Logique & Analyse **133-134** (1991), 221-250

LOGICS OF CONSCIOUSNESS EXPLAINED AND COMPARED: PARTIAL APPROACHES TO ACTUAL BELIEF(*)

Elias G.C. THIJSSE

1. *Introduction*

What is consciousness? I shall not even try to give a full answer to this question, in fact I consider it to go well beyond the capacity of human beings. However, some aspects of consciousness can and will be studied. Here the focus is on awareness and belief. In particular, what are the logical properties of actual belief?(1)

It is claimed in this article that partial semantics provides a very intuitive and sound approach to conscious belief, especially when it is combined with classical semantics. This so-called hybrid system, which can be given a more psychological twist by imposing syntactic filters, is compared to other proposals made in the literature, including Fagin & Halpern's logic of general awareness, and Levesque and Lakemeyer's 4-valued approach.

There are two possible strategies towards an adequate description of awareness and actual belief, as well as the general model theory required. One is to inspect actual belief and express its properties in modal logic, the other is to avoid so-called logical omniscience. Of course in the end these strategies converge, but I will start with the latter.

The impetus to what might be called *awareness logic* are the problems of 'logical omniscience' (LO).(2) This ironic term refers to the fact that standard logics such as *normal* modal logics fall short when they are applied to certain cognitive modes of human beings (or their simulations in AI). The problem is that these logics would force the agent to know or believe simply

(*) I acknowledge the helpful comments of Johan van Benthem, Joe Halpern, Jan Jaspars, Emiel Krahmer, John-Jules Meyer, Reinhard Muskens and Heinrich Wansing. Apart from the reconsideration of 4-valued approaches, this article is based on my dissertation [20]. For reasons of space the proofs have mostly been left out, but are covered in [22].

(1) Note that, unless stated otherwise, awareness and consciousness are not distinguished, nor 'actual', 'active', 'explicit' and 'conscious' belief. Sometimes 'believe' is replaced by 'know' when more convenient.

(2) See also the excellent introductions in [3] and [6].

too much. More precisely, they would oblige a person to know all the consequences of his or her knowledge. For example, all number theorists would 'know' whether Goldbach's conjecture (or some other open problem) holds or not, assuming they know the postulates for ordinary arithmetic.([3]) This is surely not the case, in any realistic sense of the word 'know': though these mathematicians may be said to *implicitly* know the answer to this classical query, nobody is aware of the answer, i.e. nobody knows it *explicitly*, so far. Or, more simply and perhaps even more convincingly, if somebody believes p, he need not (explicitly) believe p *or* q. In fact both arguments can already be given for the minimal normal logic K, in particular due to the principles K and I, respectively. These and some other forms of omniscience are listed below.

N	$\vdash \varphi \Rightarrow \vdash B\varphi$
K	$\vdash B(\varphi \to \psi) \to (B\varphi \to B\psi)$
C	$\vdash (B\varphi \wedge B\psi) \to B(\varphi \wedge \psi)$
I	$\vdash \varphi \to \psi \Rightarrow \vdash B\varphi \to B\psi$
E	$\vdash \varphi \leftrightarrow \psi \Rightarrow \vdash B\varphi \leftrightarrow B\psi$

Notice that, within classical propositional logic, these principles are ordered by the consequence series NK $\Rightarrow$ I $\Rightarrow$ E and the fact that K and C are equivalent modulo I. The weakest principles, such as E and C, will be the hardest to eliminate.

Now it may seem easy to circumvent the problems of logical omniscience by limiting the inferential power: simply drop the LO principles from the deductive system. Although this is precisely what awareness logics do, there are a number of complications.

One is that there are many sorts of awareness and logical omniscience, and it appears to be difficult to capture all of them in one fell swoop. Not all forms of LO are contained in the above list. So here 'positive thinking' may be productive: what principles do constitute the inferential system for actual belief? At least one wants to keep good-old classical propositional logic (pL) and its modal instantiations. For example, $Bp \vee \neg Bp$ should be valid, but $B(p \vee \neg p)$ should not. In other words, pL should hold in the *external* part of the logic while avoiding omniscience in the *internal* part. The minimal awareness logic is thus simply pL applied to the modal lan-

([3]) Supposing the conjecture is not independent of Peano's axioms.

guage. However, some weak principles such as the converse of C

$$C_c \quad \vdash B(\varphi \wedge \psi) \rightarrow (B\varphi \wedge B\psi)$$

also seem fully acceptable for active belief.([4]) Yet notice that, *modulo* pL and the innocent principle C_c, the implication rule I is equivalent to the 'extensionality principle' E.([5]) So with regards to omniscience, E and I should be put on a par.

Just presenting the intuitively correct inference rules will not do. For many purposes, such as a quick and easy method for showing that some argument is invalid, one would like to present a sound and complete class of models. This opens the quest for a suitable model theory, which is one of the main themes in this paper. Here some subtlety is required. On the one hand, a simple syntactic interpretation (essentially treating modal formulas such as $B\varphi$ as propositional variables) does not lead to concise and insightful models. Moreover, the relation between implicit and explicit belief is rather obscure in 'syntactic semantics'. On the other hand, a straightforward partial logic, which eliminates some forms of LO, will also destroy pL.

This article presents a discussion and comparison of several partial approaches to conscious belief. In the next section some motivation for 'going partial' is given. Then there is an outline of my own proposal, worked out in [20] and [22]. This so-called *hybrid system* and its modification called the *hybrid sieve system* are subsequently compared to the logics of special and general awareness presented in [3]. Next my three-valued approach will be compared to the four-valued proposals by Levesque and others. Finally it is shown that the non-standard semantics of [4] is isomorphic to a four-valued logic.

2. *Going partial*

First I give some motivation for 'going partial'. After all, it was shown in [3], [18], [24], and [21] that there are very powerful total frameworks (viz. that of *sieve* semantics and *non-normal* world semantics) that can solve the problem of modelling an arbitrary modal logic that extends the classical

([4]) Yet in [16] C_c is rejected for resource bounded knowledge in distributed systems.

([5]) See [2], theorem 8.11 (1).

propositional calculus. Although the problem of modelling weak logics for such psychological notions as awareness and actual belief is thus solved on a technical level, one would prefer a more compelling and natural representation device. A more natural approach to the virtues of consciousness and the vices of logical omniscience is to move to partial semantics, where the classical truth value (*true* and *false*) may be *undefined* and sometimes even *overdefined*, leading to an, essentially, 3- or 4-valued logic. After all, the very idea of partiality is that one conceives or considers only part of the world, i.e. the part one is aware of in one's perception or belief. Such a partial world will be called a *situation* henceforth. I proceed by reinspecting the standard version of partial modal logic and investigate into its suitability for modelling belief and awareness.

The language

The initial representation language $\mathcal{L}_{\neg,\wedge,\{B_i\}_i}(\mathcal{P})$ of multi-modal propositional logic contains the usual connectives, $\neg$ and $\wedge$, propositional variables from $\mathcal{P}$, and modal operators B_i standing for 'agent i explicitly believes that'.(6) The other operators are introduced by definition: $\varphi \vee \psi = \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi = \neg\varphi \vee \psi$, and $\hat{B}_i\varphi = \neg B_i \neg\varphi$.

Standard semantics

First consider a partial semantics for $\mathcal{L}_{\bar{B}}$ where the situations are *coherent*.(7) $M = \langle S, \bar{B}, V\rangle$ is a partial multi-modal Kripke model in which S is a set of situations (or: partial worlds), $B_i \subseteq S \times S$ is an accessibility relation for each i and V a partial valuation function, i.e. V: $\mathcal{P} \times S \xrightarrow{\text{prt}} \{0,1\}$. The standard truth and falsity conditions are as follows: ($B_i[s] = \{t \mid sB_it\}$, for convenience)

$s \models p \Leftrightarrow V(p, s) = 1$ $(p \in \mathcal{P})$ $\quad$ $s \not\models p \Leftrightarrow V(p, s) = 0$ $(p \in \mathcal{P})$

$s \models \neg\varphi \Leftrightarrow s \not\models \varphi$ $\quad$ $s \not\models \neg\varphi \Leftrightarrow s \models \varphi$

$s \models \varphi \wedge \psi \Leftrightarrow s \models \varphi$ & $s \models \psi$ $\quad$ $s \not\models \varphi \wedge \psi \Leftrightarrow s \not\models \varphi$ or $s \not\models \psi$

$s \models B_i\varphi \Leftrightarrow \forall t \in B_i[s] : t \models \varphi$ $\quad$ $s \not\models B_i\varphi \Leftrightarrow \exists t \in B_i[s]: t \not\models \varphi$

(6) Since the number of agents is usually finite, the language is henceforth symbolized as $\mathcal{L}_{\bar{B}}$, or whatever suits the context.

