-theoretic side (3.3). We only give a short sketch here, the reader is referred to [4, 18] or [7] for further details. We start out by repeating the notion of *maximal consistent sets*.

A set Φ is maximal consistent (m.c.) in a logic L if it is: (i) consistent (in L) and (ii) for all $\varphi, \Phi \cup \{\varphi\}$ is consistent $\Rightarrow \varphi \in \Phi$. Due to a theorem of Lindenbaum (cf. [3]), such maximal consistent sets do exist for the logic KB and its variants that we discuss here. Moreover, each consistent formula φ is contained in a m.c. set. We assume familiarity with m.c. sets (cf. [4, 18]), but summarize their vital properties in Lemma 3.5. Then we proceed by giving the definition of the *canonical model* for a (normal) modal logic (3.6) and recall some of its properties in 3.7. These notions and results will be needed in the sequel.

LEMMA 3.5
Let L be any normal modal logic (cf. 2.3). Then:

1. Every L-consistent set Φ can be extended to a m.c. set Σ
2. Suppose Σ is m.c. in L. Then:
   (a) $\varphi \notin \Sigma \Leftrightarrow \neg\varphi \in \Sigma$
   (b) $(\varphi \wedge \psi) \in \Sigma \Leftrightarrow \varphi \in \Sigma$ and $\psi \in \Sigma$
   (c) $(\varphi \vee \psi) \in \Sigma \Leftrightarrow \varphi \in \Sigma$ or $\psi \in \Sigma$
   (d) $\Phi \vdash_L \varphi$ iff $\Sigma \vdash_L \varphi$ for every m.c. set $\Sigma \supseteq \Phi$.

DEFINITION 3.6
The canonical model $\mathcal{M}^c = \langle W^c, R^c, \pi^c \rangle$ for a modal logic L is defined as follows:

- $W^c = \{\Sigma | \Sigma$ is a maximal L-consistent set$\}$
- $R^c = \{(\Sigma, \Delta) |$ *for all* $\Box\varphi \in \Sigma \Rightarrow \varphi \in \Delta\}$
- $\pi^c(\Sigma)(p) = \mathbf{true}$ iff $p \in \Sigma$.

LEMMA 3.7 ([7, 18])
For all φ and m.c. sets $\Sigma \in \mathcal{M}^c$:

- $\Box\varphi \in \Sigma \Leftrightarrow \forall\Delta \in \mathcal{M}^c(R^c\Sigma\Delta \Rightarrow \varphi \in \Delta)$
- $\underline{\Box}\varphi \in \Sigma \Leftrightarrow \exists\Delta \in \mathcal{M}^c(R^c\Sigma\Delta \wedge \varphi \in \Delta)$
- $R^c\Gamma\Delta \Leftrightarrow$ for all $\varphi : (\varphi \in \Delta \Rightarrow \Box\varphi \in \Gamma)$

LEMMA 3.8 ('fundamental theorem' [4, 18])
$(\mathcal{M}, \Sigma) \models \varphi$ iff $\varphi \in \Sigma$.

PROOF. For atomic formulas, this is immediate from the definition of π^c. For conjunctions and negations it follows from 3.5.(2a) and 3.5.(2b), respectively. If $\varphi = \Box\psi$: $(\mathcal{M}, \Sigma) \models \Box\psi$ iff (by 3.3) for all Δ with $R^c\Sigma\Delta$, $(\mathcal{M}, \Delta) \models \psi$ iff (by induction) for all Δ with $R^c\Sigma\Delta$, $\psi \in \Delta$ iff (3.7) $\Box\varphi \in \Sigma$. ∎

COROLLARY 3.9
$\vdash_K \varphi$ iff $\models \varphi$.

PROOF. The 'only if' part is 3.4. For the 'if' part, suppose $\not\vdash_K \varphi$, i.e., $\neg\varphi$ is K-consistent. Then, by 3.5.1, $\{\neg\varphi\}$ is contained in a m.c. set Σ. By 3.8, $(\mathcal{M}^c, \Sigma) \models \neg\varphi$, implying $\not\models \varphi$. ∎

Now we start to rig our bare model to models for KB. Of course we have to add a number of binary relations, so that our KB-models will be tuples

$$\langle W, \pi, S_1, \ldots, S_n, S_E, T_1, \ldots, T_n, T_F \rangle,$$

where S_i is the relation for K_i, T_i for B_i, S_E for E, and T_F for F, respectively. More interestingly, we will see that the axioms of KB force special properties upon those relations (in the canonical model).

EXAMPLE 3.10
As an easy example, consider the axiom $K_i\varphi \rightarrow B_i\varphi$. In $\mathcal{M}^c$ this leads to: $T_i^c\Gamma\Delta \Leftrightarrow \{\varphi | B_i\varphi \in \Gamma\} \subseteq \Delta \Rightarrow \{\varphi | K_i\varphi \in \Gamma\} \subseteq \Delta \Leftrightarrow S_i^c\Gamma\Delta$.

DEFINITION 3.11
A *KB-model* $\mathcal{M}$ is a tuple $\langle W, \pi, S_1, \ldots, S_n, S_E, T_1, \ldots, T_n, T_F \rangle$ satisfying:

1. S_i is an equivalence relation (cf. Definition 4.2)
2. $\forall x \exists y T_i xy$
3. $T_i \subseteq S_i$
4. $\forall x, y, z \in W((S_i xy \wedge T_i yz) \Rightarrow T_i xz)$
5. $S_E = S_1 \cup \ldots \cup S_n, T_F = T_1 \cup \ldots \cup T_n$.

We denote the *class of* KB*-models* with $\mathcal{KB}$.

THEOREM 3.12
Each KB-consistent formula is satisfied in some $\mathcal{KB}$-model.

PROOF. If φ is KB-consistent, it is contained in some KB-m.c. set Γ. So it is true in $(\mathcal{M}^c, \Gamma)$. We thus only have to show that $\mathcal{M}^c$ is a model in $\mathcal{KB}$, i.e. that it satisfies 1–5 of 3.11.

1. S_i^c is an equivalence: S_i^c is reflexive, $S_i^c\Gamma\Gamma$, by definition of S_i^c and, using A2, $K_i\varphi \in \Gamma \Rightarrow \varphi \in \Gamma$. It is seen to be transitive, by an argument similar to that in the proof of item 4 of this theorem. Finally, it is symmetric: suppose $S_i^c\Gamma\Delta$, i.e. $K_i\varphi \in \Gamma \Rightarrow \varphi \in \Delta$ (*). If not $S_i^c\Delta\Gamma$, we have a ψ with $K_i\psi \in \Delta$, but $\psi \notin \Gamma$. By item 2a of 3.5 then, $\neg\psi \in \Gamma$, implying (using A2) $\neg K\psi \in \Gamma$. Axiom A3 guarantees $K_i\neg K_i\psi \in \Gamma$, so, by (*), $\neg K_i\psi \in \Delta$, which contradicts $K_i\psi \in \Delta$.
2. By A9, $(\underline{B_i}true) \in \Gamma$, so, by 3.7, for some Δ: $T_i\Gamma\Delta$.
3. This was argued in 3.10.
4. Suppose $S_i^c\Gamma\Delta$ and $T_i^c\Delta\Sigma$. Then, $B_i\varphi \in \Gamma \Rightarrow K_iB_i\varphi \in \Gamma$ (by A15), so (by definition of T_i^c), $B_i\varphi \in \Delta$ and hence (since $S_i^c\Delta\Sigma$) $\varphi \in \Sigma$. All in all, we have $T_i\Gamma\Sigma$.

5. Since $\vdash E\varphi \rightarrow K_i\varphi$, as in 3.10, we conclude $S_i^c \subseteq S_E^c$ for all $i \leq n$, and hence $S_1^c \cup \ldots \cup S_n^c \subseteq S_E^c$. Now suppose $S_E^c \not\subseteq S_1^c \cup \ldots \cup S_n^c$, then for some $\Delta : S_E^c \Gamma \Delta$ and for no $i \leq n$, we have $S_i^c \Gamma \Delta$. Then, for all $i \leq n$, there is some φ_i for which $K_i\varphi_i \in \Gamma$, but $\varphi_i \notin \Delta$. The former gives us $K_i(\varphi_1 \vee \ldots \vee \varphi_n) \in \Gamma$ for all $i \leq n$ (and hence $E(\varphi_1 \vee \ldots \vee \varphi_n) \in \Gamma$), and the latter $(\varphi_1 \vee \ldots \vee \varphi_n) \notin \Delta$ (cf. 3.5.2b). This contradicts $S_E^c \Gamma \Delta$, so $S_E^c \subseteq S_1^c \cup \ldots \cup S_n^c$. ∎

In [9], it is claimed that, if we would add the axioms for C to the $S5$-logic for knowledge, the necessity operator for C may be seen as the transitive reflexive closure of S_E, i.e., $R_C uv$ iff there is some S_E-path from u to v. From [7], where a similar operator ($\Box^*$) is studied in the area of dynamic logic, we know that the canonical model for such a system need not have this property. However, the canonical model is transferred into a *finite* model, which then is still a model of the proper kind and in which the relation that belongs to $\Box^*$ *is* the reflexive transitive closure of the relation for $\Box$. It may be shown that for KB_{CD} there are similar problems, but also that a finite canonical model (of the appropriate kind) can be obtained in this case (cf. [17]). However, for the sequel, we need the unaffected canonical model as defined in 3.6.

