ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Applied Logic 5 (2007) 492–518

JOURNAL OF APPLIED LOGIC

www.elsevier.com/locate/jal

Only knowing with degrees of confidence

Arild Waaler [a,b,*], Johan W. Klüwer [c], Tore Langholm [b], Espen H. Lian [b]

[a] *Finnmark College*
[b] *Department of Informatics, University of Oslo, Norway*
[c] *Department of Philosophy, University of Oslo, Norway*

Available online 3 May 2006

Abstract

A new logic of belief (in the "only knowing" family) with confidence levels is presented. The logic allows a natural distinction between explicit and implicit belief representations, where the explicit form directly expresses its models. The explicit form can be found by applying a set of equivalence preserving rewriting rules to the implicit form. The rewriting process is performed entirely within the logic, on the object level, provided we supply an explicit formalization of the logical space. We prove that the problem of deciding whether there exists a consistent explicit form is Σ_2^p-complete, a complexity class to which many problems of nonmonotonic reasoning belong. The article also contains a conceptual analysis of basic notions like belief, co-belief and degrees of confidence.

Keywords: Logic of belief; Non-monotonic logic; Autoepistemic logic; Complexity; Prioritized defaults

1. Introduction

This article presents the propositional modal logic Æ, a logic in the "only knowing" family of logical systems pioneered by Levesque [8]. Compared to other propositional systems in this family the Æ system contributes on three levels: *conceptually* by the introduction of a richer set of epistemic concepts, both for the description of the system itself and for use in representation of common-sense patterns of reasoning within Æ; *logically* by being closed under uniform substitution and being axiomatized entirely at the object-level; by *an increased expressive power* which enables the representation of a certain sort of prioritized defaults.

The paper is structured as follows. Basic syntax and semantics are presented in Section 2; soundness, completeness, the finite model property and decidability are established by means of techniques which are standard in modal logic. Section 3 contains a philosophical interpretation of the main notions underlying the formal semantics of Æ, in particular the notions of belief, co-belief and degree of confidence. From Section 4 onwards we continue on the assumption that the language is finite. Finite languages allow us to define a particular formula which we shall call "the logical space", the effect of which is to turn the necessity operator into a true representation of logical necessity with no ambiguity as to whether Boolean propositions are necessary, contingent or impossible. Finite languages also have the capacity to represent exact belief states in a form which permits reduction into an explicit and particularly salient

[*] Corresponding author.
E-mail address: arild@ifi.uio.no (A. Waaler).

1570-8683/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jal.2006.03.007

form, a form which reflects its models transparently. This is the content of the Modal Reduction Theorem, the result which forms the center of this study. In Section 4 we exploit the model theory of Æ to give a semantical proof of the theorem.

In the rest of the paper we focus on computational aspects of the Modal Reduction Theorem. The first thing we show is that the logical space, the size of which is exponential in the number of propositional letters, can be implicitly defined and generated by means of meta-rules which generalize the special rule introduced by Levesque in his initial paper on "only knowing". In fact the propositional fragment of Levesque's logic is a special case of the system $Æ_\rho$ introduced in Section 5.1; more precisely it corresponds to adopting what we shall call the *maximal* logical space and having only one degree of confidence in the language. Just as in Levesque's system the proof checking problem for $Æ_\rho$ is NP-complete (and not linear). Moreover, these logics are not closed under uniform substitution, a property which holds for Æ and which traditionally has been used to distinguish a logic from a theory.

On the computational side there are, however, strong reasons for accepting $Æ_\rho$ as it gives an economic representation of the logical space, which otherwise is exponential in the number of propositional variables. The satisfiability problem for Levesque's system is Σ_2^p-complete [13,14], the same as for $Æ_\rho$ without confidence levels.

We finally identify a set of provable equivalences within Æ and prove that they are sufficient for proving the Modal Reduction Theorem syntactically. The procedure is examplified in Section 6 to give an account of the restricted class of prioritized supernormal defaults within Æ. For an encoding of full prioritized default logic, with a prescriptive interpretation of the preference relation, see [3].

2. The logic Æ

2.1. Syntax

The object language contains a stock of propositional letters, the constants $\top$ and $\bot$, and connectives $\neg$, $\vee$, $\wedge$, $\supset$ and $\equiv$. If Γ is a finite set of formulae, $\bigwedge\Gamma$ ($\bigvee\Gamma$) is the conjunction (disjunction) of the elements of Γ in arbitrary order. Modal operators are $\Box$ (necessity) and modalities B_k (belief) and C_k (co-belief) for each k in a finite index set I partially ordered by a relation $\preceq$. A *modal atom* is a formula of the form $B_k\varphi$, $C_k\varphi$, or $\Box\varphi$, and a *modal literal* is either a modal atom or its negation. We employ the following dual modalities: $\Diamond\varphi$ is $\neg\Box\neg\varphi$ (φ is possible), $b_k\varphi$ is $\neg B_k\neg\varphi$ (φ is compatible with belief at k), and c_k is $\neg C_k\neg\varphi$ (φ is compatible with co-belief at k).

The intuitions behind the syntactical operators shall be addressed at length in Section 3, but some explanation at this point facilitates the introduction of concepts that we shall use throughout the text. In short, $\Box$ is intended to express personal necessities. We shall read the formula $\Box(\varphi\supset\psi)$ as φ *entails* ψ and say that φ is *at least as strong as* ψ. The formula $\Diamond\varphi\wedge\Diamond\neg\varphi$ expresses that φ is *contingent*. Despite the fact that $\Box$ can easily be defined by means of the belief and co-belief operators we shall take it as primitive; this choice is made both to give the operator conceptual priority and to facilitate technical arguments.

The indices in I are intended to represent various degrees of *confidence* or *conviction*. $B_k\varphi$ expresses that φ is believed with degree of confidence k; $C_k\varphi$ expresses that φ is co-believed with degree of confidence k. In the following, we will sometimes abbreviate "degree of confidence" as *doc* and write that φ is believed *at k* when $B_k\varphi$ holds.

The belief and co-belief operators are complementary. $C_k\varphi$ expresses a notion of *caution*, and can generally be read as expressing that *at most* $\neg\varphi$ is believed with degree of confidence k; or, what amounts to the same, that $\neg\varphi$ is at least as strong as everything that is believed at k (see Section 3.2 for a caveat regarding this interpretation). The "all I know at k" expression $O_k\varphi$ is central; it abbreviates $B_k\varphi\wedge C_k\neg\varphi$, meaning that *precisely* φ is believed with *doc k*.

Some syntactical concepts: A formula φ is *completely modalized* if every occurrence of a propositional letter is within the scope of a modal operator. It is *purely Boolean* if it contains no occurrences of modal operators. The *propositional substitution operator* $[\cdot/\cdot]$ distributes over connectives and modalities in the obvious way. $\varphi[\psi_1/\psi_2]$ is φ with every subformula occurrence of ψ_1 substituted with ψ_2. A *tautology* is a substitution instance of a formula valid in classical propositional logic (such as $\Box\varphi\supset\Box\varphi$); if φ is a tautology of propositional logic we shall write $\vdash_{\mathrm{PL}}\varphi$. The logic Æ is defined as the least set that contains all tautologies, contains all instances of the following schemata for each $k\in I$:

$$
\begin{array}{llll}
\mathrm{Def}\Box: & \Box\varphi\equiv(B_k\varphi\wedge C_k\varphi) & T: & \Box\varphi\supset\varphi \\
K_B: & B_k(\varphi\supset\psi)\supset(B_k\varphi\supset B_k\psi) & K_C: & C_k(\varphi\supset\psi)\supset(C_k\varphi\supset C_k\psi)
\end{array}
$$

$B_\Box$:	$B_k\varphi \supset \Box B_k\varphi$	$C_\Box$:	$C_k\varphi \supset \Box C_k\varphi$
$\overline{B}_\Box$:	$\neg B_k\varphi \supset \Box\neg B_k\varphi$	$\overline{C}_\Box$:	$\neg C_k\varphi \supset \Box\neg C_k\varphi$
P_B:	$B_i\varphi \supset B_k\varphi$ for all $i \prec k$	P_C:	$C_k\varphi \supset C_i\varphi$ for all $i \prec k$

and is closed under all instances of the rules:

$$\frac{\varphi}{\Box\varphi}\ \text{(RN)} \qquad \frac{\varphi \ \ \varphi \supset \psi}{\psi}\ \text{(MP)}$$

P_B and P_C are the *Persistence* axioms for B and C respectively. We write $\vdash \varphi$ if φ is a theorem of Æ. If $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \supset \psi$, we sometimes write $\varphi_1, \ldots, \varphi_n \vdash \psi$ and refer to $\varphi_1, \ldots, \varphi_n$ as *premises*.

The logic of $\Box$ is S5. To see this, note that the T schema is an axiom, that we can combine axioms Def$\Box$, $B_\Box$ and $C_\Box$ to yield the 4-schema $\Box\varphi \supset \Box\Box\varphi$ and that Def$\Box$, $\overline{B}_\Box$ and $\overline{C}_\Box$ yield the 5-schema $\neg\Box\varphi \supset \Box\neg\Box\varphi$. Similarly, we can show that B_k and C_k are both K45.

