Logique & Analyse **127-128** (1989), 177-194

POSSIBLE LOGICS FOR BELIEF

W. van der Hoek and J.-J. Ch. Meyer,

Abstract

Taking the (modal) logic for belief *KD45* as a base, we show how some properties of belief depend on particular axioms (of that system), while other properties (which can be unified under the name *logical omniscience*) seem to be typical for modal logics that define belief as a necessity. We examine the effects of a variety of axioms in a systematic way and give a sketch, along the lines of a contribution of Fagin and Halpern, of the problems that seem hard to overcome in such a system. However, it appears, that switching from the view of belief-as-necessity to belief-as-possibility solves some of those problems. In particular, closure under implication and belief of valid formulas can be invalidated, and an agent's belief set needs no longer be consistent. We explain how we consider belief-as-possibility and belief-as-necessity as two extreme notions, allowing for a variety of 'beliefs' in between. However, this switch of view in *KD45* does not invalidate properties like *positive* and *negative introspection*. We show that both views allow extensions with Fagin and Halpern's theory of *awareness*, in which an agent needs to be aware of a formula before he can explicitly believe it. Next, we introduce a notion of *principles*, or *prejudices*, that allows an agent to have implicit beliefs. This can model his 'reasoning against the facts'. While awareness can prevent an agent from believing φ, that he would believe if he were logically omniscient, principles even enable him to believe $\neg\varphi$ in such a case.

0. *Introduction.*

We examine a variety of logics for *belief*: logics that are defined in a framework of *modal logic* ([Ch], [HC]), in which *epistemic logic* is treated easily ([Hi], [Mo], [MHV]). An important issue that we shall address is that of *logical omniscience* (l. o.) which, roughly speaking, says that an agent's beliefs are closed under implication. We investigate these problems and study several (sometimes partial) solutions. In particular, it will appear, that many problems can be solved if one changes his view in a modal framework from 'belief as a *necessity*' (which is the traditional one) to 'belief as a *possibility*', as was already suggested in [Me]. Moreover, we will see how additional features can offer a great variety of possible logics for belief.

Before we say more about l. o, we briefly (re-) introduce some terminology. The language L of our (propositional) epistemic formulas is built from a set of atoms p, q, r, ..., the constant *true*, the connectives $\neg$, $\wedge$ and $\vee$ ($\rightarrow$, $\leftrightarrow$ and *false* can be added as abbreviations), and an operator L. This modal operator is also written in the literature as '$\Box$' or 'K' (with intended meaning of $K\varphi$: "φ is known" cf. [HM]). Here, the intended meaning of $L\varphi$ is "φ is believed". These formulas are interpreted on (Kripke) structures (or, models), which are tuples $\mathbf{M} = \langle S, R, \pi \rangle$, with S a set of worlds or states, π an assignment of truth values to the primitive propositions for each state s in S, and R a binary relation on S.

We assume familiarity with the definition of $\models$, with $(\mathbf{M}, s) \models \varphi$ read as "formula φ is *true* in $\mathbf{M}$ at s". We just give the L-clause: $(\mathbf{M}, s) \models L\varphi$ iff for all t such that Rst: $(\mathbf{M}, t) \models \varphi$ (*). If, for given φ, S and R there exists an $s \in S$ and a π such that $(\mathbf{M}, s) = (\langle S, \pi, R \rangle, s) \models \varphi$, we say that φ is satisfiable (at s) in $\mathbf{M}$. If for all s $(\mathbf{M}, s) \models \varphi$, we say that φ is valid on $\mathbf{M}$ and write $\mathbf{M} \models \varphi$. Finally, φ is called (just) valid if it is valid on all Kripke structures $\mathbf{M}$.

A *Kripke frame* $\mathbf{F}$ is a model with no valuation function π yet: $\mathbf{F} = \langle S, R \rangle$. $(\mathbf{F}, w) \models \varphi$ means that for all π, $(\langle S, R, \pi \rangle, w) \models \varphi$ and $F \models \varphi$ is shorthand for (for all w, $(\mathbf{F}, w) \models \varphi$). We say that a modal formula φ *corresponds with* first order property ϕ if it holds that for all frames F, $F \models \varphi \Leftrightarrow F \models \phi$. ϕ will be generally a property of the accessibility relation R.

We define, for $s \in S$, the belief set BS(s) at s as $BS(s) = \{\varphi \mid L\varphi \text{ is true at } s\}$. (When convenient, we leave out reference to s.) From (*), we see that BS(s) consists of all formulas that are true in all (from s) accessible worlds. Since these worlds are (propositional) models, we have:

i) BS(s) is consistent, if s has R-successors,
ii) BS contains all tautologies and
iii) BS is closed under logical deductions.

For many applications of a theory of belief, these properties makes a believer too perfect a reasoner. i) is perhaps less critical, for it suggests a solution: if one does not want just consistent belief sets, one changes the constraints on R. This will indeed be one of the issues we address, but this remedy will not always be sufficient. The cures against an agent's omniscience that we will discuss can be categorised as follows:(1)

1) Explicitate l. o. in its aspects, and find conditions on the structure that affect (omit) them.
2) Give an alternative definition of belief within the modal framework.
3) Make modifications on the belief set BS, given a fixed definition of belief. We will define:
   a) φ is *explicitly* believed iff ($\varphi \in$ BS and φ meets some additional (*awareness*) condition).
   b) φ is *implicitly* believed iff ($\varphi \in$ BS *or* φ meets some alternative (*prejudices*) condition).

1. *Belief-as-Necessity: the System KD45.*

In "Belief, Awareness and Limited Reasoning", Fagin and Halpern [FH] mention some short-comings of the standard way of defining belief. We will not use their 'n-agents' system here, but the following can easily be extended to it. [FH]'s logic of belief (for the case n = 1) is characterized by:

A1 All instances of propositional tautologies.
A2 $(L\varphi \wedge L(\varphi \to \psi)) \to L\psi$.
A3 $\neg L$*false*.
A4 $L\varphi \to LL\varphi$.
A5 $\neg L\varphi \to L\neg L\varphi$.
R1 $\vdash \varphi,\ \vdash \varphi \to \psi \Rightarrow \vdash \psi$.