Note how the particular properties of the binary relations in the canonical model are guaranteed by particular axioms of our logic. For instance, A2, $K_i\varphi \rightarrow \varphi$ forces S_i to be reflexive, (3.12.1) and the definition of E guarantees that E may be understood as the necessity operator for the union of the operators S_i for K_i. We emphasize that although $K_i\varphi \rightarrow \varphi$ is true on all S_i-reflexive models, the converse is not true: let $\mathcal{M}$ consist of two worlds u and v, with $S_i = \{(u,v),(v,u)\}$ and $\pi(u) = \pi(v)$. Then, $\mathcal{M}$ is not reflexive and still $\mathcal{M} \models K_i\varphi \rightarrow \varphi$, because of a particular property of a particular π. To abstract from the actual assignment π, the notion of *frame* is introduced, on which the interaction between axioms and properties on the binary relation can be studied more clearly.

DEFINITION 3.13
A *frame* $\mathcal{F}$ is a Kripke model without valuation

$$\pi : \mathcal{F} = \langle W, S_1, \ldots, S_n, S_E, T_1, \ldots, T_n, T_F \rangle$$

(in shorthand, $\mathcal{F} = \langle W, S_i, T_i \rangle$). We write $\mathcal{F} \models \varphi$ iff for all $\pi, \langle \mathcal{F}, \pi \rangle \models \varphi$. If Φ is any (first-order) property of $\mathcal{F}$, we say that multi modal formula φ (which is generally understood to be a *schema*) *corresponds* with Φ, if $\mathcal{F} \models \varphi \Leftrightarrow \mathcal{F}$ satisfies Φ. We then write $\varphi \sim_{co} \Phi$. If this is only true for frames $\mathcal{F}$ in some class $\mathcal{D}$ of frames, we say that we have *relative correspondence* $(\varphi \sim_{co(D)} \Phi)$. For an introduction to this topic, we refer to [2]. We denote the class of models based on $\mathcal{F}$ by $\mathcal{M}(\mathcal{F})$. A given model $\mathcal{M}$ is understood to be based on its underlying frame $\mathcal{F}_{\mathcal{M}}$: the underlying frame of the canonical model is called the *canonical frame*. Finally, we say that a logic L is sound and complete with respect to $\mathcal{D}$, or $(L \vdash \varphi \Leftrightarrow \models_{\mathcal{D}} \varphi)$ if for all $\mathcal{F}$ in $\mathcal{D}$, $L \vdash \varphi \Leftrightarrow \mathcal{F} \models \varphi$ (we then say that $\mathcal{F}$ is a frame for L).

DEFINITION 3.14
Let M_L be some multi modal language for a normal modal logic L. We say that (the scheme) $\varphi \in \mathsf{M}_\mathsf{L}$ is *canonical* (can(φ)) if the canonical frame for L satisfies φ.

As is known, (and as will be a consequence of the following section), on the level of frames, A2 *does* correspond to reflexivity. From 3.12.1 we know that S_i^c in the canonical model for KB is reflexive (forced by A2), and thus the canonical frame is. Since A2 $\sim_{co}$ reflexivity, we conclude that A2 is canonical. We stress that in general, the fact that an axiom A corresponds to property Φ is not equivalent to saying that A is canonical[2]. We know that A5 ∧ A6 ∧ A7 corresponds to

[2] I thank one of the anonymous referees for pointing out that I did not distinguish between these two notions properly in an earlier version of this paper

the property that S_C (the relation for which C is the necessity operator) is the reflexive transitive closure of S_E, whereas the canonical model for KB need not have this property at all (cf. [7, 17]). Conversely, it may be that the canonical frame has some property Φ that is 'coincidental', i.e. that is not forced by any axiom. As an example, we saw that (A2 $\wedge$ A3) (φ) forces S_i^c to be an equivalence relation. If $n = 1$, since $\Gamma' = \{K_1 p, p\}$ and $\Delta' = \{K_1 \neg p, \neg p\}$ are both consistent sets, they give rise to worlds Γ and Δ which are not S^c-accessible from each other. In other words, in the canonical frame $\exists x \exists y(\neg Sxy \wedge \neg Syx)(\Phi)$ is true, although this property does not correspond to any modal formula (If $\mathcal{F} = \langle \{w\}, \{(w,w)\} \rangle$, then $\mathcal{F} \models \varphi$, but $\mathcal{F} \not\models \Phi$). It will appear that all the multi-modal schemes φ in which we are interested here, *are* canonical.

REMARK 3.15
The fact that no modal formula φ corresponds to a given $\neg\Phi$ is sometimes exploited to make a shift from a class of models $\mathcal{C}$ for which some logic L is a complete axiomatization, to the class of models in $\mathcal{C}$ *that do satisfy* Φ (and for which L is still a complete axiomatization)! For instance, on $S5$-frames, $\neg\forall x \forall y(Sxy \wedge Syx)$ does not (relatively) correspond to any φ; however, a move from the canonical model for $S5$ to *generated* models gives models for which $\forall x \forall y(Sxy \wedge Syx)$ holds (cf. [23, 24]). Similarly, adding $(K_1\psi \vee \ldots \vee K_n\psi) \rightarrow D\psi$ (where the operator D denotes 'distributed knowledge' cf. [9] or [14] in which this operator was represented with I) to our logic KB would not give immediately a canonical model for which $S_1 \cap \ldots \cap S_n = S_D$ holds (cf. [14]), where S_D is the accessibility relation for which D is the necessity operator. Now, the fact that $S_1 \cap \ldots \cap S_n \neq S_I (= \neg\Phi)$ is not multi modally definable may be used to knead this canonical model into a model for which Φ is true, so that completeness of $KB \cup \{(K_1\psi \vee \ldots \vee K_n\psi) \rightarrow D\psi\}$ with respect to Φ-models can be obtained (cf. [14]).

REMARK 3.16
A typical question we want to address using this machinery is the following. Suppose we have some epistemic logic KB^* and we want to know whether adding one of our favourite properties for knowledge and belief implies having to accept another, perhaps less preferable property, i.e. we ask whether

$$(*) \quad KB^* \cup \{\varphi_1\} \vdash \varphi_2.$$

The answer is positive, if, for example, we can show that can(φ_1), and find Φ_1 and Φ_2 such that $\varphi_1 \sim_{co} \Phi_1, \varphi_2 \sim_{co} \Phi_2$, and $\Phi_1 \Rightarrow \Phi_2$. It is negative if we can find Φ_h with $\varphi_h \sim_{co} \Phi_h$ ($h = 1, 2$) and a KB^*-frame for Φ_1 that does not satisfy Φ_2, a question about first-order properties on Kripke frames. (Note that the seemingly semantical question whether the canonical model for $KB^* \cup \{\varphi_1\}$ is a model for φ_2 has a syntactical back bone: the answer is no iff $\neg\varphi_2$ is true at some world in $\mathcal{M}^c$ iff (by the fundamental theorem) $\neg\varphi_2$ is consistent in $KB^* \cup \{\varphi_1\}$.)

We end this section by mentioning an alternative semantics for our notions of knowledge and belief. In [8], Halpern shows that *probabilistic Kripke models* are also suitable to interpret $S5$-like knowledge or $KD45$-like belief. Such a model (for simplicity, we assume to have only one agent) $\mathcal{N}$ is of the form $\mathcal{N} = \langle W, P, \pi \rangle$, where W is a finite or countably infinite set (of, again, worlds) and P: $\mathcal{P}(W) \rightarrow$ [0,1] is a *discrete probability function*. In particular, $\Box\varphi$ is true at w iff $P(\{v | (\mathcal{N}, v) \models \varphi\}) = 1$ (cf. also [12, 22]). It appears that, when no additional constraints are made upon P, the logic for '$\Box$' is just $KD45$, so that, in that case, 'belief' is the same as 'certainty'. If we want that 'certainty' is 'knowledge', we have to import the property $\Box\varphi \rightarrow \varphi$, which is valid if we additionally assume that P satisfies $\forall w P(w) > 0$. In other words, $B\varphi \wedge \neg\varphi$ is satisfiable in world w, iff the measure of w equals 0 (and w is not taken into account when verifying $B\varphi$ at w). The techniques that we develop in the following section to characterize

several properties for knowledge and belief, are easily extended to the models of this kind (which we make clear at the end of Section 4).