2.2. Models

An *Æ-model M* is a quadruple (U, U^+, U^-, V), where U is a non-empty set of *points*; U^+ and U^- are functions which assign a subset of U to each index in I. $U^+(k)$ is denoted U_k^+; U_k^- denotes $U^-(k)$. V is a valuation function which assigns a subset of U to each propositional letter in the language.

An Æ-model is intended to represent a doxastic subject (cf. Section 3). U is the universe of the subjectively possible states of affairs, i.e., the range of states of affairs that the subject can conceive of. Equating possibility and conceivability, we occasionally refer to U as the *space of conceivability*. The subject has a *belief state* at each degree of confidence k, modeled by U_k^+, and a *co-belief state*, modeled by U_k^-. We abbreviate 'degree of confidence' as *doc*. Any point in U_k^+ is a candidate, with *doc k*, for being the actual world. Points in U_k^+ are called *k-plausible*, or *doxastic alternatives at k*. U_k^- is, accordingly, the set of *k-implausible* points, i.e., the set of worlds that are ruled out, at k, as not actual.

For each $k \in I$, we require that

$$U_k^+ \cup U_k^- = U. \qquad \text{(SU)}$$

SU, named for *subjective universe*, expresses that the universe is just the set of points that are, for any given k, either plausible or implausible to the doxastic subject. While belief states may vary between degrees of confidence, the universe does not.

The more alternatives the subject can rule out, the stronger the belief state. We require that greater confidence is never accompanied by stronger belief:

$$U_k^+ \subseteq U_i^+ \quad \text{and} \quad U_i^- \subseteq U_k^- \quad \text{for each } i \prec k. \qquad \text{(Persistence)}$$

The model is *bisected* if, for each $k \in I$,

$$U_k^+ \cap U_k^- = \emptyset. \qquad \text{(Bisection)}$$

This is not a model condition; not a property that is forced by the axiom system. However, we can force Bisection to be satisfied by syntactic means; this is proved in Lemma 2 below and further discussed in Section 4.2.

A satisfaction relation can be defined for each point x:

$M \vDash_x p$	iff	$x \in V(p)$ for a propositional letter p
$M \vDash_x \Box\varphi$	iff	$M \vDash_y \varphi$ for each $y \in U$
$M \vDash_x B_k\varphi$	iff	$M \vDash_y \varphi$ for each $y \in U_k^+$
$M \vDash_x C_k\varphi$	iff	$M \vDash_y \varphi$ for each $y \in U_k^-$

and as usual for Boolean connectives. A formula is *satisfied* in a model if it is true at one of its points. If $M \vDash_x \varphi$ for all $x \in U$, we write $M \vDash \varphi$ and say that φ is *true in M*. If φ is true in all models, we shall write $\vDash \varphi$.

Observe that all points in a model agree on the truth value of every completely modalized formula. Hence, for such formulae the notions of satisfaction and truth in a model coincide. This justifies use of the notation $M \vDash \varphi$ whenever a completely modalized φ is satisfied in M. Note that even though neither B_k nor C_k satisfies the D schema from modal logic, for each formula φ, either $B_k\varphi \supset b_k\varphi$ or $C_k\varphi \supset c_k\varphi$ is true in any given model.

It follows directly from the truth definition that a proposition is necessary if and only if it holds in all conceivable alternatives. We omit the easy proofs of the properties in the following lemma, in which we write $\|\varphi\|$ for the *truth set* of φ in M: $\{x \in U \mid M \vDash_x \varphi\}$.

Lemma 1. (1) *$M \vDash B_k\varphi$ iff $U_k^+ \subseteq \|\varphi\|$ and $M \vDash C_k\varphi$ iff $U_k^- \subseteq \|\varphi\|$.*
(2) *If $M \vDash O_k\varphi$, then $U_k^+ = \|\varphi\|$ and $U_k^- = \|\neg\varphi\|$.*
(3) *$\|\varphi\| \subseteq \|\psi\|$ iff $M \vDash \varphi \supset \psi$.*
(4) *$\|\varphi\| = \|\varphi[\psi_1/\psi_2]\|$ if $\|\psi_1\| = \|\psi_2\|$.*

Lemma 2. *Let M be a model which, for each $k \in I$, satisfies a formula $O_k\varphi_k$. Then M is bisected.*

Proof. By Lemma 1(2). □

Lemma 3. *Let β be completely modalized. Then $\vDash \varphi \equiv (\varphi[\beta/\top] \wedge \beta) \vee (\varphi[\beta/\bot] \wedge \neg\beta)$.*

Proof. Let M be an arbitrary model. If β is completely modalized, then $M \vDash \beta \equiv \top$ or $M \vDash \beta \equiv \bot$. Applying Lemma 1(4), we obtain $M \vDash \varphi \equiv \varphi[\beta/\top]$ or $M \vDash \varphi \equiv \varphi[\beta/\bot]$, respectively, and thus in either case we have

$$M \vDash \big((\beta \equiv \top) \wedge \big(\varphi \equiv \varphi[\beta/\top]\big)\big) \vee \big((\beta \equiv \bot) \wedge \big(\varphi \equiv \varphi[\beta/\bot]\big)\big),$$

which is tautologically equivalent to the formula in the lemma. □

The next lemma expresses a simple reduction property of nested occurrences of modalities.

Lemma 4. *Any formula is equivalent to a formula without nested occurrences of modalities.*

Proof. The lemma is trivial for formulae without nested modalities; for an induction on the length of formulae, suppose φ contains nested modalities. Then φ contains a modal atom β as a proper subformula. By Lemma 3, φ is equivalent to a Boolean combination of the shorter formulae $\varphi[\beta/\top]$, $\varphi[\beta/\bot]$ and β, and hence we are done by the induction hypothesis. □

We shall say that formulae without nested occurrences of modalities are in *normal form*. Formulae in normal form that are not purely Boolean shall be said to be of *modal depth* 1.

Lemma 5. *Let $M \vDash O_i\varphi_i$ and $M \vDash O_k\varphi_k$. If $i \prec k$, then $M \vDash \varphi_k \supset \varphi_i$.*

Proof. Assume that $O_i\varphi_i$ and $O_k\varphi_k$ both hold in M and that $i \prec k$. By Lemma 1(2) and the Persistence requirement on models we then obtain $\|\varphi_k\| \subseteq \|\varphi_i\|$, which yields the conclusion of the lemma by an application of Lemma 1(3). □

Theorem 6. *Æ is sound, i.e., $\vdash \varphi$ implies $\vDash \varphi$ for any φ.*

Proof. By routine induction on the length of proofs. □

2.3. Completeness and the finite model property

To facilitate the proofs of completeness and decidability we shall in this section introduce an alternative semantics for Æ. The alternative interpretation is a standard relational semantics with a structure that directly reflects the axioms.

Let us say that a *frame* is a quadruple (W, E, R, S), where the universe W is non-empty, E is a binary relation over W, and R and S are functions that assign binary relations over W to each *doc* in I. R and S satisfy the following constraints:

(f1) $X \circ Y \subseteq Y$, where X is either R_k or S_k, and Y is either R_k, S_k, or one of their complements $\overline{R_k}$, $\overline{S_k}$,
(f2) E is reflexive,
(f3) $E = R_k \cup S_k$ for each $k \in I$,
(f4) if $i \prec k$, then $R_i \subseteq R_k$ and $S_k \subseteq S_i$.

As usual R_k denotes $R(k)$ and $(x, y) \in X \circ Y$ iff there is a z such that xXz and zYy. Two of the eight instances of (f1) state that R_k and S_k are transitive, e.g. $R_k \circ R_k \subseteq R_k$, while two of them state that they are Euclidean, e.g. $R_k \circ \overline{R_k} \subseteq \overline{R_k}$. A *relational model* is a frame equipped with a valuation function V which assigns a subset of the universe to each propositional letter. A relational model M evaluates modal formulae by quantifying over R_k and S_k: $M \vDash_x B_k\varphi$ iff $M \vDash_y \varphi$ for each y such that xR_ky; $M \vDash_x C_k\varphi$ iff $M \vDash_y \varphi$ for each y such that xS_ky.

Lemma 7. *E is an equivalence relation.*

Proof. Reflexivity is (f2). Transitivity: Let k by any index in I. Then $(R_k \cup S_k) \circ (R_k \cup S_k) \subseteq R_k \circ R_k \cup R_k \circ S_k \cup S_k \circ R_k \cup S_k \circ S_k$. By (f1) each composition is included in $R_k \cup S_k$. By (f3), $E \circ E \subseteq E$. Euclideanness: $(R_k \cup S_k) \circ \overline{(R_k \cup S_k)} \subseteq (R_k \circ \overline{R_k} \cap R_k \circ \overline{S_k}) \cup (S_k \circ \overline{R_k} \cap S_k \circ \overline{S_k})$, which, by (f1), is included in $\overline{R_k} \cap \overline{S_k} = \overline{R_k \cup S_k}$. By (f3), $E \circ \overline{E} \subseteq \overline{E}$. □

A *cluster* is an equivalence class of W modulo E. Let C be an E-cluster. We define the *belief part with regard to k*, C_k^+, and the *co-belief part with regard to k*, C_k^-, of C by: $C_k^+ = \{x \in C \mid xR_kx\}$ and $C_k^- = \{x \in C \mid xS_kx\}$.