(1) One direction we typically do not investigate here, is that of allowing 'incoherent', or 'partial' worlds. Such an approach can be found in e.g. [Le]. Fagin an Halpern object against this approach (in [FH]), that believers would become again perfect reasoners, now w.r.t. to this new (relevance) logic. We think this will ultimately be the (only) criticism one can raise against any logic that soundly and completely corresponds with (a notion of) belief.

R2 $\vdash\varphi\Rightarrow\vdash L\varphi$.

Fact 1. This system is sound and complete w. r. t. structures that are serial, transitive and Euclidean.

The system is known as *weak S5* or *KD45* ([Ch], [HC]). Adding A3': $L\varphi\rightarrow\varphi$ (stronger than A3 and corresponding with reflexivity), yields the system *S5* and the corresponding structures become equivalence classes. A3' is also known as the *knowledge axiom*, because it is usually taken to distinguish knowledge from belief. In *KD45*, an agent believes his beliefs behave like knowledge: $L(L\varphi\rightarrow\varphi)$ is valid.

The power of systems like these lies in the flexibility of, syntactically, adding or removing axioms and, semantically, making constraints on the (relation of the) structure. The property of belief represented by A4 is known as *positive introspection*, that of A5 as *negative introspection*. They characterise transitivity (i. e, $\langle S, R, \pi\rangle\models$A4 for all $\pi\Leftrightarrow$R is transitive) and Euclidicity, respectively. Their reverses, $LL\varphi\rightarrow L\varphi$ and $L\neg L\varphi\rightarrow\neg L\varphi$ characterise density and *selective transitivity* ($\forall x\exists y\forall z(Rxy \wedge (Ryz\rightarrow Rxz))$), respectively. For a systematic overview of the connections between validity of several modal formulas (like the introspection formulas) on the one hand and properties of the underlying Kripke models on the other, we refer to [Ho1].

It is of course also interesting to study knowledge (K) and belief (B) in one and the same logical system. One way to do so is to define knowledge and belief as separate entities with some interaction axioms. Such an approach is to be found in [KL]. The question about what should be reasonable interaction properties, and in particular, how many can be added without yielding that knowledge and belief become the same, is then investigated in [Ho1].

An alternative approach to combine knowledge and belief 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 sixties by e.g. [LP]) but an opposite view is taken in [SM], where belief (or rather $B(\varphi, \varphi_{ass})$, the belief in φ relative to some 'unusuallity assertion') is defined in terms of knowledge. Here, we will study both knowledge and belief, but hardly consider to have both notions in one system.

Fact 2. ([HM], [Ho1]). $KD45\vdash LL\varphi\leftrightarrow L\varphi$ and $KD45\vdash\neg L\varphi\leftrightarrow L\neg L\varphi$.

Corollary. (*Manageability of modalities*). Let X be a sequence of symbols from {L, $\neg$} and φ any epistemic formula. Then Xφ can be rewritten to X'φ, with X' a subsequence of $\neg L \neg$.

2. *Logical Omniscience.*

We saw that A3-A5 correspond with particular conditions on R. However, A2 and R2 (and also A1 and R1) seem inevitable if we interpret our system on Kripke structures. This causes *logical omniscience*; l. o. partly: suppose φ is believed, i. e, Lφ. Then φ is true in all accessible worlds, being propositional models, so we can make logical derivations in them. Deriving ψ in them gives Lψ, i.e. ψ is believed. So, belief sets are closed under logical derivations. We list some aspects of l. o. , using 'B' as belief operator for later reference:

LO1	$B\varphi \wedge B(\varphi \rightarrow \psi) \rightarrow B\psi$	*Closure under implication.*
LO2	$\vdash \varphi \rightarrow \psi \Rightarrow \vdash B\varphi \rightarrow B\psi$	*Closure under valid implication.*
LO3	$\vdash \varphi \Rightarrow \vdash B\varphi$	*Belief of valid formulas.*
LO4	$(B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$	*Closure under conjunction.*
LO5	$B\varphi \rightarrow B(\varphi \vee \psi)$	*Weakening of Belief.*
LO6	$B\varphi \rightarrow \neg B \neg \varphi$	*Consistency of Beliefs.*
LO7	$B\varphi \rightarrow \varphi$	*Having no false Beliefs.*
LO8	$B(B\varphi \rightarrow \varphi)$	*Belief of having no false beliefs.*
LO9	B *true*	*Believing truth.*
LO10	More epistemic alternatives yield fewer beliefs.	*Anti-monotonicity in epistemic alternatives*

Fact 3. KD45 suffers from LO1-LO6, and LO8-LO10. Moreover, LO1-LO5, LO9 and LO10 are valid in each system that defines belief as a *necessity operator*, and that satisfies A1, A2, R1 and R2.

3. *Belief-as-Possibility: the System Dual KD45.*

In [FH], the inevitability of A2 and R2 is mentioned: "...No matter how we modify the R-relation, the fact that we say that an agent knows or *believes a fact exactly if this fact is true in all the worlds the agent considers possible*, seems to force us to the situation where an agent knows all tautologies and his knowledge is closed under implication". The italics are ours, and they suggest a way out: we can stick to the possible worlds approach, but escape the consequences of A2 and R2, if we drop the constraint in italics and define a fact to be believed if it is true in (at least) *one* possible world. Or, equivalently, to have a belief operator, say M, that is the *dual* of L, i.e., $M = \neg L \neg$. This has the effect we wanted; a formula $M\varphi$ is true if $L\neg\varphi$ is not: i.e., if $\neg\varphi$ is not true in all possible worlds, so if there is at least one world in which φ is true. In a system in which M is defined, we can simultaneously have a necessity operator which we will interpret as knowledge and denote by K. On serial structures, we have $(K\varphi \rightarrow M\varphi)$.

Considered as a description of belief, the operator M corresponds with a belief of a very credulous person or agent: φ is believed, as long as $\neg\varphi$ is not known for sure. M means believing in the sense of considering it possible, not excluding the possibility(²). If we dualise *KD45*, we get the following:

dual A1 A1, since A1 does not mention any L.
dual A2 $M\psi \rightarrow (M\varphi \vee M(\neg\varphi \wedge \psi))$(³).
dual A3 M true.
dual A4 $MM\varphi \rightarrow M\varphi$.
dual A5 $M\neg M\varphi \rightarrow \neg M\varphi$.
dual R1 R1.
dual R2 $\vdash \varphi \Rightarrow \vdash \neg M \neg \varphi$.