(7) See [19] and [20], chapter 4.

$s \models \varphi$ should be read as: 's supports (verifies) φ' or 'φ is true in s', and $s \models\!\!\!| \varphi$ as: 's rejects (falsifies) φ' or 'φ is false in s.' The notion of validity is *verification*:

$$\Sigma \models \varphi \quad \text{iff} \quad M, s \models \Sigma \Rightarrow M,s \models \varphi \text{ for all } M \text{ and } s.$$

If $\Sigma = \varnothing$ the definition amounts to *absolute* validity of the formula φ, otherwise one deals with *relative validity*, in other words, with *strong consequence*. (Strong) equivalence is defined as mutual strong consequence:

$$\varphi \equiv \psi \text{ iff } \varphi \models \psi \;\&\; \psi \models \varphi$$

Discussion

The standard semantics has a remarkable feature: there are *no* valid formulas any more. For example the singleton model with a self-accessible situation in which each atom is undefined does not support any complex formula either. The absence of absolute validities entails that one type of overidealization of belief and knowledge has been removed. In other words, the usual types of omniscience connected to the modal schemes K and C are circumvented. Moreover, though the inference rules N, I and E are vacuously valid (since the validity of the premise cannot be realized), they are innocuous now: these rules have no input, and therefore no output either. For example, $B_i(p \vee \neg p)$ and $B_i(B_i p \vee \neg B_i p)$ are neither valid nor produced by N.

Although the logic deals with belief rather than with awareness, it also provides an indirect route to awareness (or, rather, acquaintance): somebody may be said to be aware of (or, acquainted with) a fact φ, if every basic fact p in φ has a definite truth value (1 or 0) in every situation the agent considers possible from the situation she is in, in other words, if she explicitly believes $p \vee \neg p$.

$$A_i\varphi = \bigwedge\!\!\!\!\bigwedge_{p \text{ in } \varphi} B_i(p \vee \neg p)$$

Deriving awareness from explicit beliefs is a promising way to reintroduce one of the central notions in the field.

So, the purely partial semantics for this multi-modal logic seems quite successful, and this could be the end of the story. However, there are a number of difficulties:

- the impossibility of absolute validity apparently excludes the incorporation of additional properties which are needed to model various types of knowledge and belief. *Positive* and *negative introspection*, i.e. 'knowing of what you (do not) know that you (do not) know it', *truth* of knowledge, and *consistency* of belief ('not believing contradictions') cannot be encoded in the usual schemes 4, 5, T and D, respectively.([8]) This will prove to be a minor point.
- the impossibility of absolute validity also excludes intuitively correct *objective* facts such as $B_i p \vee \neg B_i p$.([9]) More generally, one would prefer a logic that at least contains (the modal substitutions of) classical propositional logic. This is a major point.
- unlike absolute validity, relative validity is obtained. Then it turns out that many of the eliminated forms of LO pop up again in relativized form. This is also a major point.

In a way, the first point is cancelled by the third: if the usual types of LO, which are captured by basic modal schemes, are obtainable in a relative shape, this may also hold for schemes such as 4. Instead of $\vdash \varphi \rightarrow \psi$ one may then consider $\varphi \vdash \psi$. For frame completeness one usually has to include its contrapositive $\neg\psi \vdash \neg\varphi$, for model completeness single rules qualify.([10])

The second point is more serious than the first, since this may involve other formulas than implications: the closest counterpart of *tertium non datur* $\vdash \varphi \vee \neg\varphi$ seems to be $\varphi \vdash \varphi$, which is valid in the purely partial semantics under consideration, but hardly reflects the original scheme. As will be shown in the sequel, there are fairly easy ways to solve the problem of incorporating propositional logic. However, containment of tautologies may involve the restoration of the deduction theorem and so a solution to the second problem may reinforce the lurking danger observed in the third point: a revival of omniscience connected to, for example, K and I.

([8]) 4 stands for $B\varphi \rightarrow BB\varphi$, 5 for $\hat{B}\varphi \rightarrow B\hat{B}\varphi$, T for $K\varphi \rightarrow \varphi$, and D for $B\varphi \rightarrow \hat{B}\varphi$.

([9]) This may be contrasted to a *subjective* assertion such as $B_i(p \vee \neg p)$. I use the terms objective/subjective in an intuitive sense. The distinction involved does not correspond to non-modal vs fully modalized (as e.g. in [15]), but to whether or not the formula is independent of the agent's state of mind.

([10]) See [20, ch.4] for frame completeness of, e.g., the partial system T (with both $\Box\varphi \vdash \varphi$ and $\varphi \vdash \Diamond\varphi$), and [9] for model completeness of single rules such as only $\Box\varphi \vdash \varphi$.

The third point is a very serious one. It may easily be overlooked, since one tends to focus on principles such as N and K. To make the point entirely explicit, I will shortly review the deductive system which corresponds to the purely partial semantics.

Omniscience regained

The core system corresponding to the purely partial semantics for modal logic consists of the rules of M^+:([11])($\varphi \dashv\vdash \psi$ abbreviates $\varphi \vdash \psi$ & $\psi \vdash \varphi$)

double negation	$\neg\neg\varphi \dashv\vdash \varphi$	
de Morgan's laws	$\neg(\varphi \wedge \psi) \dashv\vdash \neg\varphi \vee \neg\psi$	
	$\neg(\varphi \vee \psi) \dashv\vdash \neg\varphi \wedge \neg\psi$	
$\wedge$*-elimination*	$\varphi \wedge \psi \vdash \varphi$	$\varphi \wedge \psi \vdash \psi$
$\vee$*-introduction*	$\varphi \vdash \varphi \vee \psi$	$\psi \vdash \varphi \vee \psi$
$\vee$*-elimination*	if $\varphi, \rho \vdash \chi$ and $\psi, \rho \vdash \chi$ then $\varphi \vee \psi, \rho \vdash \chi$	
$\wedge$*-introduction*	if $\chi \vdash \varphi, \rho$ and $\chi \vdash \psi, \rho$ then $\chi \vdash \varphi \wedge \psi, \rho$	
ex falso	$\varphi \wedge \neg\varphi \vdash \psi$	
transitivity	if $\varphi \vdash \psi$ and $\psi \vdash \chi$ then $\varphi \vdash \chi$	
finiteness	$\Phi \vdash \Psi$ iff there are finite $\Phi' \subseteq \Phi$, $\Psi' \subseteq \Psi$ such that([12]) $\bigwedge\Phi' \vdash \bigvee\Psi'$	
dualization	$\hat{B}_i\neg\varphi \dashv\vdash \neg B_i\varphi$	$B_i\neg\varphi \dashv\vdash \neg\hat{B}_i\varphi$
C_r + *dual*	$B_i\varphi \wedge B_i\psi \vdash B_i(\varphi \wedge \psi)$	$\hat{B}_i(\varphi \vee \psi) \vdash \hat{B}_i\varphi \vee \hat{B}_i\psi$
I_r + *dual*	if $\varphi \vdash \psi$ then $B_i\varphi \vdash B_i\psi$ and $\hat{B}_i\varphi \vdash \hat{B}_i\psi$	
K_r + *dual*	$B_i(\varphi \vee \psi) \vdash \hat{B}_i\varphi \vee B_i\psi$	$B_i\varphi \wedge \hat{B}_i\psi \vdash \hat{B}_i(\varphi \wedge \psi)$
modal ex falso	$\hat{B}_i(\varphi \wedge \neg\varphi) \vdash \psi$	

C_r is the relativized counterpart of C, I_r relativizes I. Modulo the other rules, K_r amounts to $B_i(\varphi \to \psi) \vdash B_i\varphi \to B_i\psi$, which is a relativized form of K. Finally *modal ex falso* is the modal counterpart of the well-known *ex falso (sequitur quodlibet)* rule.

([11]) Cf. [9] for a concise sequential formulation. The text format is for the language $\mathcal{L}_{\neg,\wedge,\vee,\{B_i,\hat{B}_i\}_i}$; for the initial language $\mathcal{L}_{\neg,\wedge,\{B_i\}_i}$ de Morgan's laws and dualization are redundant, but e.g. K_r looks bad.

([12]) If $\Sigma = \{\varphi_1,\ldots,\varphi_n\}$ then the finite conjunction and disjunction over Σ are defined by $\bigwedge\Sigma = \varphi_1 \wedge \ldots \wedge \varphi_n$ and $\bigvee\Sigma = \varphi_1 \vee \ldots \vee \varphi_n$, respectively (omitting parenthesis, licensed by associativity).