Lemma 8. *$R_k \cap (C \times C) = C \times C_k^+$ and $S_k \cap (C \times C) = C \times C_k^-$.*

Proof. Let $x, y \in C$. Assume first that xR_ky. By Euclideanness, yR_ky. This shows that $R_k \cap (C \times C) \subseteq C \times C_k^+$. Conversely, assume yR_ky. The assumption that $x, y \in C$ implies, by (f2), that either xR_ky or xS_ky. In the latter case $S_k \circ R_k \subseteq R_k$ gives xR_ky; hence xR_ky in any case. S_k is treated symmetrically. □

Clusters provide the link between the relational models in this section and the Æ-models introduced in Section 2.2. The basic intuition is that a cluster corresponds exactly to a subjective universe. Note that if x is a point in a cluster C, then C consists of all and only those points which x can see through E, i.e., through either R_k (which gives the k-plausible points) or S_k (which gives the k-implausible points).

Lemma 9. *Each Æ-model is isomorphic to a cluster in a relational model, and* vice versa. *Hence a formula has an Æ-model iff it has a relational model.*

Proof. From an Æ-model (U, U^+, U^-, V) we can construct a relational model $(U, U \times U, R, S, V)$, where $R_k = U_k^+$ and $S_k = U_k^-$. It is easy to verify that the four frame conditions hold in the model; this is left to the reader. By Lemma 8 the relational model has one and only one cluster, i.e. U; it is hence immediate that the two models agree on the truth value of every formula at any point in U. Conversely, let C be a cluster in a relational model (W, E, R, S, V) and let V_C be a valuation function such that $V_C(p) = V(p) \cap C$. Let C^+ be a function which to each *doc k* assigns the set C_k^+ and C^- be defined similarly. Consider the quadruple $M = (C, C^+, C^-, V_C)$. Isomorphy follows from Lemma 8. To see that M is an Æ-model, note that since $R_k \cup S_k$ is reflexive, either xR_kx or xS_kx. This shows that the SU condition is met. Persistence holds by (f4). Clearly, truth coincides in the two models for every point in C. □

Lemma 9 justifies a technical focus on relational models as all results established on the basis of these models immediately transfer to Æ-models (and hence to Æ). This holds in particular for completeness and the Finite model property, addressed next.

A set s of formulae is *maximal* if it is consistent, and every proper extension of it is inconsistent. Two maximal sets s and t *agree* on a formula φ if $\varphi \in s$ iff $\varphi \in t$. Let U^c be the set of all maximal sets, $V^c(p) = \{s \in U^c \mid p \in s\}$, where p is a propositional letter, and the binary relations R_k^c, S_k^c on U^c be defined by:

$$sR_k^ct \quad \text{iff} \quad \text{for all} \quad \varphi,\ B_k\varphi \in s \rightarrow \varphi \in t,$$

$$s S^c_k t \quad \text{iff} \quad \text{for all} \quad \varphi,\ C_k\varphi \in s \to \varphi \in t,$$
$$s E^c t \quad \text{iff} \quad \text{for all} \quad \varphi,\ \Box\varphi \in s \to \varphi \in t.$$

The *canonical model* M^c is defined as $(U^c, E^c, R^c, S^c, V^c)$, where R^c assigns R^c_k to each k and S^c similarly.

Lemma 10. *Let sE^ct. Then s and t agree on all modal atoms.*

Proof. Let φ be a modal atom or the negation of a modal atom. If $\varphi \in s$, $\Box\varphi \in s$ by one of the axioms $B_\Box$, $C_\Box$, $\bar{B}_\Box$ and $\bar{C}_\Box$. By construction of E^c, $\varphi \in t$. □

Since the modalities are normal, the *Truth Lemma* holds: $M^c \vDash_s \varphi$ iff $\varphi \in s$. Completeness can then be established in the usual way:

Theorem 11. *Æ is complete, i.e., if $\vDash \varphi$ then $\vdash \varphi$.*

Proof. It is sufficient to prove that M^c is a relational model, i.e., that the frame conditions are met. Each subcondition of (f1) is imposed on M^c by one of the axioms $B_\Box$, $C_\Box$, $\bar{B}_\Box$ and $\bar{C}_\Box$ in combination with Def□. We illustrate one of them: let sS^c_kt and tR^c_ku, and assume $B_k\varphi \in s$. By $C_\Box$ and Def□, $C_kB_k\varphi \in s$. Hence $\varphi \in u$ and sR^c_ku. Condition (f2) is imposed by axiom T. As to (f3), assume first that not sR^c_kt and not sS^c_kt. There are then formulae $B_k\varphi \in s$ and $C_k\psi \in s$ such that $\neg(\varphi \vee \psi) \in t$. By Def□, $\Box(\varphi \vee \psi) \in s$, hence not sE^ct. Conversely, assume that not sE^ct, i.e., that there is a $\Box\varphi \in s$ such that $\neg\varphi \in t$. By Def□, $B_k\varphi \in s$ and $C_k\varphi \in s$. By construction, not sR^c_kt and not sS^c_kt. Axioms P_B and P_C force condition (f4) on M^c. □

A *filtration set* is a set of formulae Ψ which is closed under subformulae, and which satisfies *≺-closure:* whenever $i \prec k$, $B_k \in \Psi$ if $B_i \in \Psi$ and $C_i \in \Psi$ if $C_k \in \Psi$.

The filtration $M^\dagger = (U^\dagger, E^\dagger, R^\dagger, S^\dagger, V^\dagger)$ of M^c through Ψ is defined as follows. $U^\dagger$ is the set of equivalence classes of U^c modulo $\sim_\Psi$, where the equivalence relation $\sim_\Psi$ on U^c is defined by: $s \sim_\Psi t$ iff $s \cap \Psi = t \cap \Psi$; the equivalence class of s modulo $\sim_\Psi$ is denoted $|s|$. $V^\dagger$ is a function which, for all propositional letters p in Ψ, satisfies: $V^\dagger(p) = \{|s| \mid p \in s\}$. The binary relations are given by:

$$|s|E^\dagger|t| \quad \text{iff} \quad s \text{ and } t \text{ agree on all modal atoms in } \Psi,$$
$$|s|R^\dagger_k|t| \quad \text{iff} \quad \forall\chi(B_k\chi \in s \cap \Psi \to \chi \in t \cap \Psi) \text{ and } |s|E^\dagger|t|,$$
$$|s|S^\dagger_k|t| \quad \text{iff} \quad \forall\chi(C_k\chi \in s \cap \Psi \to \chi \in t \cap \Psi) \text{ and } |s|E^\dagger|t|.$$

Lemma 12. *$M^\dagger$ is a relational model.*

Proof. All subconditions of the frame condition (f1) are easily verified from the definition of $R^\dagger_k$ and $S^\dagger_k$; we show that $S^\dagger_k \circ R^\dagger_k \subseteq R^\dagger_k$. Let $|s|S^\dagger_k|t|$ and $|t|R^\dagger_k|u|$. Then $|s|E^\dagger|u|$; hence $B_k\varphi \in s \cap \Psi$ only if $B_k\varphi \in t \cap \Psi$ only if $\varphi \in u \cap \Psi$. $E^\dagger$ is trivially reflexive. To show (f3), assume that not $|s|R^\dagger_k|t|$ and not $|s|S^\dagger_k|t|$. There is then an $s \in U^c$, and formulae $B_k\varphi$ and $C_k\psi$, both in $s \cap \Psi$, such $\varphi \notin t$ and $\psi \notin t$. By axiom Def□, $\Box(\varphi \vee \psi) \in s$. If $|s|E^\dagger|t|$, $\varphi \vee \psi \in t$; hence either $\varphi \in t$ or $\psi \in t$, contradicting the previous conclusion. This proves that $E^\dagger \subseteq R^\dagger_k \cup S^\dagger_k$. The converse inclusion of (f3) is trivial. Condition (f4) follows from ≺-closure of the filtration set and the persistence axiom. □

The *Filtration Theorem* holds for Æ:

Theorem 13. *For all $s \in U^c$ and $\varphi \in \Psi$, $M^\dagger \vDash_{|s|} \varphi$ iff $\varphi \in s$.*

Proof. By standard theory it is sufficient to show that $E^\dagger$ satisfies the following two conditions: (i) if sE^ct then $|s|E^\dagger|t|$; (ii) if $|s|E^\dagger|t|$ and $\Box\varphi \in s \cap \Psi$, then $\varphi \in t \cap \Psi$, and that $R^\dagger_k$ and $S^\dagger_k$ satisfy the corresponding two conditions. Condition (ii) trivially holds for all of them. To prove (i) for $E^\dagger$, assume that not $|s|E^\dagger|t|$. Then $|s|$ and $|t|$ disagree on some modal atom in Ψ. By Lemma 10 not sE^ct. Condition (ii) is then easily verified for $R^\dagger_k$ and $S^\dagger_k$. □

Corollary 14. *Æ is determined by the set of finite frames and is decidable.*

Proof. By Lemmas 12, Theorem 13 and the fact that the size of the filtration set which contains a formula, say φ, can be bound by $|I| \times |\varphi|$, where $|\varphi|$ is the number of subformulae of φ. □

3. Interpretation

This section is dedicated to the informal interpretation of Æ, with an emphasis on models. For comparison, consider what we may call standard doxastic logic, the modal logic KD45 with a single belief operator B. Æ differs from the standard in three main respects: it does not have a consistency (D) schema for belief, it has an additional 'co-belief' modality C, and it introduces a set of degrees of confidence indexing the modalities.