Fact 4. In *dual KD45*, both $M\varphi \leftrightarrow MM\varphi$ and $\neg M\varphi \leftrightarrow M\neg M\varphi$ hold.

(²) This weak notion of belief can solve the paradox Shoham ([Sh]) finds in Moore's ([Mo]) 'older brother' argument. Moore states that he believes (B) he has no older brother, not because his parents told him he did not have one, but because if he did have an older brother, he would know (K) about it. So, he adopts $\neg K\varphi \rightarrow B\neg\varphi$ (or, $\neg B\neg\varphi \rightarrow K\varphi$). Shoham objects against this, that now one also has to adopt "if I did not have an older brother, I would know about it" $(\neg K\neg\varphi \rightarrow B\varphi)$. With our M, it is indeed fully consistent to believe both φ and $\neg\varphi$, as long as one neither knows $\neg\varphi$ nor φ.

(³) Or, equivalently, $(M\psi \wedge \neg M\varphi) \rightarrow M(\varphi \wedge \neg\psi)$ ("If I believe in ghosts but do not believe that they do any harm, I believe in ghosts that don't do any harm".)

Corollary. In *dual KD45*, positive and negative introspection are valid. Adding knowledge (K), we also have $(\neg M\varphi \leftrightarrow K\neg M\varphi)$ and $(K\varphi \leftrightarrow MK\varphi)$.

Remark 1. The main difference between belief-as-possibility and belief-as-necessity (M- and L-beliefs, from now on) is, that in the former an agent gets his (different) beliefs from different worlds. He can, for instance, believe φ (is possible) and believe ψ (is possible), without necessarily believing that $\varphi \wedge \psi$ (is possible in one and the same world). (Note that $(Mp \wedge M\neg p)$ is not the same as $M(p \wedge \neg p)$.)

This reminds on [FH]'s notion of *frame of mind*. In fact, it is a special case of their *local reasoning*, taking each cluster to consist of *one* world. However, we do not need their complex machinery here, and, in their approach, belief is essentially a necessity. However, it also reminds on our treatment in [MH] of (several extensions of) defaults. There, we show that it makes sense to impose an order on such frames of mind (being either worlds or sets of worlds) allowing the agent to make preferences when 'conflicting beliefs' turn up (cf. the famous "Nixon-diamond").

With the difference between L- and M-belief in mind, one easily verifies:

Fact 5. *dual KD45* suffers from LO2, LO3 and LO5. LO3 is not valid if one drops dual A3 (seriality).

Remark 2. Note the difference between dual A3', $\varphi \rightarrow M\varphi$, characterising reflexivity, and the rule $\vdash \varphi \Rightarrow \vdash M\varphi$, corresponding with (the weaker property of) seriality. For M, they both can be denied (although we do have $\vdash \varphi \Rightarrow \vdash M\mathit{true} \rightarrow M\varphi$), for L, the rule $\vdash \varphi \Rightarrow \vdash L\varphi$ is valid. If φ is a tautology, dual R2 implies $(M\neg\varphi \rightarrow \mathit{false})$, whereas A3 implies $(L\neg\varphi \rightarrow L\mathit{false})$ (and thus $L\neg\varphi \rightarrow L\psi$).

dual A3' for knowledge, $\varphi \rightarrow K\varphi$, is used by Shoham ([Sh]) to relate two inferences Moore distinguishes: *autoepistemic*- (If φ, then I'd know it) and *default*- (Most birds fly)inferences. According to Shoham, adding $\varphi \rightarrow K\varphi$ can model both kinds of inferences, because (epistemic) $\varphi \rightarrow K\varphi$ has as its contrapositive (default) $M\neg\varphi \rightarrow \neg\varphi$. We stress here, that one should not add one of them as a *scheme*, but with some instantiations for φ(4). If one adds $\varphi \rightarrow K\varphi$ as an axiom-scheme, then one introduces *Knowledge of truths* (whereas R2 introduces *Belief of validities*). Moreover, adding it to a knowledge system (with A3'), truth, belief and knowledge would collapse.

Remark 3. Property LO10 is especially interesting when studying *development* of belief and knowledge (over time). It is possible to vary the accessibility relation over points of time$\in$T. This gives a nice tool for describing a system in which, for instance, knowledge grows, and at the same time the amount of believed-but-not-known 'facts' decreases. We will not discuss such notion of time here, however.

Remark 4. Of course, L- and M- systems have the same number of undeniable formulas or rules, since for each valid L-formula, there corresponds a valid, dual, M-formula (e. g, since L suffers from LO4, M suffers from its dual, M $\vee$ distr: $M(\varphi \vee \psi) \rightarrow (M\varphi \vee M\psi)$. However, having both definitions at one's disposal one can choose with one's own application in mind.

4. *Awareness and Principles.*

In M-systems we can deny LO9, so that, e. g, $\neg M(p \vee \neg p)$ is satisfiable. This is related with a source of l. o. that is mentioned in ([FH]): *(lack of) awareness*. In [FH] it is argued that one cannot say one knows or does not know p if one is not aware of p, and we think the same holds for belief. Satisfiability of $\neg M(p \vee \neg p)$ is not a real solution, however: how can one say one does not believe p or not p if one is not aware of p? The only way out seems to decide on a meta-level what an agent is aware of: an agent cannot say that he is not aware of p (if he is not aware of p). Such a solution is offered in [FH], inspired by [Le]. The idea behind the approach is, that with the notion of awareness, one can prevent from getting in un-

(4) We don't think it is realistic, to assume for all φ, if it is true, you'd know it (it is more plausible for 'having an older brother' than for Fermat's theorem) or, for all φ, if it is possible, it is the case (it looks more reasonable to assume it for 'Tweety flies' than for 'I win the lottery').

desirable formulas into a belief set. Then, e. g, LO1 need not be valid anymore: φ and $\varphi \rightarrow \psi$ may be believed, without believing ψ. (A priori, no restrictions are made on the set the agent is aware of: he may be aware of $\varphi \rightarrow \psi$, but not of ψ.)