3. *Combining partial and classical semantics*

One way to regain the cherished tautologies within a purely partial approach is to alter the notion of validity: change 'always true' (verification) to 'never false' (non-falsification).([13]) However, the problem is not really how to recover tautologies, but how to recover them without turning the logic into a normal modal system, in other words, how to avoid attributing overly strong properties to conscious belief and knowledge. The 'falsificational' approach in fact normalizes the logic and is therefore unfit. The same holds, *mutatis mutandis*, for the supervaluation approach.([14])

In this section a hybrid approach is considered: simply combine partial and classical semantics. Some alternatives will be discussed in section 4.

3.1 *A hybrid approach to truth*

One way to incorporate tautologies is to adopt a dual perspective on semantic states. Worlds as such are complete (something must be either true or false in the real world), but from the point of view of the agent they are partial: in general, she only has an opinion about part of the world. This idea, which can be traced back to essentially Fagin & Halpern's logic of awareness, can be implemented in partial semantics by distinguishing two kinds of truth relations. One is the bivalent truth relation $\models$, reflecting objective truth, the other the trivalent truth relation $\models\!\!\models$, reflecting subjective truth. Their opposites are non-truth ($\not\models$) and falsity ($=\!\!|$), respectively. In a given situation, a proposition is thus true or false with respect to $\models$, but may be undefined with respect to $\models\!\!\models$.

The definition of a hybrid model $M=\langle S, \overrightarrow{B}, V\rangle$ and the trivalent truth/falsity conditions (for $\models\!\!\models$ and $=\!\!|$) are as in standard partial semantics. In addition, there are the following truth conditions for $\models$:

$$s \models p \Leftrightarrow V(p, s)=1$$
$$s \models \neg\varphi \Leftrightarrow s \not\models \varphi$$
$$s \models \varphi \wedge \psi \Leftrightarrow s \models \varphi \;\&\; s \models \psi$$
$$s \models B_i\varphi \Leftrightarrow s \models\!\!\models B_i\varphi \Leftrightarrow \forall t \in B_i[s]: t \models\!\!\models \varphi$$

([13]) See [20], chapters 3 and 4.

([14]) See [5] and [20], pp. 84/5.

Validity is defined as overall classical truth:

$\models \varphi$ iff $M, s \models \varphi$ for all models M and situations s.

So, when checking the validity of a formula in this *hybrid semantics*, one starts with a two-valued evaluation and is dragged into the three-valued mode only by the modal operators. In other words, it is the doxastic operator which makes one change from objective to subjective truth, as it should be. Consequently, there is a partial 'internal' logic (i.e. within B_i) and a classical 'external' logic.

A remarkable effect of the hybrid semantics is the interpretation of $\hat{B}_i$. In fact there is already a point in how to paraphrase the dual of B_i in natural language. A dull but accurate account is simply 'it is not the case that i believes that not'. Assuming consistency of explicit belief one may even agree to Hintikka's translation 'it is compatible with i's belief that' in [7], but Lenzen's 'i considers it possible that' in [14] seems too strong. For recall that B_i stands for explicit (actual, active) belief, so denying that i believes $\neg\varphi$ may be correct when i is unaware of φ. Yet the purely partial semantics would lead to Lenzen's interpretation. The classical external denial is captured by the hybrid semantics, however:

$$s \models \hat{B}_i\varphi \Leftrightarrow \exists t \in \mathrm{B}_i[s]\colon t \not\models \varphi$$

which is very close to the original meaning, and Hintikka's paraphrase.

Let us now present some of the properties of the hybrid system. To start out, notice that the relation $\models$ is indeed bivalent. As in the purely partial semantics, $\models\!\!\models$ and $\not\models$ are mutually coherent: $s \models\!\!\models \varphi \Rightarrow s \not\models \varphi$). Coherence can be strengthened to a result that relates partial and classical truth:([15])

Proposition 3.1 (*propagation*) $s \models\!\!\models \varphi \Rightarrow s \models \varphi$.

In the purely partial approach external persistence (for extensions of the model based on the same frame) is an important property. Does it still hold for the hybrid system? A model $M = \langle S, R, V\rangle$ is extended to $M' = \langle S, R, V'\rangle$ ($M \sqsubset M'$) iff for all $s \in S$ and $p \in \mathcal{P}$: if $V(p, s) = 1$ or 0, then

([15]) Notice a similar result would not hold for the 4-valued approach: coherence is of vital importance.

$V(p, s) = V'(p, s)$. Then indeed $M, s \models \varphi \Rightarrow M', s \models \varphi$ and similarly for $\not\models$. Therefore, external persistence holds for $\models$ and $\not\models$, but obviously does not hold for $\vdash$.

The logic also possesses the revealing connection between (derived basic) awareness and explicit belief that was observed by Fagin & Halpern in [3], both for Levesque's system and their own logic of (special) awareness :([16]) if someone is aware of a tautology, he believes it. Let $\mathcal{L}^0$ be the set of strictly propositional formulas, free from modalities, i.e. $\mathcal{L}^0 = \mathcal{L}_{\neg,\wedge,\vee}$.

Proposition 3.2 If $\varphi \in \mathcal{L}^0$ and $\models \varphi$, then $\models A_i\varphi \rightarrow B_i\varphi$.

Proof: (by contraposition) Let φ be a propositional formula such that $\not\models A_i\varphi \rightarrow B_i\varphi$. Then there is a model $M = \langle S, \overrightarrow{B}, V\rangle$ and a state $s \in S$ such that $V(p_k, t) \in \{0,1\}$ for all $t \in B_i[s]$ and all atoms $p_1,\ldots, p_n$ occurring in φ. Moreover, $M, t' \not\models \varphi$ for some $t' \in B_i[s]$. By induction it follows that for each $\psi \in \mathcal{L}^0\{p_1,\ldots,p_n\}$ and $t \in B_i[s]$: $M, t \models \psi$ or $M, t \not\models \psi$. Consequently $M, t' \not\models \varphi$, and by propagation $M, t' \not\models \varphi$, thus $\not\models \varphi$. ∎

The 'core logic' of the hybrid semantics is a regular modal logic, where the introduction rules for B_i only operate on inferences that have not made use of the *tertium non datur* rules $\psi \vdash \varphi \vee \neg\varphi$ and $\psi \vdash B_i(\varphi \vee \neg\varphi)$. Put more positively, the inferential system contains classical propositional logic, the conjunction scheme C, and rule I restricted to strong consequence, together with their duals:

I_{M^+}	if $\varphi \vdash_{M^+} \psi$ then $\vdash B_i\varphi \rightarrow B_i\psi$ and $\vdash \hat{B}_i\varphi \rightarrow \hat{B}_i\psi$
C	$\vdash (B_i\varphi \wedge B_i\psi) \rightarrow B_i(\varphi \wedge \psi)$
C *dual*	$\vdash \hat{B}_i(\varphi \vee \psi) \rightarrow (\hat{B}_i\varphi \vee \hat{B}_i\psi)$

Further properties of explicit belief are triggered by suitable conditions on the frame. In general, the framework of hybrid truth is a rather flexible one, like standard possible world semantics. A remarkable exception to this is the 5 scheme of *negative introspection*. The corresponding condition is extremely strong: accessibility has to be both *Euclidean* and lead to total situations, i.e. states such that every formula is either supported or rejected.

([16]) See [3], [15] and sections 4.1 and 4.3 below.

This heavy constraint turns the logic for B_i into a normal modal system (K5), which is unfit for the enterprise. Perhaps the right conclusion from this is that requiring negative introspection for explicit belief is very much nonsensical, and one should be punished for such a sin. But then one may argue that *positive introspection* is almost equally counterintuitive for explicit belief. Perhaps the converse schemata([17]) are preferable to the usual forms of introspection.

$4_c \quad \vdash B_iB_i\varphi \rightarrow B_i\varphi$
$5_c \quad \vdash B_i\neg B_i\varphi \rightarrow \neg B_i\varphi$

Axiom 4_c can be motivated by observing that it seems to be impossible to *explicitly* believe that you explicitly believe one thing or other without explicitly believing it; in Hintikka's terms, this would be a 'self-defeating' activity. A similar consideration can motivate 5_c. Notice that a somewhat stronger notion of belief, say, conviction([18]) is modelled here; perhaps for extremely uncertain belief this need not hold. To conclude, the hybrid semantics is partly successful. A number of problems is solved more or less automatically. In particular, there now are tautologies, but no N-omniscience. However, I have to confess that some of the properties attributed to belief in this way are less fortunate. One of the main points is the persisting K-omniscience: in the hybrid system people are forced to believe the conclusions derivable within their own belief. I will turn to this problem in the next section.