Although "only knowing" logic is not a new idea (Levesque's pioneering paper [8] is fifteen years old at the time of writing), we believe there is still a place for a closer look at the conceptual framework and the proper interpretation of the doxastic language in B and C. Throughout this section, *doc* indexes will be suppressed where they are not subject of discussion.

3.1. A personal universe

The first word about Æ models is that the universe U comprises every state of affairs that is conceivable to the implicit doxastic subject. Conceivability is here to be understood as the expression of "personal metaphysics": the subject's notion of what is necessarily the case, what it takes for granted, basic presuppositions, that which is beyond doubt. U is the range of worlds that the subject considers possible.

At each degree of confidence k, U_k^+ defines a *belief state*, consisting of the doxastic alternatives, the points that are consistent with what is believed at k. It is natural to view the belief modality B as expressing that points are *ruled out as implausible*: $B\varphi$ expresses that every point that has $\neg\varphi$ true is ruled out. $C\varphi$, on the other hand, expresses that *no* point at which $\neg\varphi$ is true is ruled out, that every $\neg\varphi$ point is plausible (given a bisected model; see Section 3.2).

The semantic constraint that U is non-empty implies that the truth axiom $\Box\varphi \supset \varphi$ is valid (what is necessary is true). We may interpret the constraint as a requirement that the *real* state of affairs is conceivable. The real state of affairs is however not necessarily contained in U_k^+, and while it is never *plausible* for the subject that any of its beliefs is false, situations in which some are may still be *conceivable*; if so, then as implausible. In formal terms, while $\Diamond(B\varphi \wedge \neg\varphi)$ is satisfiable, $b(B\varphi \wedge \neg\varphi)$ is not.

Where we draw the line between the necessary and the contingent is a matter of practical application. The minimal notion of metaphysics is easily identified: to count only logical truths as necessary. This requires a maximal universe, with every logically consistent proposition represented at some point. For a comprehensive metaphysics, U should be restricted so that no point validates anything the subject considers impossible. One case in point is the analytic relationships between concepts. There is a great difference between the belief that "cats are not made of stone" and a contingent belief such as "cats make ideal pets". Certainly, the fact that what is taken for granted varies from context to context is of considerable relevance to common-sense reasoning. It is therefore a notable feature of Æ models that they can accommodate stricter notions of what is necessary than the logically true.

3.2. At most *and O*

In the "only knowing" literature, from Levesque's [8] onwards, the focus has mainly been on applications of the O operator, with $O\varphi$ expressing that precisely φ, i.e., φ, and at most φ, is believed. 'At most φ is believed' means that φ is at least as strong as the strongest believed proposition, and $C\neg\varphi$ has been, appropriately, taken to express this notion.

With Æ, the situation is less straightforward. An 'at most' interpretation of $C\neg\varphi$ is supported by the observation that whenever $C\neg\varphi$ is true, the truth set of φ is a subset of the belief state,

$$\text{If } M \models C\neg\varphi \quad \text{then } \|\varphi\| \subseteq U^+. \tag{1}$$

The lack of Bisection as a requirement for Æ models means, however, that the converse does not hold, which implies that what is believed at most is only partly characterized by C formulae. While a set of C formulae can approximate the

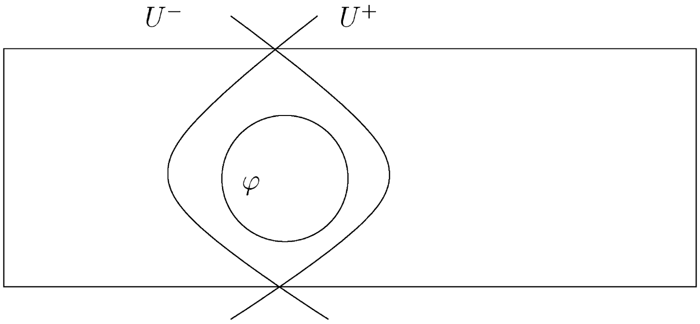

Fig. 1. A non-bisected model.

belief state, as evidenced by (1), in cases where the set of formulae has a model which is not bisected the approximation can never supply a complete characterization.

In a non-bisected model, $U^+ \cap U^-$ is nonempty, a set of points that are both plausible and implausible, and the belief state does not properly complement the co-belief state. Every proposition φ whose truth set $\|\varphi\|$ is a subset of $U^+ \cap U^-$ (cf. Fig. 1) should be said to be believed 'at most'. (Too see this, note that if $\|\varphi\|$ is a subset of $U^+ \cap U^-$, it is a subset of the belief state U^+. Hence φ is at least as strong as the strongest believed proposition, which by definition means that at most φ is believed.) $C\neg\varphi$ will not hold in such a case. Furthermore, there is then no formula ψ such that $B\psi \wedge C\neg\psi$ (i.e., $O\psi$) is true.

It is clear that we cannot provide a full account of the meaning of $C\varphi$ by means of the notion of belief 'at most' alone. For the purposes of providing an informal account of meaning, we shall associate the attitude of *caution* with co-belief, and hence with the set of implausible alternatives U^-. Accordingly, let the notion of being skeptical apply to belief and the set of plausible alternatives U^+. In a bisected model, caution and skepticism as employed here will be extensionally equivalent.[1]

In order to find an interpretation of $C\varphi$ that doesn't rely on belief 'at most', consider the roles of belief and co-belief in a model. According to the model definition, beliefs serve to exclude conceivable alternatives from the set of plausibles, while co-beliefs serve to exclude conceivable alternatives from the set of *im*plausibles. The more a subject co-believes, the more cautious the subject is.

We propose the following reading of $C\varphi$, for which to 'rule out' refers to rendering implausible.

$C\varphi$: Every alternative that caution permits to be ruled out has φ true.

To see why this is appropriate, consider again how a belief state is determined. The more the subject believes, the smaller is the set of plausible points. Each belief (in a contingent proposition) contributes to the approximation of the belief state by eliminating points from the set of plausible alternatives. With co-belief, the situation is inverted. The more is co-believed, the less is considered implausible. A maximally cautious co-believer will consider no alternative implausible, just as the maximally skeptical believer considers every alternative plausible. The adoption of a co-belief that φ amounts to an increase in caution: to excluding all non-φ points from the set of implausible alternatives.

A non-bisected model may be interpreted as a model of a subject that has a discrepancy between caution and belief; between caution to not accept evidence and actual acceptance. In a bisected model, however, there is a co-belief to match every belief; just the right amount of caution to match what is believed.

Only when we have a bisected model is it appropriate to say that co-belief determines the belief state U^+, and that belief characterizes the co-belief state U^-. Similarly, while b and c restrict U^+ and U^-, respectively, from below, when the model is bisected, and only then, we can appropriately say that c is an expression of belief, and b an expression of co-belief.

Non-bisected models may have a potential for representing a subtle form of doxastic underdetermination. Nevertheless, we will in the remainder of this section work under the presupposition that models are in fact bisected, in line with the common practice of interpreting C as an "at most" operator.

[1] The requirement SU (Section 2.2) is analogous to a principle of excluded middle in ensuring that every point is either plausible or implausible. Accordingly, when Bisection is satisfied the model satisfies a form of non-contradiction, in that no point is both plausible and implausible.

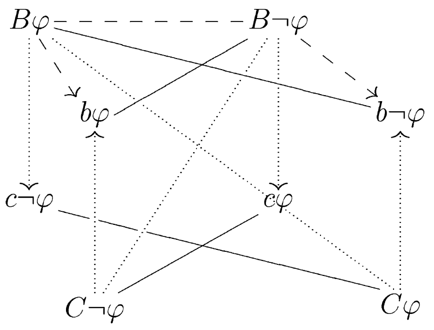

Fig. 2. Cube of oppositions: B and C.

3.3. *B* and *C*

How does the language of Æ expand on the traditional doxastic language in B, and how are the additional modalities that Æ provides related? Let us begin by pointing out what B and b, which have a standard definition in Æ, mean. $B\varphi$ is a *general and positive* belief modality, expressing that every conceivable alternative at which φ is false is implausible; $b\varphi$ is a *particular and negative* modality that says some alternative at which φ is true is plausible. Thus, $B\varphi$ expresses belief (in a weak sense, knowledge) while $b\varphi$ expresses non-belief (in a weak sense, ignorance).

In Æ, the language is enriched with a *general and negative* modality C, and a *particular and positive* modality c. Fig. 2 relates the simple modalities in a design analogous to a traditional "square of oppositions", where arrows express implication and lines without arrowheads express contradiction. Different styles of line correspond to preconditions. Whole lines illustrate unconditional relationships, dashed lines apply only given a consistent belief state, and dotted lines apply only if φ is contingent. The positive (knowledge) modalities have been placed in the rear corners of the cube, negative (ignorance) modalities in front.