In [FH] a function **A** is added to the structure, that determines what formulas the agent is aware of in each world, and an operator A to the formal system, obeying $(M, s)(=(\langle S, R, \mathbf{A}, \pi\rangle, s)) \models A\varphi$ iff $\varphi \in \mathbf{A}(s)$. Now, we can distinguish between (just) belief and explicit belief. (In [FH], the first is denoted by implicit belief, but we reserve this term for another notion). We define *explicit belief*: $L_e\varphi \equiv_{def} L\varphi \wedge A\varphi$, and $M_e \equiv_{def} M\varphi \wedge A\varphi$. Instead of properties for (just) belief, now we get properties *relativised* to awareness ([5]):

LO1A: $L_e\varphi \wedge L_e(\varphi \rightarrow \psi) \rightarrow (A\varphi \rightarrow L_e\varphi)$.
LO2A: $\vdash \varphi \rightarrow \psi \Rightarrow \vdash L_e\varphi \rightarrow (A\psi \rightarrow L_e\psi)$.
LO3A: $\vdash \varphi \Rightarrow \vdash (A\varphi \rightarrow L_e\varphi)$.
LO4A: $(L_e\varphi \wedge L_e\psi) \rightarrow (A(\varphi \wedge \psi) \rightarrow L_e(\varphi \wedge \psi))$.
LO5A: $L_e\varphi \rightarrow (A(\varphi \vee \psi) \rightarrow L_e(\varphi \vee \psi))$.
LO9A: $A\mathit{true} \rightarrow L_e$ true.

Remark 5.

1) **A** does not make L_e the same (w.r.t. l.o.) as M. $L_e\varphi \wedge L_e(\varphi \rightarrow \psi) \wedge \neg L_e\psi$ may be satisfiable, still $L_e\varphi \wedge L_e(\varphi \rightarrow \psi) \wedge L_e \neg \psi$ is not (As soon as ψ is been aware of, perfect reasoning continues!).
2) **A** cannot, e. g, invalidate $L_e p \vee \neg L_e p$. Considering formulas as *assertions*, this is more problematic (cf. the introduction of this section) than interpreting them as *descriptions of states of affairs*.
3) 'Awareness' must be interpreted in a very broad sense. It does not always behave like common awareness. For instance, $A \neg Ap$ is satisfiable, $L_e \neg Ap$ and even $(A \neg p \wedge \neg Ap)$.
4) **A** is purely syntactical. That makes it difficult to reason about, yielding often ad hoc solutions. E. g, in [FH] it is suggested to let **A** satisfy $((\varphi \wedge \psi) \in \mathbf{A}(s) \Rightarrow \varphi, \psi \in \mathbf{A}(s))$, to overcome satisfiability of $L_e(p \wedge \neg L_e p)$.

([5]) Similar things can be said about M-beliefs. For instance, now, for any tautology $\varphi \neq \mathit{true}$, $(M_e\mathit{true} \wedge \neg M_e\varphi)$ is satisfiable. Generally, if $B\varphi$ is a conclusion in a system for belief, $(B=L$ or $M)$, then so is $(A\varphi \rightarrow B_e\varphi)$.

This solution is not completely correct([6]), but also rather ad hoc: how is it related to that of satisfiability of $L_e\neg(p\rightarrow L_e p)$, now the belief is not *written* as a conjunction?

5) As showed above, **A** can invalidate 'old' rules, e.g. LO1-LO5 are invalidated by it. These properties have positive belief conclusions, i.e. of the form $L\varphi$ ($M\varphi$), with φ any formula, as opposed to *negative* belief conclusions, (of the form $\neg L\varphi(\neg M\varphi)$). **A** blocks such positive conclusions([7]).

Remark 5. 5) only applies to positive belief conclusions, not to negative ones. A can invalidate LO3, $\vdash\varphi \Rightarrow \vdash L_e\varphi$, as opposed to LO6, $L_e\varphi\rightarrow\neg L_e\neg\varphi$, and dual LO3, $\vdash\varphi\Rightarrow\vdash\neg M_e\neg\varphi$. The latter two express that φ does not get in the belief set, and **A**'s blocking function cannot change that. This suggests introducing a function that adds formulas to belief sets. We define **P** that gives a set of formulas for each world, representing the beliefs an agent wants to stick to, perhaps without having any piece of evidence for them or seeing them fit in a complete situation. They can be considered his *principles*, (or *prejudices*, for they may be inconsistent with the facts that are true or even possible).

We define P corresponding to this function **P**, in the same spirit as A corresponds to **A**, and call the resulting systems LP- and MP-systems. We define *implicit* belief: $L_i\varphi\equiv_{def}L\varphi\vee P\varphi$ and $M_i\varphi\equiv_{def}M\varphi\vee P\varphi$. Now it is clear, how an agent can have inconsistent implicit beliefs. And, remark 5, 1) is solved: $L_i\varphi\wedge L_i(\varphi\rightarrow\psi)\wedge L_i\neg\varphi$ is satisfiable. Also, $M_i(\varphi\vee\psi)$ is satisfiable both with $(\neg M_i\varphi\wedge\neg M_i\psi)$ and with $(M_i\neg\varphi\wedge M_i\neg\psi)$. LO3 is solved for M: if φ is a tautology, $M_i\neg\varphi$ is satisfiable. Relativising makes sense again: for a tautology φ, we had $A\varphi\rightarrow L_e\varphi$, and now also $\neg P\neg\varphi\rightarrow\neg M_i\neg\varphi$ (these formulas can also be considered dual). Awareness can prevent tautologies from being (L-) believed, principles can make their negation be

([6]) It is satisfiable at s with: $(p\wedge\neg L_e p)\in\mathbf{A}(s),\forall t(Rst\Rightarrow\pi(t,p)=\mathit{true}$ and $p\notin\mathbf{A}(t))$. We need also $Rst\Rightarrow\mathbf{A}(s)=\mathbf{A}(t)$.

([7]) This blocking, or relativising, is very much reminding of the abnormally predicate in circumscription ([MC]). For instance LO1A, $L_e\varphi\wedge L_e(\varphi\rightarrow\psi)\rightarrow(A\psi\rightarrow L_e\varphi)$ can be read a "If Tweety is a bird and all birds fly, then, if everything is normal w.r.t. Tweety's flying, Tweety flies".

(implicitly) believed. We encourage the reader to check that problems LO1-LO10 are solved now, for M as well as for L.