3.2 *The elimination of residual omniscience*

Despite the relative success of the hybrid approach, there still are some forms of omniscience. This is sometimes argued to be inevitable: if the logic is to contain more than just the modal substitution instances of the classical propositional calculus, then these extra principles would lead to new belief or knowledge, i.e. create omniscience. This argument is not conclusive, however. The point is simply that the derived belief may be intuitively acceptable; if not, one arrives at a form of omniscience that should be

([17]) Called 'extraspection' schemata in [8].

([18]) Cf. [20], chapter 5.

exorcized. For example, C_c seems fully acceptable for explicit belief. So, in my view, every belief which follows from this principle qualifies.

On the other hand, apart from K, the (restricted) I-rule should also be eliminated, since it implies $\vdash B_i\varphi \rightarrow B_i(\varphi \vee \psi)$ which is unacceptable for actual belief. The general strategy to solve this will be to require awareness. Adding awareness to possibly unconscious belief turns it into conscious belief. In a slogan:

CONSCIOUS BELIEF = BELIEF + AWARENESS

Before proposing my actual solution, I will shortly inspect another candidate for eliminating residual omniscience.([19])

Recycling awareness

A first idea to eliminate rule I is to use the awareness created by explicit belief itself, in other words, not to waste awareness. In [12], Levesque & Lakemeyer propose to use the derived awareness $B(p \vee \neg p)$ of all the atoms contained in the formula, as defined in section 2. Then 'actual belief' is introduced by:

$$B_i^A\varphi = B_i\varphi \wedge A_i\varphi$$

Since awareness of φ need not involve awareness of $\varphi \vee \psi$, disjunctive weakening does not hold for B_i^A in general. For example, $B_i^A p \rightarrow B_i^A(p \vee q)$ is not valid. Yet, there are almost equally dubious results which are still validated by the augmented system, e.g. $\vdash B_i^A p \rightarrow B_i^A(p \vee \neg p)$. Moreover, it is easily verified that the other problematic principles, C and K, still hold in this approach.

Superimposing awareness sieves

A more radical strategy is to use so-called awareness sieves: use a syntactic filter to single out the conscious belief from the general beliefs. This mechanism enables the control of *conscious belief*. Within the area of possible world semantics, such a flexible framework is outlined in [21], essentially

([19]) In section 4 the possibility of tolerating inconsistencies as a strategy to eliminate LO is discussed.

generalizing Fagin & Halpern's logic of 'general awareness' without their structural conditions (*seriality*, *transitivity*, *Euclidicity*). This so-called *sieve semantics* turned out to be a very general and flexible framework for weak modal logics. The idea is now to superimpose the awareness sieve on the hybrid semantics.

A partial sieve model $M = \langle S, \overrightarrow{B}, \overrightarrow{A}, V\rangle$ with hybrid evaluation is defined as follows. The trivalent truth/falsity relations ($\models$ and $\nRightarrow$) and the bivalent truth relation ($\models$) are defined as in the hybrid semantics from section 3.1 (for $\mathcal{L}_{\overrightarrow{B}}$). There are additional clauses for the *conscious belief* operators C_i, which are interpreted by means of awareness sieves and accessibility relations. For each i and s the awareness sieve is a subset of formulas, i.e. $A_i(s) \subseteq \mathcal{L}_{\neg,\wedge,\overrightarrow{B},\overrightarrow{C}}$. The additional clauses for C_i are:([20])

$$\begin{aligned} s \models C_i\varphi \Leftrightarrow\ & s \models C_i\varphi \Leftrightarrow s \models B_i\varphi \ \&\ \varphi \in A_i(s) \\ & s \nRightarrow C_i\varphi \Leftrightarrow s \nRightarrow B_i\varphi \text{ or } \varphi \notin A_i(s) \end{aligned}$$

Validity is still defined as universal bivalent truth. Here are a number of observations which indicate that the 'hybrid sieve' semantics fulfils the requirements of a proper partial interpretation:

- as before, *propagation* holds ($s \models \varphi \Rightarrow s \models \varphi$), and therefore also *coherence* ($s \models \varphi \Rightarrow s \nRightarrow \varphi$);

- another useful property, also exhibited by the previous partial logics, is *inherited classicality*([21]): if V is bivalent for *all* situations, then for every s and φ, $s \models \varphi \Leftrightarrow s \models \varphi$;

- the semantics is still *externally persistent*: extension of the valuation for a fixed frame (to which the awareness sieve belongs) implies preservation of trivalent truth and falsity.

Is this semantics as general and flexible as total sieve semantics? In other words, can every logic for C_i that extends classical propositional logic still be captured? This question is answered in the affirmative.

([20]) Perhaps the falsity clause is not the most intuitive one after all. Closer to the idea of superposition seems $s \nRightarrow C_i\varphi \Leftrightarrow s \nRightarrow B_i\varphi \ \&\ \varphi \in A_i(s)$. This alternative semantics, which is not classically closed, will be discussed elsewhere.

([21]) See [20], pp. 66,92, and cf. 'reliability' in [13, p.18].

Theorem 3.3 *Hybrid sieve semantics for the restricted language* $\mathcal{L}_{\bar{c}}$ *is sound and complete for every modal system extending* pL.

Proof: Using inherited classicality and additional operators for implicit belief L_i underlying conscious belief (see section 4.1), one can reduce the theorem to a similar result for total sieve semantics.([22]) ■

The obvious generalization to the full language is:

Conjecture 3.4 *Hybrid sieve semantics is sound and complete for every modal system containing* pL, *the core logic of the hybrid system for* B_i *and* $\vdash C_i\varphi \rightarrow B_i\varphi$.

Anyway, every modal logic for conscious belief that contains tautologies can be captured by a suitable class of models. This notion of 'modal logic' is very wide: for example even the principle of extensionality (E) need not hold. Also, the notion of completeness is not very restricted. As in normal modal logic, one may be more interested in what is called frame completeness. If the sieve counts as part of the frame, then one can find corresponding conditions for intuitively valid principles, such as D* and C_c.([23])

- C and C_c hold automatically for B_i.

- D (or, equivalently, D*) for B_i, i.e. $\models B_i\varphi \rightarrow \hat{B}_i\varphi$, is captured by the condition that B_i is serial: $\forall s\, \exists t$: $sB_i\, t$, i.e. $B_i[s] \neq \varnothing$.

- C_c for C_i, i.e. $\models C_i(\varphi \wedge \psi) \rightarrow (C_i\varphi \wedge C_i\psi)$, is captured by the condition that $\varphi \wedge \psi \in A_i(s) \Rightarrow \varphi, \psi \in A_i(s)$.

- D for C_i, i.e. $\models C_i\varphi \rightarrow \hat{C}_i\varphi$, corresponds to the condition of seriality of B_i from inconsistent awareness: $\exists\varphi$: $\varphi, \neg\varphi \in A_i(s) \Rightarrow B_i[s] \neq \varnothing$

- D* for C_i, i.e. $\models \neg C_i(\varphi \wedge \neg\varphi)$ corresponds to the condition of seriality of B_i from contradictory awareness: $\exists\varphi$: $\varphi \wedge \neg\varphi \in A_i(s) \Rightarrow B_i[s] \neq \varnothing$

([22]) See [20], corollary 6.2. The proof of theorem 7.2 ibid. was erroneously applied to conjecture 3.4.

([23]) Cf. [23] for correspondence theory of total sieve semantics.

for all $s \in S$

These axioms show an interesting interplay. For example, the different consistency axioms for C_i follow straightforward by the correspondence properties from those for B_i. Also note that C_c for conscious belief, which is perfectly acceptable, implies that the D axiom for C_i is at least as strong as the D* axiom $\neg C_i(\varphi \wedge \neg\varphi)$. The other conjunction property C, that is $\vdash (C_i\varphi \wedge C_i\psi) \rightarrow C_i(\varphi \wedge \psi)$, is much less obvious.

D* is fully acceptable for conscious belief: one never (not even in dialectic philosophy) consciously believes a contradiction. Of course, one may become aware of an inconsistency within one's belief, but this involves *two* beliefs which are relatively inconsistent, rather than one. It is somewhat less clear whether D holds for conscious belief. Even in the above case of realizing an inconsistency in one's belief, at least one of the beliefs involved will presumably have been implicit. Then D would also be acceptable. The easiest implementation of this is by requiring consistency of B_i. This is in accordance with the partial semantics of B_i: the operator does not stand for 'implicit belief' but for 'derivable from explicit belief'. Then another operator L_i for 'implicit belief', underlying B_i, may allow inconsistencies. I will return to this issue in section 4.1.