4 Some correspondence results

In this section, we will prove (among other properties) that axiom A15: $B_i\varphi \rightarrow K_iB_i\varphi$ corresponds with $\forall x\forall y\forall z(S_ixy \wedge T_iyz \rightarrow T_ixz)$. Given this, it is not difficult to see that T4: $K_i\varphi \rightarrow B_iK_i\varphi$ corresponds with $\forall x\forall y\forall z(T_ixy \wedge S_iyz \rightarrow S_ixz)$, (an interchange of the K_is and B_is induces an interchange of the S_is and T_is) and also that $(K_i\varphi \rightarrow K_iK_i\varphi)$ corresponds with $\forall x\forall y\forall z(S_ixy \wedge S_iyz \rightarrow S_ixz)$, transitivity of S_i (replacing B_i by K_i induces a replacement of T_i by S_i). Obviously, inferring the last mentioned correspondence from one of the first two is easy, whereas the other way around is a much more difficult, if not impossible, task. So, for correspondence-problems, it would be nice having different operators for each occurrence in formulas like A15.

DEFINITION 4.1
We assume to have a language with sufficiently many operators $K^1, K^2, K^3, \ldots$ and equally many binary relations $R^1, R^2, R^3, \ldots$ associated to them. The K^m's ($m \in \mathbb{N}$) are just modal operators, which could be instantiated with operators from $\{K_i, B_i | i \leq n\}$.

DEFINITION 4.2
We define the following properties on binary relations R^1, R^2 and R^3, leaving universal quantification over x, y and z implicit.

(a)	seriality of R^1	$\exists yR^1xy$
(b)	reflexivity of R^1	R^1xx
(c)	transitivity of R^1 over (R^2, R^3)	$R^2xy \wedge R^3yz \Rightarrow R^1xz$
(d)	Euclidicity of R^3 over (R^2, R^1)	$R^2xy \wedge R^1xz \Rightarrow R^3yz$
(e)	weak (R^1, R^2)-density of R^3	$R^3xy \Rightarrow (\exists z(R^1xz \wedge R^2zy)$
(f)	selective transitivity of R^1 over (R^2, R^3)	$\exists y\forall z(R^2xy \wedge (R^3yz \Rightarrow R^1xz))$
(g)	R^1-postponed reflexivity of R^2	$R^1xy \Rightarrow R^2yy$
(h)	R^2-symmetry of R^1	$R^1xy \Rightarrow R^2yx$

If, for instance, we have transitivity of R over (R, R), we say that R is *transitive*. An *equivalence relation* is reflexive, transitive and symmetric.

THEOREM 4.3
Consider the following multi modal formulas (in K^1, K^2 and K^3).

(a) $\neg K^1 false$
(b) $K^1\varphi \rightarrow \varphi$
(c) $K^1\varphi \rightarrow K^2K^3\varphi$
(d) $\neg K^1\varphi \rightarrow K^2\neg K^3\varphi$
(e) $K^1K^2\varphi \rightarrow K^3\varphi$
(f) $K^1\neg K^2\varphi \rightarrow \neg K^3\varphi$
(g) $K^1(K^2\varphi \rightarrow \varphi)$
(h) $\varphi \rightarrow K^1\neg K^2\neg\varphi$

Then, for all $x \in \{a, \ldots, h\}$:

1. as a scheme, 4.3.x corresponds with 4.2.x
2. axiom 4.3.x is canonical.

PROOF. A proof for 1 is obtained by generalizing well known correspondence results for (standard) modal logic (cf. [2, 18]). In fact, both 1 and 2 follow from a theorem ascribed to Sahlqvist, but proven independently in [26] and [1] (cf. also [25]). Here we do not need that full machinery, but prove 1(d), as an example. For item 2, one needs generalizations of the construction in Section 3, from which the results for (a), (b), (c) and (g) are immediately obtained. To illustrate an existential quantified case, we prove 2(e) as a generalisation of a proof in [7].

1(d). We have to show:
$\mathcal{F} \models \neg K^1\varphi \rightarrow K^2\neg K^3\varphi \Leftrightarrow \mathcal{F} \models (R^2xy \wedge R^1xz \Rightarrow R^3yz)$.
$\Leftarrow$: Suppose for some π and $w, \langle \mathcal{F}, \pi\rangle, w \models \neg K^1\varphi$, i.e., for some v with R^1wv, $\langle \mathcal{F}, \pi\rangle, v \models \neg\varphi$. Let u be any world for which R^2wu. Then R^3uv, and hence $\langle \mathcal{F}, \pi\rangle, u \models \neg K^3\varphi$, and thus $\langle \mathcal{F}, \pi\rangle, w \models K^2\neg K^3\varphi$.
$\Rightarrow$: Suppose $\mathcal{F} \not\models R^2xy \wedge R^1xz \Rightarrow R^3yz$, i.e., there are worlds u, v, and w for which R^2wu, R^1wv, but not R^3uv. Define π such that p is false only in v. Then $\langle \mathcal{F}, \pi\rangle, w \models \neg K^1p \wedge \neg K^2\neg K^3p$, so $\mathcal{F} \not\models \neg K^1p \rightarrow K^2\neg K^3p$.

2(e). Suppose $R^3\Gamma\Delta$. We have to find a Σ in the canonical model, for which both $R^1\Gamma\Sigma$ and $R^2\Sigma\Delta$. By the definition of canonical model, and Lemma 3.5.1, it is sufficient to show that the set $\Sigma' = \{\psi | K^1\psi \in \Gamma\} \cup \{K^2\delta | \delta \in \Delta\}$ is consistent. Suppose not, then $\psi_1 \wedge \ldots \wedge \psi_m \wedge \underline{K^2}\delta_1 \wedge \ldots \wedge \underline{K^2}\delta_k \rightarrow \bot$, for some $m, k \geq 1$. This is equivalent to $\psi_1 \wedge \ldots \wedge \psi_m \rightarrow (K^2\neg\delta_1 \vee \ldots \vee K^2\neg\delta_k)$, so, using 2.5(iii), we have $\psi_1 \wedge \ldots \wedge \psi_m \rightarrow K^2(\neg\delta_1 \vee \ldots \vee \neg\delta_k)$. By 2.5(i) and (ii), we get $K^1\psi_1 \wedge \ldots \wedge K^1\psi_m \rightarrow K^1K^2(\neg\delta_1 \vee \ldots \vee \neg\delta_k)$. We now use $K^1K^2\varphi \rightarrow K^3\varphi : K^1\psi_1 \wedge \ldots \wedge K^1\psi_m \rightarrow K^3(\neg\delta_1 \vee \ldots \vee \neg\delta_k)$. By definition of the ψ_r, $K^1\psi_r \in \Gamma$ ($r \leq m$). Since $R^3\Gamma\Delta$, we have $(\neg\delta_1 \vee \ldots \vee \neg\delta_k) \in \Delta$, and (using 2c) $\neg\delta_s \in \Delta$, for some $s \leq k$, contradicting the definition of the δs. ∎

We like to stress that the proofs for these general cases ('fresh' operators for each occurrence) are not more complicated than in the standard modal case.

REMARK 4.4
Because $K^2\varphi \rightarrow \varphi$ corresponds with reflexivity of $R^2, (\forall x R^2xx)$, it is easy to see that $K^1(K^2\varphi \rightarrow \varphi)$ is valid at x if all R^1-successors y of x satisfy $K^2\varphi \rightarrow \varphi$, and so, if $\forall y(R^1xy \rightarrow R^2yy)$. This suggests a way to derive 'postponed correspondences' like 4.3(g). Suppose KB^m-formula φ corresponds *locally* with property $\phi(x)$, i.e. for all frames $\mathcal{F}$ and world $x \in \mathcal{F}, (\mathcal{F}, x) \models \varphi$ iff $(\mathcal{F}, x) \models \phi(x)$. Then, $K^i\varphi$ corresponds locally with $\forall y(R^ixy \rightarrow \phi(y))$ (and so, globally with $\forall x\forall y(R^ixy \rightarrow \phi(y))$.