Remark 6.

1) Awareness and principles serve the same function as [Le]'s partial (incoherent, respectively) worlds. They decrease (increase, respectively) the 'old' belief set([8]). Note that, although $L_i(p \wedge \neg p)$ is satisfiable, $p \wedge \neg p$ is not, distinguishing inconsistent beliefs and incoherent worlds!
2) For B=L, M: $B_e\varphi \rightarrow B\varphi \rightarrow B_i\varphi$ ($L_e\varphi \rightarrow L\varphi$ is shared with [FH] and [Le]). Maximal **A** ($\mathcal{L}$ for all s), equalises B_e and B, minimal **P** ({ } for all s), makes B_i and B collapse.
3) The relations of 2) are also achieved by considering $M = \langle S, R, R_i, R_e, \pi\rangle$ with $R_i \subset R \subset R_e$ for L- systems, ($R_e \subset R \subset R_i$ for M-systems), and requiring e. g, that $L_e\varphi$ is true, if φ is true in all R_e-accessible worlds. But then, the problems **A** and **P** are designed for, would not be solved: e. g, within all explicitly possible worlds, perfect reasoning goes on, yielding LO1 again for L_e.
4) Allowing both functions gives powerful systems. One might add a time function, and require that principles decrease and awareness grows as time goes by. We will not explore this here.

5. *Some properties of Awareness and Principles.*

Since **A** and **P** are purely syntactical, they can behave very unnatural (cf. remark 5, 4)). Additional constraints on **A** and **P** might steer this. Intuitively, **A** might be closed under subformulas, and **P** under deductions. The following constraints (with $\mathbf{A}_3 \Rightarrow \mathbf{A}_i$, and $\mathbf{P}_3 \Rightarrow \mathbf{P}_i$, $i=1, 2$) seem interesting:

A1: $\varphi \in \mathbf{A}(s)$, ψ subformula of $\varphi \Rightarrow \psi \in \mathbf{A}(s)$.
A2: φ and ψ have the same propositional primitives $\Rightarrow (\varphi \in \mathbf{A}(s) \Leftrightarrow \psi \in \mathbf{A}(s))$.
A3: Given a set of constants $\Psi(s)$: $\varphi \in \mathbf{A}(s) \Leftrightarrow$ all propositional constants

([8]) We use the triple explicit, implicit and (just) belief for those restricted, extended and 'old' belief sets, respectively, as opposed to [Le] and [FH], which use implicit belief for what we call (just) belief. (They don't have all three.)

of φ in Ψ(s).

P1: $\varphi \in \mathbf{P}(s)$ and $\vdash \varphi \rightarrow \psi \Rightarrow \psi \in \mathbf{P}(s)$.

P2: $\varphi \in \mathbf{P}(s)$, $(\varphi \rightarrow \psi) \in \mathbf{P}(s) \Rightarrow \psi \in \mathbf{P}(s)$.

P3: $\mathbf{P}(s) \vdash \varphi \Rightarrow \varphi \in \mathbf{P}(s)$.

However, one runs the risk of reintroducing l. o. problems. As remarked in [FH], A1 introduces $L_e\varphi \wedge L_e(\varphi \rightarrow \psi) \rightarrow L_e\psi$ again. In fact, the weaker $A\varphi \wedge A(\varphi \rightarrow \psi) \rightarrow A\psi$ does(9). The same does not hold for P: assuming $\mathbf{P}_2$ does not imply $L_i\varphi \wedge L_i(\varphi \rightarrow \psi) \rightarrow L_i\psi$. Regarding this issue, the notion of *compositionality* makes sense. Let $\Phi(X)$ be a property of modal operator X. Φ is *compositional in* B_e (cpe) if $\Phi(A)$, $\Phi(B) \Rightarrow F(B_e)$. Φ is compositional in B_i (cpi) if $\Phi(P)$, $\Phi(B) \Rightarrow \Phi(B_i)$ (B=L, M).

Fact 6. $F(X) = X\varphi \rightarrow X\psi$ is cpi and cpe. The same holds for $(X\varphi \vee X\psi) \rightarrow X\chi$ and $X\varphi \rightarrow (X\psi \wedge X\chi)$. $\Phi(X) = (X\varphi \wedge X\psi) \rightarrow X\chi$ is cpe. The latter is not cpi, as the avoidability of both LO1 and LO4 for implicit belief shows (Note that if Φ_1 and Φ_2 are compositional, $\Phi_1 \rightarrow \Phi_2$ need not be). Finally, $\Phi(X) = X\varphi \rightarrow (X\psi \vee X\chi)$ is cpi, not cpe. We see that, if $\Phi_1(X) \rightarrow \Phi_2(X)$ is compositional, one can lose this property for B_i if one strengthens the antecedent, and for B_e if one weakens the consequent.

A topic we investigated further in [HoM] is that of *correspondences* for implicit and explicit belief. We give some results. An easy case is that of *seriality*. The implicit form of A3: $\neg L_i\mathit{false}$ is sufficient for seriality, not necessary: $(\langle S, p, R, \mathbf{P}\rangle, s) \models \neg L_i\mathit{false}$ already if $\mathit{false} \in \mathbf{P}(s)$. We state:

Ch1 a. $\forall s(\exists t\ Rst \wedge \mathit{false} \notin \mathbf{P}(s)) \Leftrightarrow \langle S, R, \mathbf{P}, \pi\rangle \models \neg L_i\mathit{false}$.
b. R is serial $\Rightarrow \langle S, R, \mathbf{P}, \pi\rangle \models L_i\mathit{false} \rightarrow P\mathit{false}$.
c. R is serial $\Leftrightarrow (\langle S, R, \mathbf{P}, \pi\rangle \models \neg L_i\mathit{false} \Leftrightarrow \forall s\,\mathit{false} \notin \mathbf{P}(s))$.

Ch2 a. $\forall s(\exists t\ Rst \vee \mathit{false} \notin \mathbf{A}(s)) \Leftrightarrow \langle S, R, \mathbf{A}, \pi\rangle \models \neg L_e\mathit{false}$.
b. R is serial $\Rightarrow \langle S, R, \mathbf{A}, \pi\rangle \models \neg L_e\mathit{false}$.
c. R is serial $\Leftrightarrow (\langle S, R, \mathbf{A}, \pi\rangle \models \neg L_e\mathit{false}$ for all A).