In summary, there are indications to have a doxastic logic with at least three layers, and corresponding operators C_i, B_i and L_i of decreasing degrees of awareness ordered by $C_i \Rightarrow B_i \Rightarrow L_i$. On the C_i level all forms of LO can be eliminated and the logic is very weak. The intermediate B_i level already eliminates some omniscience, but should not allow inconsistencies. This level is also needed to explain certain pragmatic phenomena connected to natural language.([24]) Then, deep under the sea of awareness there is a bottom of implicit belief L_i. Although usually *less* explicit belief is connected to a *more* idealized (i.e. stronger) logic, I do not see any *a priori* reason that this has to be the case. For example, one may implicitly believe a contradiction, without being aware of it.

The resulting framework is very powerful. The drawback of this is that considerable indeterminacy of the locus of explanation is introduced in the semantics. Conditions for acceptable principles, and avoidance of others can now be triggered by means of no less than three interacting dimensions: the accessibility relations B_i, the awareness sieve functions A_i and partiality of

([24]) Such as Moore's paradox, see [20], chapter 5.

valuation (A_i and B_i for capturing acceptable principles, A_i and partiality for avoiding unacceptable ones). As indicated above, the interplay of these dimensions is not trivial, and certainly interesting.

4. *Alternatives: comparison and discussion*

In this section the hybrid systems just developed are compared to rival theories on consciousness and omniscience. The focus is on theories with a clear semantic component.([25])

4.1 *Special awareness logic vs the hybrid system*

In many respects the hybrid system (without the awareness sieves) is similar to the logic of awareness of Fagin & Halpern, here called the 'special awareness logic' (SAL) to avoid confusion with their logic of general awareness (GAL).([26]) Both SAL and the hybrid system are characterized by a twofold perspective on truth: total as well as partial. Are the two approaches equivalent? Indeed they share a large number of properties. For example, both have the rule I_{M^+}, axiom scheme C, and D is modelled by seriality, whereas for serial models a strong possibility rule qualifies:

$$\text{P}^* \qquad \models \varphi \Rightarrow \models \hat{B}_i B^* \varphi,$$

where B^* abbreviates a sequence of operators from $\{B_1,\ldots,B_m\}$. Proposition 3.2 also holds for both systems. Although one has to impose, apart from transitivity, the additional condition of upward monotonicity on the models to capture positive introspection (4), both systems collapse when requiring negative introspection (5).

One striking difference between SAL and the hybrid system is that the language of the former logic also contains operators L_i for *implicit* belief, which is a very idealized notion in [3]: the modal system of L_i is KD45, also known as 'weak S5'.([27]) But of course, addition of these operators to the

([25]) Some other alternatives as proposed in, for example, [3] (local reasoning) and [6] (morphological awareness), as well as the method of expanding the language with extra connectives are discussed in [20], but omitted here for reasons of space.

([26]) See section 4 in [3] and section 6.3.2 in [20] for a discussion of SAL.

([27]) So, as a corollary of proposition 3.2 note that $\models L_i\varphi \Rightarrow \models A_i\varphi \rightarrow B_i\varphi$.

hybrid system is feasible. The model-theoretic counterpart of L_i is an accessibility relation L_i such that $\mathrm{L}_i \subseteq \mathrm{B}_i$. If one wants the logics to be similar with respect to L_i, L_i has to be subjected to the same conditions: it should be serial, transitive and Euclidean. The new evaluation conditions for L_i are:

$$\mathrm{s} \models L_i\varphi \Leftrightarrow \forall t \in \mathrm{L}_i[s] : \mathrm{t} \models \varphi$$
$$\mathrm{s} \not\models L_i\varphi \Leftrightarrow \exists t \in \mathrm{L}_i[s] : \mathrm{t} \not\models \varphi$$
$$\mathrm{s} \models\!\!\!\!\sim L_i\varphi \Leftrightarrow \forall t \in \mathrm{L}_i[s] : \mathrm{t} \models\!\!\!\!\sim \varphi$$

Then the two approaches are equivalent with respect to implicit belief. Given the general similarity, it may not be surprising that every hybrid validity for the full language (including B_i) is also provably an SAL validity. This can be shown by a truth preserving transformation of SAL models into hybrid models. Let $M = \langle W, \overrightarrow{R}, \overrightarrow{\mathrm{A}}, V\rangle$ be an SAL model. A corresponding hybrid model $M' = \langle S, \overrightarrow{\mathrm{B}}, \overrightarrow{\mathrm{L}}, V'\rangle$ can be constructed: (Ψ is an arbitrary subset of $\mathcal{P}$)

- $S = W \times \wp(\mathcal{P})$ ($\langle w, \Psi\rangle$ is written as w_Ψ)
- $w_\Psi \mathrm{B}_i v_{\Psi \cap \mathrm{A}_i(w)} \Leftrightarrow wR_iv$
- $w_\Psi \mathrm{L}_i\, v_\Psi \Leftrightarrow wR_iv$
- $V'(p, w_\Psi) = V(p, w)$ if $p \in \Psi$ (else undefined)

Preservation of partial and total truth can be derived by induction on the structure of φ.

Lemma 4.1

(i) $M, w \models^\Psi \varphi \Leftrightarrow M', w_\Psi \models \varphi$
(ii) $M, w \not\models^\Psi \varphi \Leftrightarrow M', w_\Psi \not\models \varphi$
(iii) $M, w \models \varphi \Leftrightarrow M', w_\mathcal{P} \models \varphi$

Notice the lemma does not claim full equivalence of the models involved; in fact $M', w_\Psi \models \varphi$ may have no counterpart in M if $\Psi \subset \mathcal{P}$. By means of the last lemma one easily proves the following theorem.

Theorem 4.2 *Every formula valid with respect to the hybrid partial semantics is also valid with respect to* SAL *semantics.*

Does the converse of this theorem hold as well? No, it does not! The main

reason is that the hybrid semantics is more permissive. Unlike the SAL models it does not transfer the set of defined propositional atoms from a situation to its doxastic alternatives. This manifests itself in a formula such as:

$$B_iB_j(p \vee \neg p) \rightarrow B_i(p \vee \neg p).$$

This formula is valid in SAL; by invoking the notion of derived awareness, even $\models B_iA_j\varphi \rightarrow A_i\varphi$ holds in SAL. These formulas are invalid in the hybrid system. It is not entirely clear whether one should desire the validity of the displayed formula —it may depend on the notion of awareness involved. In all, despite these minor differences SAL and the hybrid semantics are very similar. I believe the hybrid models to be more natural, however, since there is no need to specify more of the content of an alternative doxastic state than the agent is aware of. Although I am not claiming 'psychological reality' for any of the proposals made here, it is clear, I think, which approach is more intuitive in this respect.

4.2 *General awareness logic vs the hybrid sieve system*

It follows from a very general completeness theorem in [20] or [21] on the one hand and theorem 3.3 on the other that the total sieve semantics of GAL and the hybrid sieve semantics are extensionally equivalent, in the sense that the two approaches model the same logics for the restricted language $\mathcal{L}_{\overline{C}}$: every modal logic extending pL is characterized by a class of such models.

The two systems under inspection are different in that GAL contains operators for implicit belief and awareness which are absent in the hybrid sieve system. Moreover, the B_i-operator of the former approach corresponds to the C_i-operator of the latter. So there is a clear gap between the two specific systems. Is it possible to bridge the gap?