One can now systematically list all the properties that the relations S_i and T_i of the frames in $\mathcal{KB}$ satisfy, by investigating the axioms involving K_i and B_i. For instance, for transitivity, we get, that from the c-part of 4.3.1, it follows that S_i and T_i are transitive (from Lemma 2.2 and T6). T_i is transitive over (S_i, T_i) (A15), S_i is 'maximally transitive': it is transitive over (S_i, S_i), over (T_i, S_i) (T4), over (S_i, T_i) (because $K_i\varphi \Rightarrow_{A2} B_i\varphi \Rightarrow_{A15} K_iB_i\varphi$: T11) and over $(T_i, T_i)(K_i\varphi \Rightarrow_{T11} K_iB_i\varphi \Rightarrow_{A14} B_iB_i\varphi$: T12). We can now do some reasoning about properties of binary relations in $\mathcal{KB}$ and translate the result to KB.

THEOREM 4.5
1. A reflexive, Euclidean relation is both symmetric and transitive.
2. If R^1 is Euclidean and reflexive, $R^2 \subseteq R^1$ and R^2 is transitive over (R^1, R^2), then R^2 is Euclidean.
3. A relation that is Euclidean, is also postponed reflexive.

PROOF. 1. Suppose Rxy (1) and Ryz (2). By reflexivity, Rxx (3). Euclidicity, (1) and (3) give Ryx (4). This proves symmetry of R. Finally, (2), (4) and Euclidicity give Rxz.

2. Suppose R^2xy and R^2xz. Then $(R^2 \subseteq R^1)R^1xy$ and (by ii) R^1yx. Since R^2 is transitive over (R^1, R^2), and R^1yx and R^2xz, we have R^2yz.

3. Rxy and Euclidicity give Ryy. ∎

Using arguments of 3.16, we find $KB \vdash \varphi \rightarrow K_i\neg K_i\neg\varphi$, combining results on correspondences and canonicalness in the following way: we have $KB \vdash K_i\varphi \rightarrow \varphi$ and $KB \vdash \neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ (A2 and A3); 4.3.2(b and d) guarantee that the canonical frame for KB also validates A2 and A3. Now we use 4.3.1(b and d) to conclude that S_i on this frame is both reflexive and Euclidean, and thus, by 4.5.1, symmetric. By 4.3.1(h), we know that the canonical frame (and hence, also the canonical model) for KB satisfies $\varphi \rightarrow K_i\neg K_i\neg\varphi$; so, using the fundamental theorem (3.8) we observe that $\varphi \rightarrow K_i\neg K_i\neg\varphi$ is contained in every KB-maximal consistent set and hence, by 3.5.2(d), $KB \vdash \varphi \rightarrow K_i\neg K_i\neg\varphi$.

Note that, in a similar way, we conclude that S_i is transitive, so that we again have a proof of positive introspection for K_i. Whereas in 2.2, we argued that A2 and A3 were sufficient to derive the same result within KB, we now semantically use Φ_2 and Φ_3 with A2 $\sim_{co} \Phi_2$ and A3 $\sim_{co} \Phi_3$ to find a Φ with $(\Phi_2 \wedge \Phi_3) \Rightarrow \Phi$ and $\Phi \sim_{co} K_i\varphi \rightarrow K_iK_i\varphi$. There is a similar correspondence between the proof of negative belief introspection in 2.2. and deriving Euclidicity for T_i from 4.5.2. Finally, note that 4.5.3 gives us T8 again: since $\forall x\forall y(T_ixy \rightarrow T_iyy)$ is derived for T_i, using 4.4 we conclude that $B_i(B_i\varphi \rightarrow \varphi)$ is derivable in KB; which we indeed showed in 2.8.

In particular, note that the T_is in the frames of $\mathcal{KB}$ are also transitive, Euclidean and dense (this follows from 4.3 together with T6 ('$\rightarrow$'), T7 ('$\rightarrow$') and T6 ('$\leftarrow$') of 2.6, respectively). In the opposite direction, one can make an *exhaustive list* of properties of 4.2 for the frames in $\mathcal{KB}$ (which immediately proves the following theorem), and use the absence of special properties in $\mathcal{KB}$ to show non-derivability in KB. For example, T_i is not transitive over (T_i, S_i) and also not over (S_i, S_i). By way of example, we prove the former. Figure 1 is a $\mathcal{KB}$-structure, in which T_i is denoted with thick, and S_i with thin arrows, respectively. Note that although T_iuv and S_ivw are true in that structure, T_iuw is not.

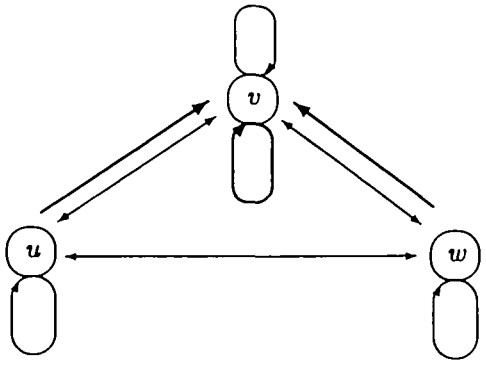

FIG. 1. A $\mathcal{KB}$-structure, in which S_i is denoted with thin, and T_i with thick arrows, respectively

From the previous paragraph, we get a lot of non-theorems of KB: in particular, because T_i is not transitive over (T_i, S_i), we have $KB \not\vdash B_i\varphi \rightarrow B_iK_i\varphi$. We give a list (writing '$\not\vdash \varphi$' instead of '$KB \not\vdash \varphi$') of non-derivable formulas that are important when studying knowledge and belief (cf. the introduction, 4.7, and, for a classification, Section 6).

THEOREM 4.6

1. $\not\vdash B_i\varphi \rightarrow B_iK_i\varphi$, $\not\vdash B_i\varphi \rightarrow K_iK_i\varphi$
2. $\not\vdash \neg K_i\varphi \rightarrow K_i\neg B_i\varphi$, $\not\vdash \neg K_i\varphi \rightarrow B_i\neg B_i\varphi$
3. $\not\vdash B_iB_i\varphi \rightarrow K_i\varphi$, $\not\vdash K_iB_i\varphi \rightarrow K_i\varphi$
4. $\not\vdash B_i\neg K_i\varphi \rightarrow \neg B_i\varphi$, $\not\vdash K_i\neg K_i\varphi \rightarrow \neg B_i\varphi$
5. $\not\vdash K_i(B_i\varphi \rightarrow \varphi)$
6. $\not\vdash \varphi \rightarrow K_i\neg B_i\neg\varphi$, $\not\vdash \varphi \rightarrow B_i\neg B_i\neg\varphi$.

REMARK 4.7
Of Theorem 4.6, 1 and 3 are of the form $XB\varphi \rightarrow YK\varphi$, with $X, Y \in \{B, K, \epsilon\}$, where ϵ is the empty (identity) operator. Properties 2 and 4 express that if φ is (believed or known to be) not known, it should also have consequences for the agent's (non-) belief about φ (they are of the form $X\neg K\varphi \rightarrow Y\neg B\varphi, X, Y \in \{B, K, \epsilon\}$). So, in KB, it is perfectly well possible (i.e. satisfiable in the system KB), that an agent (knows or believes that he) believes φ, without (knowing or believing that he is) knowing φ. The non-theorems of 4.6 neatly show some differences between knowledge and belief: 1 - 5 of 4.6 are all valid in KB if we replace each occurrence of B by K.

We end this section with the following aside. The correspondences that are obtained here, can directly be transformed to the *general probability structures* (g.p.s.) as introduced in [8] (cf. the end of Section 3). To see this, we first generalize the notion of g.p.s. A structure $\mathcal{N}$ is a $g.p.s._k$ if $\mathcal{N} = \langle W, \mathcal{P}_1, \ldots \mathcal{P}_k, \pi\rangle$, with W a (finite or countable) set of worlds, π a truth-assignment for each world and the $\mathcal{P}_i$'s families of discrete probability functions (a function $P_i(w)$ for each world w) on $W (i \leq k)$. Following [8], we define the $support_i$ relation on W as (u, v) in $support_i$ iff $\mathcal{P}_i(u)(v) > 0$. Under this definition, we can view a $g.p.s._k$ as a Kripke structure with k accessibility relations (the support relations). It is obvious that any result on modal logic has immediate implications for probability structures via this support relation. For instance, $K^1\varphi \rightarrow K^2K^3\varphi$ will be valid on those structures for which $\forall xyz(P_2(x)y > 0 \wedge P_3(y)z > 0 \Rightarrow P_1(x)z > 0)$ holds.