Next, transitivity. In [FH] is stated, that (R is transitive and (Rst$\Rightarrow \mathbf{A}(s) =$

(9) We use constraints on A and axioms in A interchangeably. Of course, this A-axiom corresponds with a constraint on A, analogous to P_2 for P. We will do the same with constraints on P and axioms for P.

A(t)) imply that $\langle S, p, R, \mathbf{A}\rangle \models L_e\varphi \rightarrow (AL_e\varphi \rightarrow L_eL_e\varphi)$ (A4, relativised to awareness). It appears that transitivity of R together with Rst $\Rightarrow$ **A**(s) $\subset$ **A**(t) is sufficient. For L_i, we immediately get a relativised result:

Ch3 a. R is transitive $\Leftrightarrow \langle S, R, \mathbf{P}, \pi\rangle \models L_i\varphi \rightarrow (\neg P\varphi \rightarrow L_iL_i\varphi)$ for all **P**
b. R is transitive and Rst implies $\mathbf{P}(s) = \mathbf{P}(t) \Rightarrow \langle S, R, \mathbf{P}, \pi\rangle \models L_i\varphi \rightarrow L_iL_i\varphi$.

To prove a, '⇒', note that $(L_i\varphi \wedge \neg P\varphi) \Rightarrow L\varphi \Rightarrow LL\varphi$. Now, the L's can be rewritten to L_i's (first the second, then the first, mind **P**'s syntactical behaviour!). For '⇐', take **P** such that **P**(s) is empty for all s, then $M \models (L\varphi \vee$ *false*$) \rightarrow (\text{true} \rightarrow L(L\varphi \vee$ *false*$) \vee$ *false*$)$, or $M \models L\varphi \rightarrow LL\varphi$, so R is transitive.

Related with the issue of correspondence is that of finding *reduction-properties* in *(dual) KD45AP*. In *KD45*, we have such reductions, cf. fact 2 (we consider $L\varphi \rightarrow LL\varphi$ also a reduction, because it rewrites $\neg LL\varphi$ to $\neg L\varphi$). Reductions keep the system manageable, especially if it is enriched with A and P. For, then $O_1O_2\varphi$ with $O_1, O_2 \in \{L, L_i, L_e, P, A\}$ has 25 instantiations, all being different if no reductions hold. By way of example, we mention some reductions that hold under certain conditions. E. g, transitivity and $\mathbf{A}_2$ are sufficient for $M_eM\varphi \rightarrow M_e\varphi$. Also, transitivity with $\mathbf{A}_2$': $(M\varphi \in \mathbf{A}(s) \Rightarrow \varphi \in \mathbf{A}(s))$ is. (Note that $\mathbf{A}_2$' yields $AM\varphi \rightarrow A\varphi$ and use fact 6 with $\Phi(X) = XM\varphi \rightarrow X\varphi$). We have:

Ch4 R is transitive and **A** satisfies $\mathbf{A}_2$' $\Rightarrow \langle S, R, \mathbf{A}, \pi\rangle \models M_eM\varphi \rightarrow M_e\varphi$, and, also,
Ch4' R is transitive $\Leftrightarrow$ (**A** satisfies **A2**' $\Rightarrow \langle S, R, \mathbf{A}, \pi\rangle \models M_eM\varphi \rightarrow M_e\varphi$).

However, in Ch4' we may not reverse '⇒' (however, cf. Ch6): consider **A**, with $\mathbf{A}(s) = \{M(p \wedge \neg p)\}$ for all s and some p. **A** does not satisfy $\mathbf{A}_2$', but still for no π, φ and s $(\langle S, R, \mathbf{A}, \pi\rangle, s) \models M_eM\varphi \wedge \neg M_e\varphi$. Either $\varphi = (p \wedge \neg p)$, giving $\mathbf{M} \models \neg MM\varphi$, thus $\mathbf{M} \models \neg M_eM\varphi$), or not, yielding $\mathbf{M} \models \neg AM\varphi$, or $\mathbf{M} \models \neg M_eM\varphi$. So, for '⇐' we need A2*: for all s and φ with $MM\varphi$ satisfiable at s: $(M\varphi \in \mathbf{A}(s) \Rightarrow \varphi \in \mathbf{A}(s))$.

Ch5 R is transitive⇔(**A** satisfies **A2*** $\Leftrightarrow \langle S, R, \mathbf{A}, \pi\rangle \models M_eM\varphi \rightarrow M_e\varphi$).

In [Hom], we derive results in the spirit of Ch4, Ch4' and Ch5 for several (combinations of) operators. Sufficient conditions for reductions in *(dual) KD45A* are often straightforward. We distinguish between conditions on the structure of **A** (e.g. , ($A\varphi \in \mathbf{A}(s) \Leftrightarrow AB\varphi \in \mathbf{A}(s)$) is sufficient for ($B_e\varphi \leftrightarrow B_eB\varphi$) and conditions on $\mathbf{A}(s)$ in relation with $\mathbf{A}(t)$ if Rst (e.g. , Rst $\Rightarrow \mathbf{A}(s)=\mathbf{A}(t)$ is sufficient for $(B_e\varphi \leftrightarrow BB_e\varphi))(B=L, M)$. The same holds for **P** in (*dual*) *KD45P* (in the two examples above, replace **A** by **P**, A by P and $_e$ by $_i$ to get results for B_i). As an illustration, we prove in *dual KD45* that (Rst $\Rightarrow \mathbf{P}(s)=\mathbf{P}(t)$) implies $(M_i\varphi \leftrightarrow MM_i\varphi)$. Rst $\Rightarrow \mathbf{P}(s)=\mathbf{P}(t)$ corresponds to $(P\varphi \rightarrow (M\mathit{true} \rightarrow MP\mathit{true})) \wedge (MP\varphi \rightarrow P\varphi)$. So, to *dual KD45P*, we add A6: $(P\varphi \leftrightarrow MP\varphi)$. Let MI be $M_i\varphi \equiv M\varphi \vee P\varphi$:

$\rightarrow$: $MM_i\varphi \Rightarrow_{MI} M(M\varphi \vee P\varphi) \Rightarrow_{\text{rev } M\vee\text{-distr.}} MM\varphi \vee MP\varphi \Rightarrow_{\text{dual } A4} M\varphi \vee MP\varphi \Rightarrow_{A6} M\varphi \vee P\varphi \Rightarrow_{MI} M_i\varphi$.