First, addition of L_i operators and accessibility relations L_i goes as in the simple hybrid semantics. Second, the awareness operators A_i could also be added, with the following simple truth/falsity conditions:

$$\mathrm{s} \models A_i\varphi \Leftrightarrow \mathrm{s} \models A_i\varphi \Leftrightarrow \varphi \in \mathrm{A}_i(s)$$
$$\mathrm{s} \not\models A_i\varphi \Leftrightarrow \varphi \notin \mathrm{A}_i(s)$$

Then C_i could be redefined by $C_i\,\varphi = B_i\varphi \wedge A_i\varphi$. I did not take this road,

since the addition of new awareness operators A_i would lead to unacceptable interaction with the B_i-operators: $B_i(A_i\varphi \vee \neg A_i\varphi)$ would be validated, which seems intuitively wrong. It is technically possible to avoid bivalence of A_i by partializing the awareness sieves, i.e. duplicate the sieve function A_i into the pair A_i^+, A_i^-, where $A_i^+(s) \subseteq \mathcal{L}$ and $A_i^-(s) \subseteq \mathcal{L}$ and give appropriately modified truth/falsity conditions for the awareness operator. I have no intuitions about 'negative awareness' different from lack of awareness, however. Therefore the language of GAL is restricted to $\mathcal{L}_{\overline{B}\overline{L}}$.(28)

Third, if C_i is to correspond to B_i in the logic of general awareness, the operator B_i from the hybrid sieve approach needs a counterpart in general awareness logic. It is possible to add to GAL such intermediary operators B_i' for each i to the syntax, and awareness sieves to the semantics such that B_i' is interpreted by means of A_i'. Then a suitable transformation of hybrid sieve models into GAL models is feasible. Despite this technical equivalence there are differences in underlying intuitions, especially with regard to the way in which unacceptable principles are circumvented: part of the awareness which deals with knowledge of the objects and notions involved, i.e. with the conceptual information present in the agent, is accounted for by means of partiality in the hybrid sieve system. Another type of awareness, corresponding to what the agent actually thinks of at a certain moment, is accounted for by means of the awareness sieve. The awareness sieve is only effective on propositions which are (partially) true, and this accords with the intuition that one needs basic conceptual knowledge before actually being aware of something.

4.3 *Restricted validity and the four-valued approach*

Within the area of partial semantics for actual belief, Hector Levesque has introduced at least two important ideas. One is to allow *incoherent* situations, to be be studied later on in this section. Until further notice situations will be coherent. The other main idea is to *restrict* the set of situations to which the validity test applies.(29)

(28) To obtain completeness for GAL, the axiom $B_i\varphi \leftrightarrow L_i\varphi \wedge A_i\varphi$ has to be replaced by $B_i\varphi \rightarrow L_i\varphi$.

(29) For total semantics the idea of restricted validity can be found in, for example, [10] and [18].

Restricted validity

Possible world validity amounts to:

$$\models \varphi \quad \text{iff} \quad M, w \models \varphi \text{ for all } M \text{ and possible worlds } w.$$

Perhaps surprisingly, this does not produce all the substitution instances of classical tautologies: for example $\not\models Bp \vee \neg Bp$, witness a simple counter-example with one possible world and one *empty* alternative. The point is that bivalence of propositional formulas does not imply bivalence of modal formulas. An immediate solution to this problem is to alter either the truth or the falsity condition of B_i, making modal formulas bivalent in all situations. Levesque chooses to modify the falsity clause:([30])

$$s \models B_i\varphi \Leftrightarrow s \not\models B_i\varphi \Leftrightarrow \exists t \in \mathrm{B}_i[s]\colon t \not\models \varphi$$

Yet in the context of partial semantics this falsity clause is an anomaly: apart from being counterintuitive it deprives the logic of the possibility to distinguish mere absence of belief from disbelief. In the latter case there is an accessible situation in which φ is false, reflecting a higher degree of awareness than in the former case, where φ is not known to be true in some accessible situation.

Within a partial context invocation of possible worlds for validity also seems a drastic move, which was criticized in [3]:

> *While restricting to complete situations [possible worlds, ET] ensures that all propositionally valid formulas continue to be valid in Levesque's logic, it seems inconsistent with the philosophy of looking at situations.* [3, p.48]

Although seemingly right, this judgement turns out to be rather harsh, since one can alter the validity type.

Proposition 4.3 *A formula is valid in the coherent variant of Levesque's semantics iff it is never false.*

([30]) [15] deviates from the standard approach in other ways too, for example the syntax is constrained to one agent and formulas of modal depth 1, and the models contain a set of doxastic alternatives rather than accessibility relations.

Proof: Cf. [19], proposition 7, where it was shown that for Levesque's original 4-valued semantics, verification on possible worlds amounts to non-falsification on coherent situations. ∎

What worries me, however, is not Levesque's notion of validity, but his solution to the lurking absence of bivalence of modal formulas in possible worlds: the modified falsity clause disturbs the nice uniform appearance of standard truth and falsity conditions. More importantly, the solution seems rather *ad hoc* and *brute force*: it makes fully modalized formulas bivalent in *every* situation, not just in possible worlds. This produces counterintuitive validities such as $B_i(B_jp \vee \neg B_jp)$. Now formulas of modal degree 2 or more are forbidden in Levesque's syntax, yet for the many agents case these are the formulas of interest.

Lakemeyer in [11] adapts Levesque's semantics by essentially splitting the accessibility relation into one relation dealing with belief and one dealing with disbelief. The distinction is motivated by the alleged "different modes of thinking when it comes to positive versus negative beliefs" [11, p. 403]. Although this move is indeed in the spirit of partial semantics, it is difficult to grasp the intuition behind split accessibility. The point is that any accessible state is indiscernible from the initial state according to someone's belief. Now splitting the accessibility relation also implies different criteria of 'sameness' among worlds. Also, though technically possible and interesting, the twin relations may lead to a less efficient computation, since two classes of alternatives have to be inspected in order to determine (absence of) belief.

Anyway, in Lakemeyer's proposal two relations B_i and B'_i are used to model B_i. The truth condition for B_i is as before, and the falsity condition for B_i is changed into:

$$s \nVdash B_i\varphi \Leftrightarrow \exists t \in \mathrm{B}'_i[s]: t \nVdash \varphi$$

To guarantee bivalence of modal formulas on possible worlds, Lakemeyer imposes the condition that B_i and B'_i coincide as seen from possible worlds: $\mathrm{B}_i[w] = \mathrm{B}'_i[w]$, i.e. $w\mathrm{B}_i s \Leftrightarrow w\mathrm{B}'_i s$ for all worlds w and situations s.([31])

([31]) The requirement is used by Lakemeyer to restrict the extension of 'possible world', rather than the set of admissible frames, as is done here.

The modified framework is rich enough to capture bivalence on worlds([32]) without making modal formulas universally bivalent. Consequently the formula $\neg B_i(B_j p \vee \neg B_j p)$ is now satisfiable. As far as this goes Lakemeyer's adaptation is technically successful.

Remaining problems are related to the presence of additional operators L_i for 'agent i implicitly believes'. L_i is interpreted (bivalently) by restricting the set of B_i-alternatives to those that are worlds. The conditions of *transitivity* and *Euclidicity* (largely restricted to possible worlds) turn the logic of L_i into the normal modal system K45. The additional operator may however cause trouble with respect to formulas such as $B_i(L_i p \vee \neg L_i p)$, the validity of which is highly undesirable. Like Levesque, Lakemeyer removes this problem by a syntactic constraint, now to the effect that no L_i may occur in the scope of a B_j.

Apart from this drawback Levesque's and Lakemeyer's proposals for the falsity clause of B_i fail to capture the right intuition. I agree with Fagin & Halpern in [3, p. 51] that the standard falsity clause is intuitively the correct one. Moreover, the problem of absence of modal bivalence is easily solved in the hybrid approach, without the need to modify the falsity condition beyond intuition. Therefore it is no coincidence that the counterintuitive $B_i(B_i p \vee \neg B_i p)$ and $B_i(L_i p \vee \neg L_i p)$, which were problematic for Levesque and Lakemeyer, respectively, are invalid in the hybrid system.

The four-valued approach

Now turn to the other innovation of Levesque: include incoherent situations. The fourth truth value *overdefined* represents the state of affairs in which an atomic proposition is both true and false, in other words, in which the situation contains inconsistent information. It is convenient to formulate a four-valued valuation as a multi-valued function from situated atoms to classical truth values, i.e. V: $\mathcal{P} \times S \longrightarrow \wp\{0,1\}$.([33])

The multiple-valued valuation function requires a few modifications in,

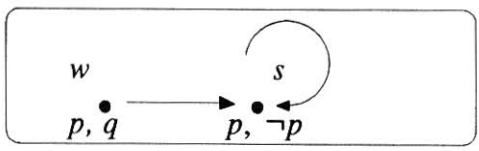
([32]) Muskens [17] proposes a similar solution, with split accessibility and restricting validity in his notion of weak consequence by meaning postulates urging the initial world to be classical. Unlike Levesque and Lakemeyer however, Muskens' falsity condition is standard, which implies that $B\varphi \vee \neg B\varphi$ is invalid in his semantics. See [22] for further exposition.