5 Conscious beliefs, believed consciousness

Our system KB verifies A15: $B_i\varphi \rightarrow K_iB_i\varphi$ (beliefs are 'conscious', in the sense of 'known'). This demonstrates that B_i represents a rather *explicit* belief, in the sense that the agent is *aware* of adopting them—the terms 'explicit belief' and 'implicit belief' are introduced in [22] and also used in [20]; in [5, 15] these notions are related to 'awareness'. Here, one may just associate 'implicit' with 'weak' and 'explicit' with 'strong'. Knowledge might be considered a very explicit notion of belief. If B_i would represent a notion of *implicit* belief, it seems reasonable to let $(B_i\varphi \wedge B_i\neg\varphi)$ be satisfiable simultaneously with $\neg B_i\bot$ (cf [15]; however, at this point, our use of 'implicit belief' diverges from that in [5, 20, 22], where it is assumed to be some logically closed set of beliefs—facts that *implicitly* follow from the agent's beliefs, although he need not be aware of it). But assuming (satisfiability of) $(B_i\varphi \wedge B_i\neg\varphi)$, A15 would yield $K_i(B_i\varphi \wedge B_i\neg\varphi)$. This again demonstrates that A15 is reasonable for *explicit* beliefs (in our sense); if agent i *knows* that he has inconsistent beliefs, he should retract some of them.

Kraus and Lehmann remark that it would be interesting to also have $B_i\varphi \rightarrow B_iK_i\varphi$, implying that agent i believes that his beliefs are conscious. (In Section 6, we pay some more attention to the kinds of belief these two formulas would apply to.) However, adding $B_i\varphi \rightarrow B_iK_i\varphi$ to KB would give $B_i\varphi \rightarrow K_i\varphi$. Now, we concentrate on finding KB-like systems that *do* allow $B_i\varphi \rightarrow B_iK_i\varphi$, without yielding $(B_i\varphi \leftrightarrow K_i\varphi)$. (We will say that such a system solves the B(elieved) C(onsciousness) of B(eliefs) problem.) The latter property ($\vdash K_i\varphi \leftrightarrow B_i\varphi$) corresponds with $S_i = T_i$, for which we will give a sufficient condition. Recall from Theorem 4.3 that $B_i\varphi \rightarrow B_iK_i\varphi$ corresponds with $\forall s \forall t \forall u : T_isu \wedge S_iut \Rightarrow T_ist$

THEOREM 5.1
Let S and T be two binary relations on a set W, and consider the properties:

(a) T is transitive over (T, S)

(b) T is contained in S

(c) T is serial, and

(d) S is Euclidean

Then:

1. (S and T satisfy (a)–(d)) $\Rightarrow$ (S equals T)
2. For each proper subset $A \subset \{a, b, c, d\}$, we can find relations S and T that satisfy A, but for which $S \neq T$.

PROOF. We prove 1, and refer to Fig. 2 for an example of a structure that satisfies a, b and c, but for which $S \neq T$. So suppose Sxy. Using c, we find a z for which Txz and, by b, Sxz. By d, we get Szy. Now apply a to Txz and Szy to conclude Txy. ∎

Semantically, we now know when S and T do collapse. What does this mean for knowledge and belief? From 4.3 we know that $B_i\varphi \rightarrow B_iK_i\varphi$ ($\underline{1}$) characterizes 5.1(a), that $K_i\varphi \rightarrow B_i\varphi$ ($\underline{2}$) characterizes 5.1(b), that $\neg B_i false$ ($\underline{3}$) characterizes 5.1(c) and that $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ ($\underline{4}$) characterizes 5.1(d). Now it is clear, that, if we want $B_i\varphi \rightarrow B_iK_i\varphi$ but not $B_i\varphi \rightarrow K_i\varphi$, we have to give up one of the three last properties of $\underline{2}, \underline{3}$, and $\underline{4}$ for knowledge and belief, because of the following:

THEOREM 5.2
$\models (\underline{1} \wedge \underline{2} \wedge \underline{3} \wedge \underline{4}) \rightarrow (B_i\varphi \rightarrow K_i\varphi)$.

PROOF. Apply 3.16 to 5.1. ∎

Theorem 5.2 implies that adding $B_i\varphi \rightarrow B_iK_i\varphi$ to KB does yield $B_i\varphi \equiv K_i\varphi$, because KB satisfies $\underline{2}$, $\underline{3}$ and $\underline{4}$. Together with 5.1 it also offers solutions: if one wants to have $B_i\varphi \rightarrow B_iK_i\varphi$ but not $B_i\varphi \rightarrow K_i\varphi$, one has to give up one of the properties expressed by A14, A9 or A3. Summarizing, Theorem 5.1.1 implies that, in order to add $(B_i\varphi \rightarrow B_iK_i\varphi)$ and at the same time avoiding $(B_i\varphi \equiv K_i\varphi)$, it is *necessary* to give up one of A3, A9 and A14, whereas 5.1.2 expresses that this may also be *sufficient* (whether this is indeed so, depends on the axioms we do *add* to such a system; in the sequel, we will investigate several possibilities).

Giving up A14, $K_i\varphi \rightarrow B_i\varphi$, or semantically, $T_i \subseteq S_i$, makes $(K_ip \wedge \neg B_ip)$, and even $(K_ip \wedge B_i\neg p)$ satisfiable. Then, B_i represents an implicit notion of belief—a notion that we studied in [15]—and then the whole system KB needs revision. (See also [28] for an epistemic logic in which $(K_i\varphi \rightarrow B_i\varphi)$ is not valid.) We doubt whether, for instance, A15 would be

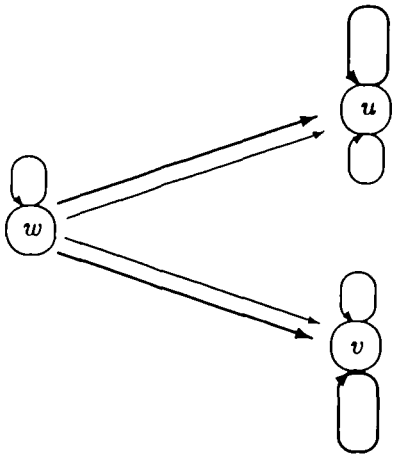

FIG. 2. A structure in which S is denoted with thin, and T with thick arrows

a desirable property for implicit belief, and probably the same holds for $B_i\varphi \rightarrow B_iK_i\varphi$, the formula that urged us to drop A14 in the first place.

One could also give up A9, but similar remarks as in the previous paragraph can be made here. For instance, from A15, we get $B_i false \rightarrow K_iB_i false$, but why should an agent hold on to false beliefs if he knows he has them? Moreover, dropping $\neg B_i false$ cannot invalidate $\neg B_i false \rightarrow (K_i\varphi \leftrightarrow B_i\varphi)$. So, either agent i beliefs in falsehood, or his beliefs equal his knowledge. Dropping A3, $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ is the third alternative. Note that a knowledge agent that satisfies A3 is very much aware of all the facts that are around: if he does not know φ, he knows that he does not. This would imply, that a Bantu tribesman knows that he does not know that personal computer prices are going down. For a discussion about 'awareness', we refer to [5], where the Bantu tribesman example is taken from, and to [15].

From Remark 2.2, we know that $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ (A3) implies $K_i\varphi \rightarrow K_iK_i\varphi$ (A3′). We could try to see what happens if we replace A3 by A3′ (a discussion on these axioms can already be found in [10]). We know that A3′ corresponds with transitivity of S_i.

DEFINITION 5.3
Let KB^- be the system consisting of all the axioms of KB, but with A3 replaced by A3′: $K_i\varphi \rightarrow K_iK_i\varphi$ and with A17 : $B_i\varphi \rightarrow B_iK_i\varphi$, added to it.

THEOREM 5.4
$KB^- \not\vdash B_i\varphi \rightarrow K_i\varphi$.