$\leftarrow$: $M_i\varphi \Rightarrow_{MI} M\varphi \vee P\varphi \Rightarrow_{\text{fact2}} MM\varphi \vee P\varphi \Rightarrow_{A6} MM\varphi \vee MP\varphi \Rightarrow_{LO5} M(M\varphi \vee P\varphi) \vee M(M\varphi \vee P\varphi) \Rightarrow_{A1} M(M\varphi \vee P\varphi) \Rightarrow_{MI} MM_i\varphi$.

Why is it that we only seem to be able to state *sufficient* conditions in, for instance, Ch4? If we would try to prove '$\Leftarrow$' in Ch4, we might start by assuming that R is not transitive and then try to that in our model $\mathbf{M}=\langle S, R, \mathbf{A}, \pi\rangle$ there is some state s for which for some formula φ, $(\mathbf{M}, s) \models M_eM\varphi \ldots \neg M\varphi$. However, $\mathbf{A}(w)$ may be empty for all $w \in S$ (and still satisfy $\mathbf{A}_2$'), so that it appears to be impossible to find an s for which $M_eM\varphi \in \mathbf{A}$-(s). Therefore, it seems appropriate to abstract from the particular function **A** in the model.

In order to do so, we need some definitions. $\langle S, R, \mathbf{A}\rangle \models \varphi$ means that for all π, $\langle S, R, \mathbf{A}, \pi\rangle \models \varphi$. Let φ be any scheme in our extended language (possibly containing L_e, M_e, L_i, M_i). Let $\varphi_{\mathbf{A}}$ be some constraint on awareness functions, and φ_R a property of R. We say that φ *characterises* φ_R *under* φ_A iff for all frames $\langle S, R\rangle$:
R satisfies $\varphi_R \Leftrightarrow$ for all **A** for which $\varphi_{\mathbf{A}}$ holds, $\langle S, R, \mathbf{A}\rangle \models \varphi$.
These definitions are easily extended to cases in which we add P or both A and P to a frame.

Ch6 To see that using the notions defined above give complete characterisation results, we state the following. Let φ_R be 'R is transitive'.

i $MM\psi \rightarrow M\psi$ (φ) characterises φ_R under 'true' ($\varphi_{\mathbf{A}}$)
ii $M_eM_e\psi \rightarrow M_e\psi$ (φ) characterises φ_R under $\varphi_{\mathbf{A}} = \forall s(M_e\alpha \in \mathbf{A}(s) \Rightarrow \alpha \in \mathbf{A}(s))$
iii $M_eM\psi \rightarrow M_e\psi$ (φ) characterises φ_R under $\varphi_{\mathbf{A}} = \forall s(M\alpha \in \mathbf{A}(s) \Rightarrow \alpha \in \mathbf{A}(s))$
iv $MM_e\psi \rightarrow M_e\psi$ (φ) characterises φ_R under $\varphi_{\mathbf{A}} = \forall s\forall t(Rst \ldots \alpha \in \mathbf{A}(t) \Rightarrow \alpha \in \mathbf{A}(s))$.

To show what we have gained with respect to Ch4, we prove iii, i.e. we show that R is transitive$\Leftrightarrow$ for all $\mathbf{A}$ such that $\forall s(M\alpha \in \mathbf{A}(s) \Rightarrow \alpha \in \mathbf{A}(s))$: $\langle S, R, \mathbf{A}\rangle \models M_e M\psi \rightarrow M_e\psi$.

'$\Rightarrow$' Suppose for some π and s $(\langle S, R, \mathbf{A}, \pi\rangle, s) \models M_e M\psi$. Then both $M\psi \in \mathbf{A}(s)$ and for some t with Rst there is some u with Rtu and $(\langle S, R, \mathbf{A}, \pi\rangle, u) \models \psi$. By transitivity of R, we have Rsu and hence $(\langle S, R, \mathbf{A}, \pi\rangle, s) \models M\psi$. Since (by $\varphi_{\mathbf{A}}$) $\psi \in \mathbf{A}(s)$, we have $(\langle S, R, \mathbf{A}, \pi\rangle, s) \models M_e\psi$.

'$\Leftarrow$' Suppose R is not transitive, i.e. for some s, t and $u \in S$ we have Rst, Rtu but not Rsu. We have to define an $\mathbf{A}$ that satisfies $\varphi_{\mathbf{A}}$, a formula ψ and a state v such that $(\langle S, R, \mathbf{A}, \pi\rangle, v) \models \neg(M_e M\psi \rightarrow M_e\psi)$. Choose $v=s$, $\psi=p$, and π such p is true only in state u. Finally, let $\mathbf{A}(s) = \{Mp, p\}$. Then $\mathbf{A}$ can be extended (for the other states) in a way that it satisfies $\varphi_{\mathbf{A}}$. We conclude that $(\langle S, R, \mathbf{A}, \pi\rangle, s) \models M_e Mp \wedge \neg M_e\psi$.

6. *Conclusions.*

We have dicussed a number of problems that classical logic of belief-as-necessity systems suffer from. All these problems more or less had to do with logical omniscience, which generally says that an agent's beliefs are closed under implication. For many applications (for instance, when modelling human beliefs) these properties make the agent too perfect a reasoner with respect to his beliefs.

We showed that several problems of this kind are solved by switching from viewing belief as a necessity to belief as a possibility. This enables an agent having inconsistent beliefs, without him having to belief inconsistencies. We mentioned a peculiar resemblance with (having several extensions in) default logic: in fact, in [MH], our treatment of defaults is inspired by allowing several frames of mind for one agent to reason within: allowing him to have several "preferred", or "working" beliefs at a time. We also show how one can, by having several operators (and perhaps imposing a preference relation upon them) keep information about which frame the belief comes from.

Next, we discussed Fagin & Halpern's notion of awareness, a syntactical utility which can solve most of the problems of logical omniscience. However, it only acts as a filter on belief sets: if φ would be in the belief set (in a system without awareness) the awareness function can prevent it from getting in the (explicit) belief set. In that case, φ is not believed, but it is still impossible to (explicitly) believe $\neg\varphi$. We defined a notion of priciples, or prejuidices, to overcome the latter problem. With this utility, an agent can reason "against the facts".