Let a *doxastic attitude* toward a proposition be the attitude of belief, disbelief, undecidedness, and so forth as given by a model. The addition of C to the doxastic language in B expands the range of doxastic attitudes that we can express. For illustration, consider the following as examples of propositions that a doxastic subject can relate to.

φ: "David Beckham hates public attention"

ψ: "David Beckham is a cat"

χ: "David Beckham is a cat and David Beckham is not a cat"

φ is generally believed to be false, but could conceivably be true (the popular image of Mr. Beckham may well be misleading). Barring extreme skepticism, ψ is an impossibility, although not a logical impossibility by most standards. χ is a logical impossibility according to any relevant standard. These are distinctions that are readily grasped, and it is desirable that a doxastic language should be able to express them.

Using C and B together, we can distinguish belief in propositions that the subject takes to be necessarily true from belief in contingent propositions. φ is an example of a proposition that is believed to be false, although it is perfectly possible to imagine a world in which it is true; this doxastic attitude is captured in formula (2).

$$B\neg\varphi \wedge c\varphi \tag{2}$$

For any proposition ψ that is considered impossible, we can make use of the following representation.

$$B\neg\psi \wedge C\neg\psi \tag{3}$$

(3) is by definition equivalent to $\Box\neg\psi$ – "ψ is impossible". Semantically, $B\neg\psi$ expresses that no alternative that has ψ true is plausible, while $C\neg\psi$ says that no alternative with ψ true is implausible; by (SU), then, there can be no alternative in which ψ is true.

The C operator enables us to express the distinction between belief in necessities and belief in contingent propositions. It might also be useful to be able to distinguish "metaphysical" from *logical* necessity, such as with χ above, but

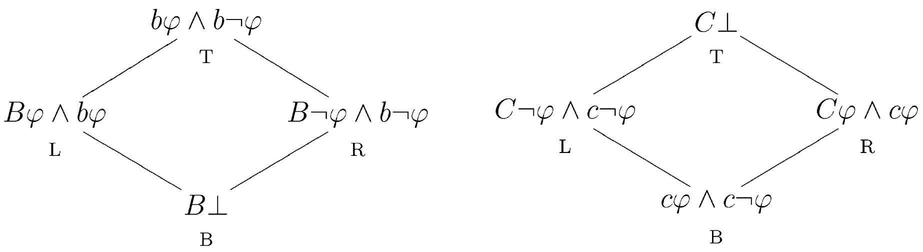
$b\varphi \wedge b\neg\varphi$
T
$B\varphi \wedge b\varphi$
L
$B\neg\varphi \wedge b\neg\varphi$
R
$B\bot$
B
$C\bot$
T
$C\neg\varphi \wedge c\neg\varphi$
L
$C\varphi \wedge c\varphi$
R
$c\varphi \wedge c\neg\varphi$
B

Fig. 3. The doxastic positions for B and C.

this desideratum is not fulfilled. We cannot distinguish, by Æ expressions, non-belief due to metaphysical impossibility from non-belief due to logical impossibility. (We do however provide an axiomatic method for specifying what is necessary and what is contingent, in Section 4.)

The attitude expressed by $B\varphi$ is often called "knowledge". The lack of a truth schema for B implies, however, that Æ should be seen as a logic of belief only, as the notion of knowledge clearly requires that what is known is also true.[2] A similar consideration applies to calling C an "ignorance" operator: only part of what goes by the name of ignorance is captured by expressions such as $C\varphi$. Knowledge and ignorance are on a par in presupposing a notion of *correctness* that is not required for simple belief. Two prominent cases of ignorance are, to believe φ while φ is false, and to not believe φ while φ is true. The truth value of φ is essential to both variants.

Because it does satisfy the truth axiom T, $\Box$ may be considered a candidate for a knowledge operator in Æ. However, because $\Box$ also satisfies the negative introspection schema $\Diamond\varphi \supset \Box\Diamond\varphi$, it is arguably too strong for a proper knowledge operator (see [7, p. 79]).

3.4. Doxastic positions

Expressions of doxastic attitude toward a proposition can be more or less precise. For instance, an expression of belief $B\varphi$ can be specified further by combining it with $b\varphi$, to express that φ is consistently believed, or with $C\neg\varphi$ to say that precisely φ is believed.

Referring to Fig. 2, we see that there are four consistent pairs of modal literals for each of B and C, the pairs linked by horizontal edges of the cube. Each pair corresponds to a distinct doxastic attitude toward the subject proposition, and each pair of formulae expresses this doxastic attitude as precisely as can be done using the respective operator. Forming conjunctions of each pair (for B and C, respectively), we obtain a set of consistent and mutually exclusive formulae, the disjunction of which is a tautology. We will refer to members of such a set of formulae as *doxastic positions*.[3]

A set of doxastic positions corresponds to the range of doxastic attitudes that can be distinguished by means of the respective operator. It is therefore useful for the purpose of describing the expressive power of the doxastic language.

The sets of doxastic positions for B and C are shown in Fig. 3, ordered by weakness of belief state. Labels T, L, R, and B stand for "top", "left", "right", and "bottom". The top positions are expressions of lack of belief, of weakness of the belief state. For B, the T position expresses undecidedness with regard to φ. The T position for C expresses maximal weakness of belief, that no contingent proposition is believed. Employing a useful analogy, we may say that $C\bot$ is true when the "database" of beliefs is empty. The B positions are the strongest expressions of belief. For B, the B position corresponds to inconsistent belief (every proposition is believed). The B position for C says that there are some φ alternatives as well as some non-φ alternatives among those that are inconsistent with what is believed. The L and R positions are intermediate with regard to strength of the belief state. For B, position L represents the attitude

[2] The reader may well find this comment uncongenial to what is implied by the title of this paper, namely, that Æ should be a logic of knowledge. Our choice of title was dictated by the standard naming practice for "only knowing" logics. Perhaps in favor of this practice, observe that the notion of belief expressed by $B\varphi$ is clearly stronger than the commonsensical use of the term "belief", in expressing a notion of *conviction* that φ is true, and not a weaker attitude that φ is, e.g., more likely to be true than non-φ. The common association of $B\varphi$ with knowledge that φ may be by analogy to the strong connection between having a belief and possessing evidence for that belief, which amounts to having learned that φ is true.

[3] The treatment here is inspired by the theory of *normative* positions in deontic and action logic, in particular [16], which see for a formal framework and references. Little has been published regarding doxastic positions. The first use of the term may be in [11].

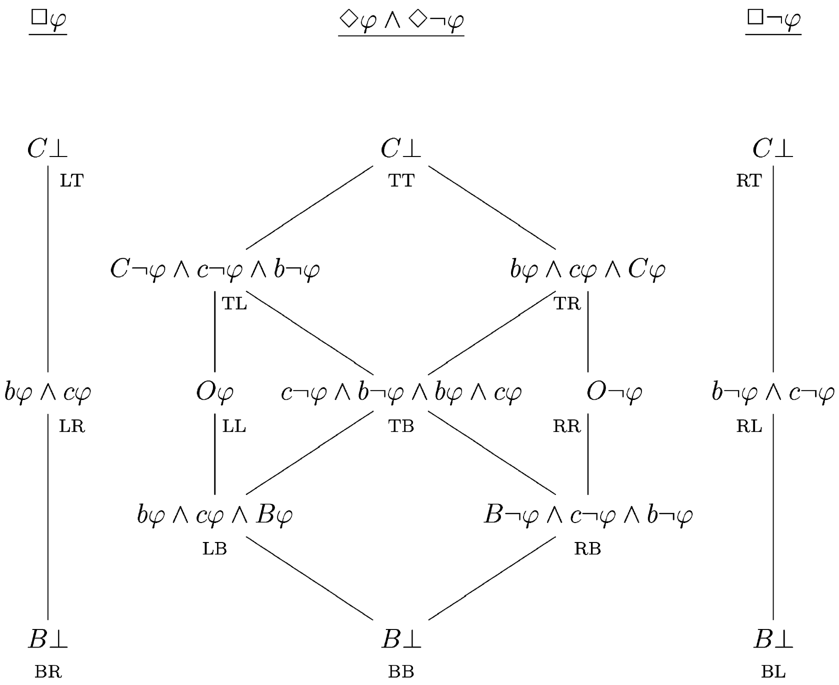
$\Box\varphi$
$\Diamond\varphi \wedge \Diamond\neg\varphi$
$\Box\neg\varphi$
$C\bot$
LT
$C\bot$
TT
$C\bot$
RT
$C\neg\varphi \wedge c\neg\varphi \wedge b\neg\varphi$
TL
$b\varphi \wedge c\varphi \wedge C\varphi$
TR
$b\varphi \wedge c\varphi$
LR
$O\varphi$
LL
$c\neg\varphi \wedge b\neg\varphi \wedge b\varphi \wedge c\varphi$
TB
$O\neg\varphi$
RR
$b\neg\varphi \wedge c\neg\varphi$
RL
$b\varphi \wedge c\varphi \wedge B\varphi$
LB
$B\neg\varphi \wedge c\neg\varphi \wedge b\neg\varphi$
RB
$B\bot$
BR
$B\bot$
BB
$B\bot$
BL

Fig. 4. The 15 doxastic positions for B and C combined.

of consistent belief that φ. For C, the L position expresses that at most φ is believed, that nothing is believed that is stronger than φ ($C\neg\varphi$), while $\neg\varphi$ holds in some alternative that is incompatible with what is believed; i.e., that the database is not empty.