([33]) Instead of V, Levesque uses two functions T (for 'truth') and F (for 'falsity') from $\mathcal{P}$ to $\wp(S)$, which are related to the text format by: $s \in T(p) \Leftrightarrow 1 \in V(p, s)$ and $s \in F(p) \Leftrightarrow 0 \in V(p, s)$.

for example, the basic evaluation clauses.

$$s \models p \Leftrightarrow 1 \in V(p, s) \qquad s \not\models p \Leftrightarrow 0 \in V(p, s)$$

Incoherent doxastic alternatives eliminate K: belief need not be closed under implication, since both the believed antecedent and its negation may be verified at a situation in a model. So, a simple counterexample to $B(p \to q) \to (Bp \to Bq)$ is:([34])



Interestingly, the other combination scheme C is valid in Levesque's quadrivalent semantics. Therefore (see section 3.2), rule I cannot be valid in general. But rule I does hold with respect to relevant entailments, i.e. strong 4-valued consequence:

$$\varphi \models \psi \Rightarrow \models B\varphi \to B\psi$$

The restriction of I to relevance logic rL is defined by:

$$\mathrm{I}_{\mathrm{rL}} \qquad \varphi \vdash_{\mathrm{rL}} \psi \Rightarrow \vdash B\varphi \to B\psi.$$

For the language $\mathscr{L}^1$ (modal formulas of depth at most 1) axiom scheme C and rule I restricted to rL jointly constitute a complete axiomatization:([35])

([34]) The model has the double, dyadic accessibility of [11] with $\mathrm{B} = \mathrm{B}^{\smile} = \{\langle w, s\rangle, \langle s, s\rangle\}$, rather than the single, monadic accessibility of [15]. The 'monadic' counterexample is similar, cf. [19], p. 573.

([35]) $\mathrm{I}_{\mathrm{rL}}\mathrm{C}$ may lead to several equivalent axiomatizations. The concrete systems given in [15] contain a number of redundancies. For example, to one system, formed by applying $\mathrm{I}_{\mathrm{rL}}\mathrm{C}$ to the natural deduction style introduction/elimination rules for rL, derivable schemes for *commutativity*, *associativity* and *distributivity* of $\wedge$ and $\vee$ are added as axioms.

Theorem 4.4 *The modal system for explicit belief with constrained syntax and Levesque's 4-valued semantics is given by (the modal instantiations of)* pL, *Modus Ponens and* $I_{rL}C$.

Despite this nice completeness result and the fact that Levesque's system avoids N, K and I omniscience, his solution is unsatisfactory.

Whereas N is running against counterexamples with *partial* alternatives, for K one needs *incoherent* alternatives, and rule I can be eliminated by either partial or incoherent alternatives. To me only the arguments depending on partiality are fully convincing: lack of awareness is felicitously represented by the absence of a classical truth value. The arguments depending on incoherence are much less conclusive. Explicitly believing in inconsistent states seems counterintuitive. To quote Fagin & Halpern:

> [...] *to the extent that* B *is viewed as the set of situations the agent considers possible, it seems unreasonable to allow incoherent situations.* [3, p.47]

As far as K is concerned, one may even speculate that Levesque's explanation stems from a peculiarity of the English language: the ambiguity of the word *inconsistent*. One of its meanings is related to the existence of contradictions (i.e. incoherence), another to the fact that people may not draw (the right) conclusions from their beliefs (i.e. K-failure). Moreover, though the principles K and I are not generally valid, a number of valid formulas related to these principles are equally unacceptable, for example:

- $\models B\varphi \rightarrow B(\varphi \vee \psi)$
- $\models (B\varphi \wedge B(\varphi \rightarrow \psi)) \rightarrow B(\psi \vee (\varphi \wedge \neg\varphi))$ [3, p.46]
- $\models B(B\varphi \vee \neg B\varphi)$ (for Levesque's semantics)

To conclude, note that, despite the initial goal of SAL to simulate Levesque's logic in augmented possible worlds semantics, the logic is quite different from both SAL and the hybrid system. For example, unlike the collapse of the latter two systems when 5 is required for B_i, this property simply holds in Levesque's system with free syntax. Therefore these approaches are incompatible.

4.4 *Non-standard semantics vs the four-valued approach*

Both Lakemeyer's and Muskens' approach are related to a *bivalent* non-standard semantics suggested by Fagin, Halpern & Vardi in [4], where models contain twin worlds, i.e. every world w has a unique (negative) counterpart w^*. Technically * is a self-inverse operation on worlds.([36]) Apart from * the models are classical Kripke structures. So, let $M = \langle W, \vec{R}, V, ^* \rangle$ be a non-standard Kripke model. The truth conditions for atoms, $\wedge$ and B_i([37]) are standard-type. The negation clause is different, and so are some of the derived truth conditions:

- $M, w \models \neg\varphi$ iff $M, w^* \not\models \varphi$;
- $M, w \models \varphi \rightarrow \psi$ iff $M, w^* \models \varphi \Rightarrow M, w \models \psi$;
- $M, w \models \hat{B}_i\varphi$ iff $M, v \models \varphi$ for some v such that $w^*R_iv^*$.

Fagin et al. notice that omniscience is avoided in a rather drastic way: there are no valid formulas in this non-standard semantics. This is reminiscent of the situation in the standard partial approach. In fact it is noticed [4, p.46,47] that the non-standard semantics and that of Lakemeyer's are locally equivalent in the sense that for every non-standard model N and world w there is a partial (Lakemeyer-style) model M and situation s such that $N, w \models \varphi \Leftrightarrow M, s \models \varphi$ and $N, w^* \not\models \varphi \Leftrightarrow M, s \not\models \varphi$, and *vice versa* from partial models to non-standard models.

This does not guarantee full equivalence of these two approaches, for Lakemeyer's semantics validates a large number of formulas, such as the classical tautologies. This difference is caused by the notion of restricted validity in the Lakemeyer-Levesque approach. Now Fagin et al. could have captured tautologies by restricting validity to *standard worlds*, i.e. worlds w such that $w^* = w$. Perhaps surprisingly, they do not take this route, but introduce a new implication, named 'strong implication' (here symbolized by) $\rightharpoonup$, which formalizes strong consequence. The new implication has the following interpretation:

- $M, w \models \varphi \rightharpoonup \psi$ iff $M, w \models \varphi \Rightarrow M, w \models \psi$

([36]) I.e. $w^{**} = w$; *a fortiori*, * is a bijection. * is attributed to R. Routley, V. Routley and R.K. Meyer and stems from one way of doing the semantics of relevance logic.

([37]) In [4] knowledge is considered instead of belief, but this need not bother us.

The gain of adding $\rightharpoonup$ is that many formulas are valid again([38]), the cost that K-omniscience reenters, which is considered an advantage of the system.([39]) The proper partial interpretation of $\rightharpoonup$ can be obtained by means of the so-called *dual negation* ∂ from [20]. If $\varphi \rightharpoonup \psi = \partial\varphi \vee \psi$, then the truth conditions for ∂ imply those for $\rightharpoonup$:

$$s \models \partial\varphi \Leftrightarrow s \not\models \varphi \qquad s \nVdash \partial\varphi \Leftrightarrow s \nVdash \varphi$$
$$s \models \varphi \rightharpoonup \psi \Leftrightarrow (s \models \varphi \Rightarrow s \models \psi) \qquad s \nVdash \varphi \rightharpoonup \psi \Leftrightarrow (s \nVdash \varphi \,\&\, s \nVdash \psi)$$

It is now possible to extend the noticed correspondence of non-standard models with Lakemeyer's models to the full language (including $\rightharpoonup$). This equivalence also holds for Muskens' semantics with standard evaluation conditions. For such 'split models' one encounters the following clauses for B_i:

$$s \models B_i\varphi \Leftrightarrow \forall t \in \mathrm{B}_i[s]\colon t \models \varphi$$
$$s \nVdash B_i\varphi \Leftrightarrow \exists t \in \mathrm{B}'_i[s]\colon t \nVdash \varphi$$

Lemma 4.5 *For every (local) non-standard model* $\langle N, w\rangle$ *there is a truth equivalent split model* $\langle M, s\rangle$, *i.e.* $N, w \models \varphi \Leftrightarrow M, s \models \varphi$ *for all* $\varphi \in \mathcal{L}_{\neg,\wedge,\vec{B},\rightharpoonup}$. *And vice versa, for every split model* $\langle M, s\rangle$ *there is a truth equivalent non-standard model* $\langle N, w\rangle$.

For unrestricted (strong) validity this leads to equivalence of the logics.