PROOF. To prove this, from arguments given in this section, it is clear that it is sufficient to find a $\mathcal{KB}^-$ model $\mathcal{M}$ in which the S_is are reflexive and transitive, the T_is are serial and transitive (not Euclidean; note that 4.5.2 cannot be applied in $\mathcal{KB}^-$), $T_i \subseteq S_i$ and in which the T_is are transitive over both (S_i, T_i) and (T_i, S_i), but at the same time $S_i \not\subseteq T_i$. Such a structure is given in Fig. 2. ∎

From the model of Fig. 2, we see that, since T_i is not Euclidean over (S_i, T_i), we also have $KB^- \not\vdash \neg B_i\varphi \rightarrow K_i\neg B_i\varphi$. We will investigate the (non-) theorems of 'KB-like systems' a

bit more systematically in the next section. Of course, it is easy to define a system that does not verify $B_i\varphi \rightarrow K_i\varphi$ but that does yield A15 and A17. However, we want a system S that is 'close(st) to $KB \cup \{A17\}$' and such that $S \nvdash B_i\varphi \leftrightarrow K_i\varphi$. For such an S, some theorems of KB must be sacrificed. For example, $B_iK_i\varphi \rightarrow K_i\varphi$ (implied by T4), with $B_i\varphi \rightarrow B_iK_i\varphi$ immediately yields $B_i\varphi \rightarrow K_i\varphi$. In order to study these problems more systematically and to get a clearer notion of 'close to KB' we will explore the fact that we now know how the properties of knowledge and belief, as expressed in the axioms and theorems T1–T10 of KB act upon the structure of its Kripke models.

6 Introspection and extraspection

Now, before we take up the BCB-problem itself, we will investigate some general properties of knowledge and belief. We will see how they are present in KB, and we show some combinations of those properties that are possible in a system that defines knowledge and belief as two necessity operators.

DEFINITION 6.1
Let X, Y and Z range over epistemic operators. Then, formulas of the form:

(a) $X\varphi \rightarrow YZ\varphi$ are called *positive introspection (p.i.-) formulas*
(b) $\neg X\varphi \rightarrow Y\neg Z\varphi$ are called *negative introspection (n.i.-) formulas*
(c) $XY\varphi \rightarrow Z\varphi$ are called *positive extraspection (p.e.-) formulas*
(d) $X\neg Y\varphi \rightarrow \neg Z\varphi$ are called *negative extraspection (n.e.-) formulas*
(e) $X(Y\varphi \rightarrow \varphi)$ are called *trust formulas.*

We will call instantiations of (a)–(d) *inspection-formulas*, and we will denote the set of all instantiations of (a)–(e) with IT. Each of the above defined notions (a)–(e) determines a subclass of IT.

Note that all the axioms and theorems that were discussed or given in Section 2 were equivalent to either an IT-formula, or of one of the forms $X\varphi \rightarrow Y\varphi$ and $X\varphi \rightarrow \varphi$.

THEOREM 6.2
In any system, if $(K_i\varphi \rightarrow B_i\varphi)$ (A14) is valid, each class of IT is partially ordered, with $\varphi \leq \psi$ iff $\varphi \Rightarrow_{A14} \psi$. For each class of IT, there is a smallest element (modulo equivalence).

PROOF. We define the notions of *positive* and *negative* occurrences of operators X in formulas. If φ does not contain X, X occurs positively in $X\varphi$. Each positive (negative) occurrence of X in φ is a positive (negative) occurrence of X in $Y\varphi$ (Y may be X, ϵ, or any other modal operator) and $\psi \rightarrow \varphi$. Each positive (negative) occurrence of X in φ is a negative (positive) occurrence of X in $\neg\varphi$ and $\varphi \rightarrow \psi$. Now we can show that $\varphi \geq \psi$ iff ψ can be obtained from φ by replacing negative occurrences of B_i in φ by B_i or K_i, and replacing positive occurrences of K_i by B_i or K_i. Instead of a proof, we give an example: in Fig. 3, '$\geq$' is the transitive reflexive closure or the relation denoted with arrows in the class NI (we do not write the subscript i; formulas in rectangles are non-theorems of KB). ∎

In the next paragraphs, we will spend some words on positive introspection, followed by a paragraph about negative introspection. The discussion can easily be extended to the other inspection properties. To start, we want to point out the difference between $(X\varphi \rightarrow YZ\varphi)$ and $Y(X\varphi \rightarrow Z\varphi)$. Note that the latter is purely a property of Y-beliefs, whereas one could interpret the former as a property noted by an observer from outside. Compare the difference between $(K_i\varphi \rightarrow K_jK_k\varphi)$ and $K_j(K_i\varphi \rightarrow K_k\varphi)$: in the latter formula, the fact that agent k

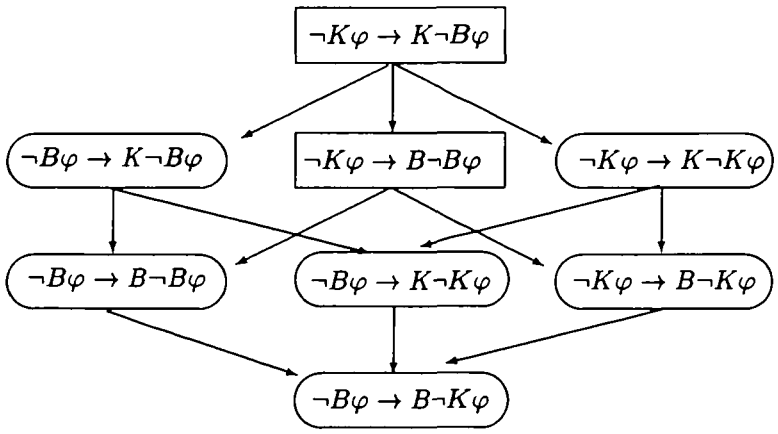

FIG. 3. '$\geq$' in the class NI

knows everything that i knows, is known by agent j (i.e. in the scope of K_j!). Even clearer is the distinction between $(B\varphi \to KB\varphi)$ and $K(B\varphi \to B\varphi)$.

The positive introspection formula $(X\varphi \to YZ\varphi)$ has, if $(K\varphi \to B\varphi)$ is valid, as a strongest instantiation $(B\varphi \to KK\varphi)$, and as its weakest $(K\varphi \to BB\varphi)$. For 'ordinary' belief and knowledge, the first is indeed too strong. The latter presumes introspection in one's own beliefs. We doubt, however, whether people use phrases like 'I believe that I believe that ...', and if so, they probably mean 'I very weakly believe that ...'. It would be interesting to have a system with two (possibly the same) notions of belief, say explicit belief (B^e) and implicit belief (B^i), such that $(B^iB^e\varphi \wedge \neg B^e\varphi)$ is satisfiable.

Because $K\varphi \to KK\varphi$ is true for most notions of knowledge, it seems reasonable to expect that $(B\varphi \to BK\varphi)$ is true for notions of belief that resemble knowledge, i.e. for strong notions of belief. We can be a bit more precise here, and ask for which X and Y, $(X\varphi \to YK\varphi)$ should be true. We might expect $(X\varphi \to YK\varphi)$ to hold for 'strong' X-belief, and 'weak' Y-belief. For instance, the choices $X \equiv$ 'I am convinced' and $Y \equiv$ 'I suspect' is a more acceptable than the other way around. (In KB, $(K\varphi \to BK\varphi)$ is valid, whereas $(B\varphi \to KK\varphi)$ is not.)

Instead of 'strong' belief, we could also write 'expensive' (having serious consequences, e.g. the belief of a judge or surgeon), and instead of 'weak' we could say 'cheap' (e.g. the belief of some gossip-paper). (The introspection property $B\varphi \to KB\varphi$ seems desirable when B denotes an 'expensive' belief; for instance, if a judge believes that females are bad car-drivers, he had better know that he believes so when he has to judge about Alice's role in an accident.) Furthermore, $(B\varphi \to BK\varphi)$ models the attitude of an agent who thinks (believes) that he is very critical in adopting beliefs: he only believes φ if he believes that he knows φ.

However, $(B\varphi \to BK\varphi)$ is not a property of *all* notions of belief. For instance, we can imagine a mathematician believing Fermat's theorem is true, without believing that he knows it is true. Moreover, $\neg(B\varphi \to BK\varphi)$ might be satisfiable in systems that interpret belief as a 'practical, working belief'. If I leave home on a bright day, I may adopt the working belief that it will not rain that day (so leave my raincoat at home), although I need not believe that I know that it will stay dry. Also, it seems that, if B is interpreted as some religious belief, $(B\varphi \to BK\varphi)$

need not hold: $(B\varphi \wedge \neg BK\varphi)$, even $(B\varphi \wedge B\neg K\varphi)$ seems perfectly consistent then.

Negative introspection formula $(\neg X\varphi \rightarrow Y\neg Z\varphi)$ has, if $(K\varphi \rightarrow B\varphi)$ is valid, $(\neg K\varphi \rightarrow K\neg B\varphi)$ as its strongest instantiation, and $(\neg B\varphi \rightarrow B\neg K\varphi)$ as its weakest. Negative introspection is closely related to the problem of 'awareness' (cf. [5, 15]). $\neg X\varphi$ could be true because the agent is not aware of φ. Now, if Y is the belief or knowledge of the same agent, and $(\neg X\varphi \rightarrow Y\neg Z\varphi)$ is true, he becomes aware of φ. Note that A3 $(\neg K\varphi \rightarrow K\neg K\varphi)$ is a strong property of knowledge: by contraposition, it implies, that the agent's ignorance of his ignorance is sufficient to have knowledge: $(\neg K\neg K\varphi \rightarrow K\varphi)$.