We now look at the possible combinations of B and C positions, in order to see what C adds to the language with just B with regard to a finer partitioning of the range of expressible attitudes.

Forming conjunctions of the B and C positions of Fig. 3, we find that 15 of 16 conjunctions are consistent; the exception is $B\bot \wedge C\bot$ (implying $\bot$ by axiom schemata Def$\Box$ and T). This set of doxastic positions is presented in Fig. 4.

Fig. 4 has been designed for comparison between positions rather than to show the full form of each position. It should be read as follows. Each node in the graph represents the combination of a B and a C position. The node labels LT, LR, and so forth show which combination of basic B and C positions is represented at each node: the left letter stands for the B position, the right letter for the C position. The top row indicates the alethic modality of φ that is implied by the position, and formulae at nodes in the graph have been simplified accordingly. For instance, the node LT, which corresponds to an empty belief state, is shown as $C\bot$, suppressing the conjunct $B\varphi \wedge b\varphi$ in order to facilitate comparison to the other nodes TT and RT that also represent empty belief states. As in Fig. 3, nodes are ordered by implied strength of belief state, with weaker belief states at higher nodes.

Explanations of what is expressed by each position follow. We will occasionally draw on the natural analogy between belief states and databases.

TT, LT, RT. These positions express that the subject has no contingent belief—that the database is empty. The positions differ only with regard to the modality of φ. LT expresses that φ is a necessity; RT, that it is an impossibility. While the positions will typically express a lack of belief, only the position TT necessarily implies a degree of ignorance. The exception obtains when every truth is necessary according to the subject's metaphysics, hence also believed. Such a subject has perfect belief. LT or RT will then properly express the belief state, depending on whether φ is true in the single conceivable alternative or not.

BR, BB, BL. These positions all represent an inconsistent belief state, and differ only in the modality assigned to φ, accordingly as in LT, TT, RT.

TL. For this position, $C\neg\varphi$ expresses that the subject believes *at most* φ, meaning that every φ-compatible proposition is plausible. From $b\neg\varphi$, we see that the position is consistent, while $c\neg\varphi$ informs us that the database is not empty, and φ is contingent: some $\neg\varphi$-alternatives are not plausible. The conjunction $C\neg\varphi \wedge b\neg\varphi$ may be interpreted as saying that *less than* φ is believed.

LL. In this position, precisely φ is believed. All and only the alternatives that have φ true are plausible.

TB. This position corresponds well with the common-sense notion of suspension of judgment with regard to φ. φ is believed neither true nor false ($b\varphi \wedge b\neg\varphi$). Furthermore, the database is not empty: some φ- as well as some $\neg\varphi$-alternatives are implausible.

LB. This position is the best candidate for expressing the default, common-sense notion that φ is believed. φ, *and more* ($c\varphi$), is believed, and the belief state is consistent ($b\varphi$).

LR. For this position, $B\varphi$ and $C\varphi$ combine to express that φ is necessary. $b\varphi$ ensures that belief is consistent, and $c\varphi$ that the database is not empty.

(The positions TR, RR, and RB differ from TL, LL, and LB only with respect to the negation of φ.)

Given that I is non-singleton, Persistence implies that the subject's attitude toward a proposition φ—the doxastic position of the subject with regard to φ—can only remain unchanged or get stronger as degree of confidence is lowered. The vertical ordering of nodes displayed in Fig. 4 therefore corresponds to possible changes of doxastic attitude towards φ from *doc* i to a lesser *doc* j.

3.5. Degrees of confidence

We now consider the question, to what extent does Æ provide an appropriate framework for representing the common-sense notion of degrees of confidence? With Æ, the set of *doc*'s I is ordered by the relation of conviction $\prec$, and $i \prec j$ is intended to mean that i is at least as great a *doc* as j.

The direction of the partial order relation $\prec$ deserves a comment. Following most authors in the field we write $i \prec k$ to denote that i is at least as great a degree of confidence as k. Persistence implies that the belief state corresponding to k is at least as strong as the belief state corresponding to i when $i \prec k$, a perspective which justifies the direction of the sign. There is an intrinsic tension in the use of the symbol because we want it to reflect that in general, lesser degrees of confidence are accompanied by increase in strength of the belief state. The direction of the sign necessarily has to reflect one and only one of these perspectives.

In practical application, the degrees of confidence we acknowledge, and the relations between them, vary widely with the situation. On the roughest possible approach, only one degree of confidence is recognized. This means we don't distinguish the reliability of mathematical beliefs from the reliability of beliefs based on hearsay. When we do make a distinction, a common approach is to recognize two degrees of confidence, one for beliefs based on empirical evidence and another, greater, degree to beliefs based on a priori insights. We can go on to make arbitrarily fine-grained distinctions; indeed, it may be possible to argue that every belief should be assigned its own separate degree of confidence. Finding an appropriate resolution is a pragmatic, application-dependent issue.

There is little reason, in general, to expect that a relation of degrees of confidence will be linear. Beliefs typically come with different degrees of confidence because the sources of the evidence upon which the beliefs are based vary with regard to their reliability. Just as it may be difficult to compare the reliability of sources, it may be difficult or impossible to determine which of two degrees of confidence should be considered greater.

The model condition that governs degrees of confidence in Æ is Persistence (Section 2.2), corresponding to the axioms P_B and P_C. Persistence of belief means that what is believed with a given degree of conviction is also part of what is believed at any comparably lesser degree of confidence.

Persistence is clearly a required property for a relation of strength of belief. Note that only what is expressed by *positive* belief modalities is preserved in lesser *doc*'s. The following formulations of the positive modalities are helpful in revealing why this is appropriate.

$B_k\varphi$: I am k-confident that φ is true.

$c_k\varphi$: I am k-confident that φ is compatible with a falsity.

This yields the following explanations of $b_k\varphi$ and $C_k\varphi$, emphasizing that they express lack of belief.

$b_k\varphi$: I am not k-confident that φ is false.

$C_k\varphi$: I am not k-confident that $\neg\varphi$ is compatible with a falsity.

Persistence applies to positive belief modalities only. b and C are expressions that there is no belief—hence, no belief to be preserved either.

It is difficult to account for the origin and character of degrees of confidence using only the notions of belief states and persistence. Intuitively, we prefer to consider belief states at varying *doc*'s as *outcomes*, as consequences of evidence processing that is sensitive to distinctions between more, and less, reliable sources of evidence. Degrees of confidence apply in the first instance to evidence, the "input", and only derivatively to belief, the outcome of a process that merges the input with antecedent doxastic attitudes. It is, for instance, insufficient to say that what is believed at priority i stems from a source (sources) of evidence whose reliability matches i, because what is believed at *doc* i depends not only on the evidence that is i-quality, but essentially also on whether that evidence is compatible with evidence that is more reliable than i. What is *believed* at priority i is typically only a subset of the evidence that has priority i. On this view, the property of Persistence should be understood on the basis of evidence priorities, and not as a principle of belief priorities. It is improper to say, for instance, that belief at *doc* j is "carried over" to a lesser *doc* i. The property of persistence is valid for belief at degrees of confidence as a consequence of the processing of prioritized evidence into prioritized beliefs. It does not reflect a property of beliefs *per se*.

The reasoning that shows why persistence should be considered valid on the basis of the structure of evidence uptake can be applied to motivate, also, that

$$B_k\varphi \supset b_i\varphi \quad \text{where } i \prec k \tag{Prudence}$$

should be a valid principle. Prudence is a consistency requirement that may be an attractive addition to Æ in applications. It expresses that what is believed at a lower *doc* must still be an option at every greater *doc*. In terms of models, this implies that the belief state should never be allowed to become empty.

Due to contraposition, Prudence implies that

$$B_k\varphi \supset b_i\varphi \quad \text{where } i, k \text{ are comparable} \tag{4}$$

and hence also a "local" consistency schema,

$$B_i\varphi \supset b_i\varphi. \tag{5}$$

This implies $b_i\top$ for any *doc* i. $C_i\neg\bot$ is always true, so Prudence implies that $C_i\neg\bot \wedge b_i\neg\bot$ holds, that *less than* $\bot$ is believed.

The consistency expressed in Prudence could not be captured with just the C operator. We express consistency of belief, in general, by using b; there is no means of expressing consistency with C or c, with the exception of the indiscriminate $C\top$.

Æ has two complementary axioms governing persistence, P_B and P_C. While P_B expresses that belief is preserved into lesser *doc*'s, P_C expresses that caution is preserved into greater *doc*'s. A principle expressed with C and c relating to Prudence like P_C relates to P_B would be

$$C_i\varphi \supset c_k\varphi \quad \text{where } i, k \text{ are comparable.} \tag{6}$$

This implies that some point is always implausible—that the database is never entirely empty.