The power of the developped notions is that they can all be combined into one and the same logical framework, i.e. modal logic, with its own natural and clear semantics. Although, at first sight, a major drawback seems to be that the awareness- and prejuidices-functions are purely syntactically defined functions on the Kripke models, yielding different systems for each added property, H. Wansing showed in [Wa] that all our approaches, as well as Fagin and Halpern's logic and Levesque's logic of implicit and explicit belief, can be uniformly modelled in the framework of Rantala's (nonnormal worlds) semantics.

We think that a closer investigation in finding special awareness and principles functions (giving natural, useful and "well-behaved" systems), as well as a study of their possible or desirable inferences (depending on the particular application), would be worthwhile. In particular, when incorporating time, some additional requirements on these two functions seem to yield interesting systems.

Beside the results of [Wa], yet we know of two places where the topic of awareness is picked up. In [HK] several sources of awareness are systematically studied. Also, given a notion of ('regular') awareness, they introduce an alternative implication, under which the explicit beliefs of the agent are closed. Furthermore, in [Th] the notion of *monotonicity* (as a consraint on awareness functions, like $Rst \Rightarrow \mathbf{A}(s) \supset \mathbf{A}(t)$) plays an imminent role. Also several semantics for systems with awareness are introduced and compared to those of [Wa].

Another interesting issue for further research seems the following ;*Belief-as-possibility* and *belief-as-necessity* can be considered two extremes on one scale: in the former, the agent beliefs φ iff φ is true in more that zero worlds he considers epistemically possible. In the latter, however, the agent believes φ iff there are at most zero worlds considered possible verifying $\neg\varphi$. It seems that this allows for a wide scope of belief operators, all differing in the number, or proportion, of verifying worlds. We think that here the so called "graded modalities" can play an important role (cf. [FC], [Ho2]).

Finally, we mention the approach in which, in stead of having a variety of belief operators, one has a variety of (qualitative) judgements over believed facts, allowing statements as "φ is believed at least as strongly (preferred over)ψ" ([Se], [Gä], [Ho3]). A notion of belief (B) may then be (re-) introduced by way of $B\varphi \equiv$ '$\varphi$is preferred over $\neg\varphi$' ([Len]). It appears, that under this definition, several elements of logical omniscience resolve.

Free University Amsterdam

ACKNOWLEDGEMENTS

We like to thank the Amsterdam working group on non-monotonic reasoning for their encouraging discussions and useful comments.

REFERENCES

[Ch] B. Chellas, *Modal Logic, an Introduction*, Cambridge University Press (1980).

[FH] R. F. Fagin and J. Y. Halpern, *Belief, Awareness, and Limited Reasoning*, in: A. I. , 34 (1988) 39-76.

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

[Hi] Hintikka, *Knowledge and Belief*, Cornell University Press, Ithaca, NY, (1962).

[HK] Z. Huang & K. Kwast, Awareness, *Negation and Logical Omniscience*, in Lecture Notes in AI, J. van Eijck (ed.), 478, (proc. JELIA), Springer (1990), 282-301.

[HM] J. Y. Halpern and Y. O. Moses, *A Guide to the Modal Logics of Knowledge and Belief*, in: Proceedings IJCAI-85, Los Angeles, CA (1985) 480-490.

[HoM] W. van der Hoek & J. -J. Ch. Meyer, *Possible Logics for Belief*, Report Free University IR-170 (1988).

[Ho1] W. van der Hoek, *Systems for Knowledge and Beliefs*, Report Free University IR -165 (1989). To appear in Proceedings of JELIA 1990, Amsterdam.

[Ho2] W. van der Hoek, *Numerieke Modale Logika*, paper University of Groningen (1986).

Ho3] W. van der Hoek, *Qualitative Modalities*, Report Free University IR- 217 (1990)

[KL] S. Kraus & D. Lehmann, Knowledge, *Belief and Time*, in: Lecture Notes in Computer Science, L. Kott (ed), 226, ICALP (1986). Extended version in Theoretical Computer Science 58 (1988), 155-

174.

[Le] H. J. Levesque, *A Logic of Implicit and Explicit Belief*, in: Proceedings AAAI-84, Austin, TX, (1984) 198 - 202, or, also, Lab. Tech. Rept. FLAIR #32, Palo Alto, CA, (1984).

[Len] W. Lenzen, *Glauben, Wissen und Warscheinlichkeit*, Wien, Springer Verlag (1980).

[LP] K. Lehrer & T. D. Paxson, Knowledge: *Undefeated Justified True Belief*, in: The Journal of Philosophy vol. 66, no. 8 (1969), 225-237.

[MC] J. McCarthy, Circumscription - *A Form of Non-Monotonic Reasoning*, in: A. I. , 13 (1980) 27-39, 171-172.

[Me] J. -J. Ch. Meyer, *A Monotonic Logic for Default Reasoning*, Rapport IR-124, Vrije Universiteit Amsterdam (1987).

[MH] J. -J. Ch. Meyer & W. van der Hoek, *Non-Monotonic Reasoning by Monotonic Means*, in: Lecture Notes in AI 478, J. van Eijck (ed.), Springer, (1990), 399 - 411. Extended version to appear in W. van der Hoek, J. -J. Ch. Meyer, Y. H. Tan & C. Witteveen (eds.), *Non-Monotonic Reasoning & Partial Semantics*, Ellis Horwood, Chichester (1991).

[Mo] R. C. Moore, *Semantical Considerations on Nonmonotonic Logic*, in: A. I. 25 (1985) 75-94.

[Sh] Y. Shoham, *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*, MIT Press, Cambridge, Mass. (1988).

[SM] Y. Shoham, Y. Moses, *Belief as Defeasible Knowledge*, in Proc. Eleventh International Joint Conference on Artificial Intelligence (IJCAI) (1989), 1168-1173.

[Th] E. Thijsse, *On Total Awareness Logics*, in Colloquium on Modal Logic 1991, M. de Rijke (ed.), Dutch Network for Language, Logic and Information (1991), 253-278.

[Wa] H. Wansing, *A General Possible Worlds Framework for Reasoning about Knowledge and Belief*, Studia Logica 49 (1990) 523-539.