The following theorem says that KB is *saturated* with respect to the classes of IT (cf. Definition 6.1).

THEOREM 6.3
KB is maximal in the sense that adding any introspection, extraspection or trust formula to it makes $B_i\varphi \leftrightarrow K_i\varphi$ a theorem.

PROOF. We carry out the proof for the classes of introspection formulas; the other cases are similar. Due to the previous theorem, in each class we can find some 'weakest' formulas that are not KB-theorems yet. The weakest p.i.- formula outside KB is $B_i\varphi \rightarrow B_iK_i\varphi$ (cf. Theorem 4.6). We have seen (in Section 5) that indeed $KB \cup \{B_i\varphi \rightarrow B_iK_i\varphi\} \vdash B_i\varphi \rightarrow K_i\varphi$. Or, for the case of negative introspection, we know from Fig. 3 that $\neg K_i\varphi \rightarrow B_i\neg K_i\varphi$ is the weakest non-theorem of KB in this class. Since, by T5, $B_i\neg K_i\varphi$ is equivalent to $\neg B_i\varphi$, we immediately have $KB \cup \{\neg K_i\varphi \rightarrow B_i\neg K_i\varphi\} \vdash B_i\varphi \rightarrow K_i\varphi$. ∎

Now we have some more equipment to look at our BCB-problem again.

DEFINITION 6.4
Let KB^+ be the system $KB \setminus \{A3\}$ together with:

$A3'$	$K_i\varphi \rightarrow K_iK_i\varphi$
$A1^+$	$B_iB_i\varphi \rightarrow B_i\varphi$
$A2^+$	$B_i\neg B_i\varphi \rightarrow \neg B_i\varphi$
$A3^+$	$B_i\varphi \rightarrow B_iK_i\varphi$
$A4^+$	$\neg B_i\varphi \rightarrow K_i\neg B_i\varphi$
$A5^+$	$\neg K_i\varphi \rightarrow B_i\neg K_i\varphi$.

REMARK 6.5
Here, we will not discuss whether KB^+ models some interesting notion of belief and knowledge. Technically, we can relate KB^+ with KB in terms of the notions developed in this section. The basic idea behind KB^+ is that it solves the BCB-problem and is quite similar to KB. An important reference for this similarity is IT. In Definition 6.4 we take benefit of the nice order in each of the IT-classes. We defined KB^+ such that it has the same IT-properties as KB, with $(B_i\varphi \rightarrow B_iK_i\varphi)$ added to it, and formulas that yield $(B_i\varphi \rightarrow K_i\varphi)$ left out. For instance, for the class PE, we take care that KB^+ lacks $B_iK_i\varphi \rightarrow K_i\varphi$ (it would yield, using $A3^+$, $B_i\varphi \rightarrow K_i\varphi$), and add $A1^+$, which is similar to KB's Theorem T6. The models for KB^+ are understood by applying Theorem 4.3. Then, in the same way as in 4.6, non-theorems of KB^+ can be found.

The next theorem compares the two systems with respect to IT. In particular, item 6 of Theorem 6.6 states that we cannot make KB and KB^+ look more alike with respect to IT. Lemma 6.7 shows that *outside IT*, KB and KB^+ can still differ.

THEOREM 6.6
KB^+ satisfies the following properties:

1. $KB^+ \vdash B_i\varphi \rightarrow B_iK_i\varphi$ and $KB^+ \nvdash B_i\varphi \rightarrow K_i\varphi$.
2. For all axioms χ of KB such that $\chi \notin IT$:
   $KB^+ \vdash \chi$
3. For all axioms χ of KB^+ such that $\chi \notin IT$:
   $KB \vdash \chi$
4. For all $\chi \in IT \setminus \{B_i\varphi \rightarrow B_iK_i\varphi\}$:
   $KB^+ \vdash \chi \Rightarrow KB \vdash \chi$
5. For all $\chi \in IT \setminus \{\neg K_i\varphi \rightarrow K_i\neg K_i\varphi, B_iK_i\varphi \rightarrow K_i\varphi\}$:
   $KB \vdash \chi \Rightarrow KB^+ \vdash \chi$
6. For all $\chi \in IT, KB' \in \{KB, KB^+\}$:
   $(KB' \vdash \chi$ or $KB' \cup \{\chi\} \vdash B_i\varphi \rightarrow K_i\varphi)$

PROOF. The first part of item 1 follows by definition of KB^+, the second part can be verified by finding a model for KB^+ (the structure of this model is immediately read off from Definition 6.4, together with Theorem 4.3) for which $S_i \not\subseteq T_i$. Items 2 and 3 are true by definition of KB^+. Finally, 4, 5 and 6 are easily verified by checking them for the strongest formula χ in each class (cf. Theorem 6.2) for which the antecedent is true (in case of 5 and 6). ∎

LEMMA 6.7
$KB^+ \vdash B_i(B_i\varphi \rightarrow K_i\varphi)$, but $KB \nvdash B_i(B_i\varphi \rightarrow K_i\varphi)$.

Solving the BCB-problem boils down to investigating the possibility of having certain combinations of IT-formulas. Of course, one can do this independently from the BCB-problem and study what kind of KB-like systems are possible anyhow. For instance, one might want a modal system modelling knowledge and belief of two agents (KB_2). Then, one might assume maximal p.i. properties (in KB_2, this amounts to $K_h\varphi \rightarrow K_iK_j\varphi, h, i, j \in \{1, 2\}$: if one agent knows φ, they both know that they both know φ) and wonder what properties can be added to them, without implying a collapse of both operators. We end this section with a theorem about possible combinations.

THEOREM 6.8
Consider the following 'extreme systems' PI, NI, PE and NE, which are systems with two epistemic operators B and K satisfying A0, R0, R1 (for K), A1 and A8 of KB, and such that:

- in PI all instantiations of positive introspection are valid
- in NI all instantiations of negative introspection are valid
- in PE all instantiations of positive extraspection are valid
- in NE all instantiations of negative extraspection are valid.

Then (in the following, the variables X, Y, Z range over $\{K, B\}$ and $K' = B$, whereas $B' = K$):

1. for any $T \in \{PI, NI, PE, NE\}, T \nvdash K\varphi \rightarrow B\varphi$ and $T \nvdash B\varphi \rightarrow K\varphi$
2. $S \nvdash K\varphi \rightarrow B\varphi$ and $S \nvdash B\varphi \rightarrow K\varphi$,
   both for $S = PI \cup NE$ and $S = NI \cup PE$.
3. adding p.e. formula $XY\varphi \rightarrow Z\varphi$ to PI yields $Z'\varphi \rightarrow Z\varphi$
4. adding n.e. formula $X\neg Y\varphi \rightarrow \neg Z\varphi$ to NI yields $Z\varphi \rightarrow Z'\varphi$
5. adding p.i. formula $X\varphi \rightarrow YZ\varphi$ to PE yields $X\varphi \rightarrow X'\varphi$
6. adding n.i. formula $\neg X\varphi \rightarrow Y\neg Z\varphi$ to NE yields $X'\varphi \rightarrow X\varphi$
7. adding n.i. formula $\neg X\varphi \rightarrow Y\neg Z\varphi$ to PI
   yields $(\neg Y false \rightarrow (X'\varphi \rightarrow X\varphi))$
8. adding p.i. formula $X\varphi \rightarrow YZ\varphi$ to NI
   yields $(\neg Y false \rightarrow (X\varphi \rightarrow X'\varphi))$
9. adding n.e. formula $X\neg Y\varphi \rightarrow \neg Z\varphi$ to PE yields $\neg X false$
10. adding p.e. formula $XY\varphi \rightarrow Z\varphi$ to NE yields $\neg X false$
11. adding $\neg X false$ to PI yields the four n.e. formulas $X\neg Y\varphi \rightarrow \neg Z\varphi$
12. adding $\neg X false$to NI yields the four p.e. formulas $XY\varphi \rightarrow Z\varphi$.

PROOF. As an example, we prove 1, 2, 3, 6, 9 and 11.