Persistence ensures that belief states are never weakened as *doc* is lowered. We could also consider a stronger property: that a decrease in *doc* should always be accompanied by a genuine increase in strength of the belief state. This would ensure validity of the following principle.

$$O_i\varphi \supset c_k\varphi \quad \text{where } i \prec k. \tag{Progress}$$

There are many interesting representations that fail to satisfy Progress (including the example of Section 6, which does however satisfy Prudence). Degrees of confidence typically correspond to which sources have been listened to, which default rules have been applied, which pieces of information have been taken into account, and so forth. A *doc* i can be considered a record of the extent of processing that went into forming the belief state at i: the more evidence is taken into account, the lesser the *doc* to accompany resultant belief. There is in general nothing exceptional if the consideration of further evidence fails to strengthen the belief state. Further evidence may be unacceptable because it is overridden by evidence that carries more weight.

Where k is $\prec$-minimal, and unique, it is natural to understand B_k as a modality of full conviction. The following dialog demonstrates that acknowledging non-unique maximal *doc*'s is counter-intuitive.

– Do you believe φ?
– Yes I do, with a maximal amount of conviction.
– So you don't believe $\neg\varphi$, then?
– Oh, I do! In fact, there is nothing I believe with a greater degree of confidence.

We can hardly make sense of this conversation unless we interpret the respondent as expressing an inconsistent doxastic attitude toward φ. It demonstrates that only maximal *and unique* elements of I can be adequately interpreted as representing the common-sense notion of a "maximal degree of confidence".

In applications, it may be natural to require that only necessities are believed with maximal *doc*. We may assign a designated index $\Box$ for the maximal *doc*, and require that $O_\Box\varphi \equiv \Box\varphi$, i.e., $O_\Box\top$. In the same vein, minimal confidence could be associated with believing $\bot$.

It is worth pointing out that whenever a doxastic subject with degrees of confidence reports a belief that φ, the belief should be reported as being believed with the strongest *doc* at which φ is believed. (For instance, say there is a degree of conviction i to match what is reported in today's newspaper. While it is reasonable to say that $2+2=4$ is believed at i, it would be misleading not to report that this is a belief that is accompanied by a maximal *doc*.) This may be considered a requirement along the lines of a Gricean maxim to be as informative as possible.

4. Logical spaces and belief state representations for finite languages

The point of a logical space is to mirror the notion of personal metaphysics introduced in Sections 2.2 and 3.1. Recall that in the Beckham example (Section 3.3), χ can never be appropriately represented as contingent, while with ψ we have a choice of specifying it as contingent, necessary, or impossible. This can be achieved by means of different logical spaces. We can also, e.g., define a logical space λ such that $\lambda \vdash \Box(\text{penguin(Tweety)} \supset \text{bird(Tweety)})$ and thereby syntactically express a constraint on conceivability. We may, of course, state this conditional as part of the agent's beliefs, but conceptually analytic statements are expressed in a better way at the level of necessity.

In terms of model theory, a logical space corresponds to, and represents, the conceivability space U. U^+ and U^- are mirrored by formal expressions which we shall call belief state representations. They can be provided in an explicit form, from which a model can be directly determined, or in an implicit form which neither needs to have a unique model nor promises a quick way in which they can be determined. The deep content of the Modal Reduction Theorem is that any implicit representation can be brought to an equivalent form as a disjunction of explicit representations, a form which unambiguously exhibits all its models.

4.1. Logical spaces

We will in the rest of this paper assume that the language has finitely many propositional letters; assume they are $p_1, \ldots, p_n$ (in this fixed order). Let us say that an *atom* is a conjunction $\pm p_1 \wedge \cdots \wedge \pm p_n$ where $\pm p_i$ means either p_i or $\neg p_i$. An atom is the syntactical counterpart to a point, characterizing the material content of a state of affairs, i.e., the "external world", neglecting the agent's cognitive state. Note that any two distinct atoms are inconsistent with each other and that the disjunction of all atoms is a tautology.

A purely Boolean formula φ determines a unique set $\widehat{\widehat{\varphi}}$ of the atoms that imply φ. Note that $\bigvee\widehat{\widehat{\varphi}}$ is a full DNF equivalent of φ. The following properties are immediate.

Lemma 15. *Let φ and ψ be purely Boolean and α be an atom.*

(1) $\alpha \vdash \varphi$ *or* $\alpha \vdash \neg\varphi$,
(2) $\widehat{\widehat{\varphi}} \subseteq \widehat{\widehat{\psi}}$ *iff* $\varphi \vdash \psi$.

Let Γ be a non-empty set of atoms; note that Γ has at most 2^n elements. The *logical space spanned by* Γ is defined as

$$\lambda(\Gamma) = \bigwedge \Diamond\Gamma \wedge \Box \bigvee \Gamma$$

where $\Diamond\Gamma = \{\Diamond\varphi \mid \varphi \in \Gamma\}$. If Γ is the set of all atoms, $\lambda(\Gamma)$ is denoted $\lambda_\top$ and is called the *maximal logical space*. We shall in the following use the symbol λ to denote an arbitrary logical space.

Lemma 16. *Let Γ be a non-empty set of atoms and $M \vDash \lambda(\Gamma)$. Then, for each atom α, $\alpha \in \Gamma$ iff α is true at some point in M.*

Proof. Note that α is true at some point in M iff $\Diamond\alpha$ is a conjunct of $\lambda(\Gamma)$, iff $\alpha \in \Gamma$. □

Any point in a model satisfies one and only one atom. Since the language is finite we can, for each point $x \in U$, define the corresponding atom $[\![x]\!]$ in a straightforward way: $[\![x]\!]$ is the conjunction of the propositional letters with each p_i negated iff false at x.

Lemma 17. *Let φ be purely Boolean and M be a model.*

(1) *φ is true at x iff $[\![x]\!] \vdash \varphi$,*
(2) *if $M \vDash \lambda(\Gamma)$, then $x \in \|\varphi\|$ iff $[\![x]\!] \in \Gamma \cap \widehat{\widehat{\varphi}}$.*

Proof. (1) Assume that $\vDash_x \varphi$. Clearly, $\vDash_x [\![x]\!]$. We cannot have that $[\![x]\!] \vdash \neg\varphi$ since this would contradict soundness. Hence $[\![x]\!] \vdash \varphi$ follows from Lemma 15(1). The converse direction follows by completeness and the fact that $\vDash_x [\![x]\!]$. For (2), note that $[\![x]\!] \vdash \varphi$ iff $[\![x]\!] \in \widehat{\widehat{\varphi}}$ (by definition). Also note that since $M \vDash \lambda(\Gamma)$, $M \vDash \bigvee \Gamma$; hence $[\![x]\!] \in \Gamma$ for each x. Thus, $x \in \|\varphi\|$ iff $[\![x]\!] \in \widehat{\widehat{\varphi}}$ and $[\![x]\!] \in \Gamma$. □

Lemma 18. *Let φ be purely Boolean and λ be a logical space. Then either $\lambda \vdash \Diamond\varphi$ or $\lambda \vdash \neg\Diamond\varphi$.*

Proof. Let λ be spanned by Γ and M be any model such that $M \vDash \lambda(\Gamma)$. If $\|\varphi\| \neq \emptyset$, $\Gamma \cap \widehat{\widehat{\varphi}} \neq \emptyset$ by Lemma 17(2). If $\Gamma \cap \widehat{\widehat{\varphi}} \neq \emptyset$, Lemmas 16 and 15(1) give that $\|\varphi\| \neq \emptyset$. Hence φ is satisfied in M iff $\Gamma \cap \widehat{\widehat{\varphi}} \neq \emptyset$, i.e., iff φ is satisfied in every model of λ. Conclude by completeness. □

4.2. Explicit and implicit belief state representations

Let φ^I be a formula of the form $\bigwedge_{k \in I} O_k\varphi_k$. We will refer to φ^I as an *O_I-block*. If λ is a logical space, $\lambda \wedge \varphi^I$ is called a *belief state representation*. If each φ_k is purely Boolean, φ^I is a *prime O_I*-block and the corresponding belief state representation is *explicit*. Otherwise, $\lambda \wedge \varphi^I$ is an *implicit* belief state representation.

If an explicit belief state representation $\lambda \wedge \psi^I$ is satisfiable, it has essentially only one model; moreover, this model can easily be defined from the formula itself. Otherwise, $\lambda \wedge \psi^I$ is inconsistent due to a clash with the persistence axioms. These observations are made precise in the next lemma, which assumes that ψ^I is $\bigwedge_{k \in I} O_k\varphi_k$ and λ is spanned by Γ. Let us, to this end, say that two models are *modally equivalent* if they agree on the truth value of all modal atoms.