Theorem 4.6 *A formula in* $\mathcal{L}_{\neg,\wedge,\vec{B},\rightharpoonup}$ *is non-standardly valid iff it is valid in split semantics.*

5. *Conclusion*

This article started out by reconsidering standard partial semantics as a candidate for an adequate logical description of awareness. Indeed this

([38]) Still many formulas, such as the *ex falso* and *tertium non datur* related $(p \wedge \neg p) \rightharpoonup q$ and $p \vee \neg p$ are invalid. Yet, contrary to a claim in [4, p.49], the 'distressing propositional tautology' $(p \rightharpoonup q) \vee (q \rightharpoonup p)$ is valid in this system.

([39]) So, being a perfect reasoner in relevance logic is rejected in [3], but implicitly supported in [4].

already provided a rather weak logic, as is required for conscious belief: a number of problematic principles, called logical omniscience, were circumvented. However, the standard partial approach suffers from two major problems: one is that intuitively valid forms such as $Bp \vee \neg Bp$ are also eliminated, the other is that most types of logical omniscience pop up again in relativized form, e.g. $Bp \Rightarrow B(p \vee q)$. So the *external* part of the logic for conscious belief had to be strengthened to essentially classical propositional logic, whereas the *internal* part had to be weakened.

The first problem (of the 'missing tautologies') was successfully solved in the hybrid system, combining total and partial truth. The second problem (of 'residual omniscience') requires a more demanding approach: add syntactic awareness sieves to the hybrid system. The resulting semantics is fully flexible in the sense that every modal logic which extends the classical propositional calculus can be modelled. For conscious belief the set of admissible models was constrained by imposing conditions on the frames.

Then the systems introduced were compared to several proposals made in the literature. Fagin & Halpern's special awareness logic was shown to be similar but not identical to the hybrid system: every SAL model can be transformed to an equivalent hybrid model, but not *vice versa*. The general awareness logic is similar to the hybrid sieve system, although the latter has the possibility to eliminate strong forms of omniscience merely by partiality, which is reflected in an additional operator expressing (direct consequences of) explicit belief. The four-valued approach of Levesque, later on improved by Lakemeyer, also restored the missing tautologies. Yet allowing incoherent situations and restricting validity to possible worlds leads to an exceptional falsity clause for explicit belief and, despite *ad hoc* syntax constraints, to undesirable omniscience, since the internal logic is closed under relevance logic. Then it was observed that the total non-standard semantics of Fagin et al. corresponds to four-valued semantics along the lines of Lakemeyer and Muskens. To summarize, I believe that the (hybrid) systems proposed here are superior, or at least not inferior to rival approaches. In fact, partiality provides a very natural explanation of why and how a great deal of logical omniscience is to be excluded.

Finally I want to counter two possible objections to the hybrid sieve system. The first is that to some logicians the logic provided by the hybrid sieve system goes well beyond what they would call a 'logic' proper, since, for example, the extensionality principle E is considered to be a prerequisite of any modal logic. To them my reply is that such a rigid conception of 'logic' excludes a logical treatment of actual belief, for even the weak

principle E, combined with classical propositional logic, leads to unacceptable consequences: consciously believing to be happy usually does not involve believing both to be happy and that Goldbach's conjecture is right or wrong. In fact, the latter inference is already blocked in the simple hybrid system, but in general a syntactic stipulation is called for. In other words, incorporation of psychological phenomena such as consciousness necessarily will lead to a 'logic' which hardly contains the cherished postulates of ordinary logic.

The second problem is connected to the richness of the hybrid sieve models. It was noticed that principles (axioms and rules) could be captured by conditions on three interacting components: accessibility, awareness sieves and partiality. So what is the exact locus of description and explanation of awareness phenomena? Although indeed different classes of hybrid sieve structures may model a principle (a situation not unusual in logic), there is an intrinsic order. Since partiality is 'for free', this option is the first in line, in particular for exclusion of invalid principles. Next comes accessibility, constrained by general conditions and possibly interacting with partiality conditions such as extension: together this accounts for purely logical (in)validity. Finally, as a last escape route, awareness sieves filter out conscious beliefs which have passed the earlier tests.

Tilburg University

BIBLIOGRAPHY

[1] van Benthem, J. - *A Manual of Intensional Logic*, CSLI Lecture Notes No. 1, 2nd edition, Stanford CA, 1988

[2] Chellas, B. - *Modal logic. An introduction*, Cambridge University Press, Cambridge UK, 1980

[3] Fagin, R. & J. Halpern - 'Belief, awareness and limited reasoning', *Artificial Intelligence 34*, pp. 39-76, 1988. Preliminary report in: IJCAI85, pp. 491-501, 1985

[4] Fagin R., J. Halpern & M. Vardi - 'A nonstandard approach to logical omniscience', in R. Parikh (ed.) *Proceedings of* TARK3 (Monterey CA), pp. 41-55, Morgan Kaufmann, 1990

[5] van Fraassen, B. - 'Singular terms, truth-value gaps, and free logic', *Journal of Philosophy* 63, pp. 481-495, 1966

[6] Gillet, E. & P. Gochet - 'La logique de la connaissance. Le problème de l'omniscience logique', *Dialectica 47 : 2-3*, pp. 143-171, 1993

[7] Hintikka, J. - *Knowledge and Belief. An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca, 1962

[8] van der Hoek, W. - 'Systems for knowledge and belief', in: J. van Eijck (ed.), *Logics in AI*, Proceedings JELIA'90, LNCS 478, Springer-Verlag, Berlin, 1991

[9] Jaspars, J. & E. Thijsse - 'Fundamentals of partial modal logic', in: P. Doherty & D. Driankov (eds.) *Partiality, Modality, Nonmonotonicity. Proceedings of the Workshop on Partial Semantics and Nonmonotonic Reasoning for Knowledge Representation, Linköping 1992*, to appear.

[10] Kripke, S. - 'Semantical analysis of modal logic II. Non-normal modal propositional calculi', in: Addison, Henkin & Tarski (eds.) *The theory of models*, pp. 206-220, North Holland, Amsterdam, 1965

[11] Lakemeyer, G. - 'Tractable meta-reasoning in propositional logics of belief', *Proceedings of* IJCAI87 (Milan), Morgan Kaufmann, Los Altos CA, pp. 402-408, 1987

[12] Lakemeyer, G. & H. Levesque - 'A Tractable Knowledge Representation Service with Full Introspection', in: M. Vardi (ed.) *Proceedings of* TARK2 (Monterey CA), pp. 145-159, Morgan Kaufmann, 1988

[13] Langholm, T. - *Partiality, Truth and Persistence*, CSLI Lecture Notes No. 15, Stanford CA, 1988

[14] Lenzen, W. - *Glauben, Wissen und Wahrscheinlichkeit*, Springer-Verlag, Vienna/New York, 1980

[15] Levesque, H. - 'A logic of implicit and explicit belief', *Proceedings of the National Conference on Artificial Intelligence*; expanded version as *FLAIR Technical Report* 32, Fairchild Laboratory of AI Research, Palo Alto CA, 1984

[16] Moses, Y. - 'Resource-bounded knowledge', in: M. Vardi (ed.) *Proceedings of* TARK2 (Monterey CA), pp. 261-275, Morgan Kaufmann, 1988

[17] Muskens, R. - *Meaning and Partiality*, dissertation University of Amsterdam, 1989

[18] Rantala, V. - 'Quantified modal logic: non-normal worlds and propositional attitudes', *Studia Logica 41*, pp. 41-65, 1982

[19] Thijsse, E. - 'Partial propositional and modal logic: the overall theory', M. Stokhof & L. Torenvliet (eds.) *Proceedings of the 7th Amsterdam Colloquium*, vol. 2, pp. 555-579, ITLI, Amsterdam, 1990. Extended version as ITK *Research Report 11*, Tilburg University, 1990

[20] Thijsse, E.G.C. - *Partial Logic and Knowledge Representation*, Eb-

uron Publishers, Delft (Holland), 1992

[21] Thijsse, E. - 'On total awareness logics (with special attention to monotonicity constraints and flexibility)', in: M. de Rijke (ed.) *Diamonds and Defaults*, pp. 309-347, Synthese Library vol. 229, Kluwer Academic Publishers, Dordrecht NL, 1993

[22] Thijsse, E. - *Partial semantics for awareness*, Tilburg University, forthcoming ITK Research Report, Tilburg NL, 1994

[23] Thijsse, E. & H. Wansing - 'A fugue on the themes of awareness logic and correspondence', in Bürckert & Nutt (eds.), *Proceedings KI-93*, DFKI report, Saarbrücken, 1993

[24] Wansing, H. - 'A general possible worlds framework for reasoning about knowledge and belief', *Studia Logica 49*, pp. 523-539, 1990