1. $T = PI$, construct a frame such that R^1 is transitive over (R^2, R^3), for all R^1, R^2, $R^3 \in \{S, T\}$, but $(S \not\subseteq T)$ and $(T \not\subseteq S)$; for instance, $W = \{v, w\}, S = \{(v, v)\}, T = \{(w, w)\}$.
2. For $S = PI \cup NE$, Let $\mathcal{F} = \langle W, S, T\rangle$, with $W = \{t, u, v, w\}$, $S = \{(t, u), (t, w), (u, w), (v, w), (w, w)\}$ and $T = \{(t, v), (t, w), (u, w), (v, w), (w, w)\}$. $\mathcal{F}$ is an S-frame, but $\mathcal{F} \not\models K\varphi \rightarrow B\varphi$ and $\mathcal{F} \not\models B\varphi \rightarrow K\varphi$.
3. Suppose we add the formula $XY\varphi \rightarrow Z\varphi$ to PI. Then immediately: $Z'\varphi \Rightarrow_{PI} XY\varphi \rightarrow Z\varphi$.
6. $X'\varphi \Rightarrow_{p.i.} YZ\varphi \Rightarrow_{\neg Y false} \neg Y\neg Z\varphi \Rightarrow_{added\ n.i.-formula} X\varphi$.
9. $X false \Rightarrow XY\varphi \wedge X\neg Y\varphi \Rightarrow_{p.i.} Z\varphi \wedge X\neg Y\varphi \Rightarrow_{added\ n.e.-formula} Z\varphi \wedge \neg Z\varphi \Rightarrow false$.
11. $\neg X false \Rightarrow (X\neg Y\varphi \rightarrow \neg XY\varphi) \Rightarrow_{p.i.} (X\neg Y\varphi \rightarrow \neg Z\varphi)$. ∎

REMARK 6.9
Theorem 6.8 has many implications. For instance, it follows from 1, 3, 7 and 11, although it is possible to have a system with two maximally p.i.-related operators, adding one p.e. instantiation to it gives either $K\varphi \rightarrow B\varphi$ or $B\varphi \rightarrow K\varphi$. The same holds for adding an n.i.-formula, provided that $\neg Y false$ holds for a suitable Y. Moreover, if we assume the latter, all n.e.-formulas are imported to the theory. Theorem 6.8 shows an asymmetry between systems with maximal introspection, and those with maximal extraspection. For example, adding n.e.-formula $X\neg Y\varphi \rightarrow \neg Z\varphi$ to PE does not yield $(\neg X false \rightarrow (Z\varphi \rightarrow Z'\varphi))$, it just gives $XY\varphi \Rightarrow_{PE} Z\varphi \Rightarrow \neg X\neg Y\varphi$.

7 Conclusions and problems

Studying the BCB-problem, I applied some correspondence theory to multi-modal epistemic logic. Studying this multi-modal system, possible combinations of epistemic properties could be examined systematically. With this general approach I showed that Kraus and Lehmann's KB is saturated with respect to many important properties (such as introspection): adding any of them to KB yields $B_i\varphi \leftrightarrow K_i\varphi$. I investigated one of the many possible systems that are 'close to

KB' and that solves the BCB-problem. This shows that the collapse of knowledge and belief one obtains by adding $B\varphi \rightarrow BK\varphi$ is not caused by the use of Kripke semantics. I argued that the techniques presented in this paper can straightforwardly be applied to probabilistic Kripke structures as well.

By allowing more epistemic operators (for each agent), many notions of belief can be combined. It seems interesting to explore this idea of having a spectrum of beliefs, ranging from weak belief, corresponding with having less alternatives (worlds) in the structure (cf. [15], where a notion of belief is defined as a possibility operator) to knowledge as some 'limit'. This idea might be extended to do a kind of 'quantitative reasoning' as follows. With respect to a relation R, define operators $L_n, n \in \mathbb{N}$, with interpretation of $L_n\varphi$: 'in all, except for at most n worlds, φ is the case'. This enables defining notions like 'φ is believed at least as strong as ψ'. At the moment, I are studying some interesting perspectives offered by this option. The idea of having such 'numerical' modal operators was suggested independently in [6] and [11]. A first application to epistemic logic is to be found in [16].

Acknowledgements

I am thankful to John-Jules Meyer for the fruitful discussions we had, and for his encouraging ideas. Gerard Vreeswijk read a preliminary version of this report. Maarten de Rijke and two anonymous referees gave valuable suggestions which amended this paper in several places.

References

[1] J. F. A. K. van Benthem. Modal correspondence theory. PhD thesis, Instituut voor Grondslagenonderzoek, University of Amsterdam, 1976.

[2] J. F. A. K. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1983.

[3] C. Chang and H. Keisler. *Model Theory*. North Holland, Amsterdam, 1973.

[4] B. F. Chellas. *Modal Logic, an Introduction*. University Press, Cambridge, 1980.

[5] R. F. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence* **34**, 39–76, 1988.

[6] M. Fattorosi Barnaba and F. de Caro. Graded modalities i. *Studia Logica* **44**, 197–221, 1985.

[7] R. Goldblatt. *Logics of Time and Computation*. Number 7 in CSLI Lecture Notes. Standford University, Stanford, 1987.

[8] J. Y. Halpern. The relationship between knowledge, belief and certainty. *Annals of Mathematics and Artificial Intelligence* **4**, 301–322, 1991.

[9] J. Y. Halpern and Y. O. Moses. A guide to the modal logics of knowledge and belief. In *Proceedings IJCAI-85*, pp. 480–490, Los Angeles, CA, 1985.

[10] J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, NY, 1962.

[11] W. van der Hoek. On the semantics of graded modalities. Technical Report IR-246, Free University of Amsterdam, 1991. To appear in *The Journal of Applied Non Classical Logics*.

[12] W. van der Hoek. Qualitative modalities. In *Proceedings of SCAI-91, Roskilde, Denmark*, ed. B. Mayoh, pp. 322–327, Amsterdam, 1991. IOS Press.

[13] W. van der Hoek. Modalities for reasoning abour knowledge and quantities. PhD thesis, Free University of Amsterdam, 1992.

[14] W. van der Hoek and J.-J. Ch. Meyer. Making some issues of implicit knowledge explicit. To appear in *Foundations of Computer Science*.

[15] W. van der Hoek and J.-J. Ch. Meyer. Possible logics for belief. Technical Report IR-170, Free University of Amsterdam, 1988. To appear in *Logique et Analyse*.

[16] W. van der Hoek and J.-J. Ch. Meyer. Graded modalities for epistemic logic. Technical Report IR 261, Free University Amsterdam, 1991. To appear in LNCS 1992.

[17] W. van der Hoek and J.-J. Ch. Meyer. A model-theoretic study of knowledge. Technical Report IRXXX, Free University Amsterdam, 1992.

[18] G. E. Hughes and M. J. Cresswell. *Introduction to Modal Logic*. Methuen, London, 1968.

[19] S. Kraus and D. Lehmann. Knowledge, belief and time. In *Proceedings ICALP*, number 226 in Lecture Notes in Computer Science, ed. L. Knott. Springer, 1986. Extended version in *Theoretical Computer Science* **58**, 155–174, 1988.

[20] G. Lakemeyer. Steps towards a first-order logic of explicit and implicit belief. In J.Y. Halpern, editor, *Proceedings of TARK Conference*, pages 325–340, 1988.

[21] K. Lehrer and T. D. Paxson. Knowledge: Undefeated justified true belief. *The Journal of Philosophy* **66**, 225–237, 1969.

[22] W. Lenzen. *Glauben, Wissen und Warscheinlichkeit*. Springer Verlag, Wien, 1980.

[23] J.-J. Ch. Meyer, W. van der Hoek, and G. A. W. Vreeswijk. Epistemic logic for computer science: A tutorial, Part I. *EATCS Bulletin* **44**, 242–270, 1991.

[24] J.-J. Ch. Meyer, W. van der Hoek, and G. A. W. Vreeswijk. Epistemic logic for computer science: A tutorial, Part II. *EATCS Bulletin* **45**, 256–287, 1991 (part II).

[25] M. de Rijke. Correspondence theory of modal logic. Technical Report, ILC, Amsterdam, 1993.

[26] H. Sahlqvist. Completeness and correspondence in the first- and second-order semantics for modal logic. In *Proceedings of the 3rd Scandanavian Logic Symposium*, ed. S. Kanger. North Holland, 1975.

[27] Y. Shoham and Y. Moses. Belief as defeasible knowledge. In *Proceedings Eleventh International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1168–1173, 1989.

[28] F. Voorbraak. The logic of objective knowledge and rational belief. In *Logics in AI*, number 478 in Lecture Notes in Artificial Intelligence, ed. J. van Eijck, pp. 499–516. Springer, 1990.

Received 22 January 1991