Lemma 19. *Let $\lambda \wedge \psi^I$ be an explicit belief state representation and suppose λ is spanned by Γ. Then*

(1) *all models of $\lambda \wedge \psi^I$ are modally equivalent,*
(2) *$\lambda \wedge \psi^I$ is consistent iff $\Gamma \cap \widehat{\widehat{\varphi}}_k \subseteq \widehat{\widehat{\varphi}}_i$ for each $i \preceq k$,*
(3) *either $\lambda \wedge \psi^I \vdash \varphi$ or $\lambda \wedge \psi^I \vdash \neg\varphi$ for each completely modalized φ.*

Proof. (1): Let $M = (U, U^+, U^-, V)$ and assume that $M \vDash \lambda \wedge \psi^I$. By Lemma 16, $\Gamma = \{[\![x]\!] \mid x \in U\}$. For each *doc* k, M partitions Γ into $\Gamma_k^+ = \{[\![x]\!] \mid x \in U_k^+\}$ and $\Gamma_k^- = \{[\![x]\!] \mid x \in U_k^-\}$. By Lemmas 17(2) and 1(2), $\Gamma_k^+ = \Gamma \cap \widehat{\widehat{\varphi}}_k$ and $\Gamma_k^- = \Gamma \setminus \Gamma_k^+$. Since Γ_k^+ and Γ_k^- are independent of M, every model of $\lambda \wedge \psi^I$ partitions Γ in this way. Clearly, all such models agree on the value of all modal atoms. (2): Assume that $\Gamma \cap \widehat{\widehat{\varphi}}_k \subseteq \widehat{\widehat{\varphi}}_i$ for each $i \preceq k$. It is then simple to construct a model for $\lambda \wedge \psi^I$; the assumption is used to prove that the model satisfies the Persistence property. By soundness, $\lambda \wedge \psi^I$ is consistent. Conversely, if $\lambda \wedge \psi^I$ is consistent, it has a model (by completeness). If $i \preceq k$, Lemma 5 gives that $\|\varphi_k\| \subseteq \|\varphi_i\|$. Conclude by Lemma 17(2). (3): The statement holds trivially if $\lambda \wedge \psi^I$ is inconsistent. Otherwise, it has a model which is modally equivalent to all its other models. In this model φ is either true or false. Conclude by completeness. □

It follows from Lemma 19(3) that the agent's attitude towards every proposition which can be expressed in the language can be determined from an explicit belief state representation. A more general explicit representation has the form

$$\lambda \wedge \left(\psi_1^I \vee \cdots \vee \psi_m^I\right)$$

where each ψ_i^I is a prime O_I-block. $\lambda \wedge \psi^I$ is a special case of this with $m = 1$; we call explicit representations of this form *unambiguous*. Otherwise, the formula (in general) conveys incomplete information about the agent's beliefs. Nevertheless the formula expresses that it has at most m models which are not modally equivalent. Note in particular that for any formula φ, we have $\lambda, \psi_1^I \vee \cdots \vee \psi_m^I \vdash \varphi$ if and only if $\lambda \wedge \psi_i^I \vdash \varphi$ for each ψ_i^I. Hence, if there are propositions towards which the agent's attitude is indeterminate, the explicit representation provides a clear method for identifying these.

If ψ^I is not prime, $\lambda \wedge \psi^I$ gives an *implicit* representation of a belief state; implicit because it will in general require some work to see what its models are. All representations of non-trivial common-sense situations will be implicit representations, some of which are addressed in Section 6. Whatever form they may have, there is a modal reduction property which applies to them to the effect that their content can be analyzed and stated in an explicit form by purely formal manipulations *within* the logic. This property is manifested in the *Modal Reduction Theorem*:

Theorem 20. *For each logical space λ and O_I-block φ^I, for some $m \geqslant 0$, there are prime O_I-blocks $\psi_1^I, \ldots, \psi_m^I$ such that*

$$\lambda \vdash \varphi^I \equiv \left(\psi_1^I \vee \cdots \vee \psi_m^I\right).$$

Every such ψ_k^I that is consistent with λ is called a λ-*expansion* of φ^I. In the case when φ^I has no λ-expansion, $\lambda \vdash \neg\varphi^I$. As each λ-expansion has a unique model, the "if" direction of the theorem states that these models are also models of φ^I. The "only if" direction states that every other model is *not* a model of φ^I. Accordingly the theorem tells us exactly which models the formula has up to modal equivalence.

4.3. A semantic proof of the Modal Reduction Theorem

The idea behind the semantical proof of the Modal Reduction Theorem is to encode part of the semantics into the syntax. Technically we shall work with the filtration of the canonical model, and the argument rests essentially on the existence of a sufficiently rich filtration set. More precisely, we define a *molecule* as a disjunction of atoms; since there are 2^n distinct atoms in our finite language, there are 2^{2^n} non-equivalent molecules. By convention $\bot$ is the disjunction of the empty set of atoms and is hence a molecule. The filtration set Σ which underlies the constructions in this section is the least set of formulae which is closed under subformulae and the following two constraints for each *doc k*:

if α is an atom, then $b_k\alpha$, $c_k\alpha$ and $\Diamond\alpha$ are in Σ,

if μ is a molecule, then $B_k\mu$, $C_k\mu$ and $\Box\mu$ are in Σ.

By construction Σ is $\prec$-closed, and it is hence a filtration set. $M^\dagger$, the filtration of the canonical model wrt. Σ, is defined as in Section 2.3. In the rest of the section all references to semantical constructions are relative to this model.

We first define a syntactical representation of a set of points in $M^\dagger$. If X is a set of points, $[\![X]\!]$ is defined by $\bigvee\{[\![x]\!] \mid x \in X\}$.

Lemma 21. *Let X and Y be subsets of a cluster C (in $M^\dagger$).*

(1) $[\![X]\!] \vdash \neg[\![Y]\!]$ *iff* $X \cap Y = \emptyset$,
(2) $[\![C]\!], \neg[\![X]\!] \vdash [\![Y]\!]$ *iff* $X \cup Y = C$.

Proof. To see the non-trivial direction of (1), note that for two points $x, y \in C$, $\llbracket x \rrbracket \vdash \neg\llbracket y \rrbracket$ iff $x \neq y$. This is because two different points must disagree on a formula in the filtration set. Since x and y belong to the same cluster, they agree on every modal atom and must hence disagree on a propositional letter. For (2), note that $\llbracket C \rrbracket \vdash \llbracket X \rrbracket \vee \llbracket Y \rrbracket$ whenever $X \cup Y = C$ and use simple propositional reasoning. □

Let C be a cluster of $M^\dagger$ and let C_k^+ and C_k^- be the belief part and co-belief part of C wrt. k. Then

$$\beta_k^+(C) = \bigwedge\{b_k\llbracket x \rrbracket \mid x \in C_k^+\} \wedge B_k\llbracket C_k^+ \rrbracket,$$

$$\beta_k^-(C) = \bigwedge\{c_k\llbracket x \rrbracket \mid x \in C_k^-\} \wedge C_k\llbracket C_k^- \rrbracket,$$

$$\lambda(C) = \bigwedge\{\Diamond\llbracket x \rrbracket \mid x \in C\} \wedge \Box\llbracket C \rrbracket.$$

We say that $\beta_k^+(C)$ is the *belief formula of C wrt. k* and $\beta_k^-(C)$ the *co-belief formula of C wrt. k*. $\lambda(C)$ denotes the logical space spanned by the set of atoms true at a point in C.

Lemma 22. *Let C be a cluster and φ be purely Boolean. Then $M^\dagger \vDash_C B_k\varphi$ iff $\llbracket C_k^+ \rrbracket \vdash \varphi$ and $M^\dagger \vDash_C C_k\varphi$ iff $\llbracket C_k^- \rrbracket \vdash \varphi$.*

Proof. We will only prove the first, since the proof of the latter is symmetrical. Suppose $M^\dagger \vDash_C B_k\varphi$. Then $M^\dagger \vDash_x \varphi$ for every $x \in C_k^+$. By Lemma 17(1), $\llbracket C_k^+ \rrbracket \vdash \varphi$. For the other direction, suppose $\llbracket C_k^+ \rrbracket \vdash \varphi$. But by definition, $M^\dagger \vDash_C B_k\llbracket C_k^+ \rrbracket$. Hence normal modal logic gives $M^\dagger \vDash_C B_k\varphi$. □

Lemma 23. *Each cluster C is uniquely characterized by $\beta_k^+(C) \wedge \beta_k^-(C)$.*

Proof. Let D be a cluster distinct from C. Then C and D must, by construction of $M^\dagger$, disagree on a formula in Σ of the form $B_k\varphi$ or $C_k\varphi$. The case where the two clusters disagree on a formula $C_k\varphi$ is symmetric to the case where they disagree on a formula $B_k\varphi$, so we will only treat the B_k-modality.

Assume first that $M^\dagger \vDash_C B_k\varphi$ and $M^\dagger \nvDash_D B_k\varphi$. By Lemma 22, $\llbracket C_k^+ \rrbracket \vdash \varphi$ and $\llbracket D_k^+ \rrbracket \nvdash \varphi$. Then $\llbracket D_k^+ \rrbracket \nvdash \llbracket C_k^+ \rrbracket$, because if $\llbracket D_k^+ \rrbracket \vdash \llbracket C_k^+ \rrbracket$, then $\llbracket D_k^+ \rrbracket \vdash \varphi$. Hence, by Lemma 22, $M^\dagger \nvDash_D B_k\llbracket C_k^+ \rrbracket$. Thus $M^\dagger \nvDash_D \beta_k^+(C) \wedge \beta_k^-(C)$.

Conversely, assume that $M^\dagger \nvDash_C B_k\varphi$ and $M^\dagger \vDash_D B_k\varphi$. Then there is a point x such that $x \in C_k^+$, $x \notin D_k^+$ and $M^\dagger \vDash_x \neg\varphi$. By Lemma 17(1), $\llbracket x \rrbracket \vdash \neg\varphi$. Since $\beta_k^+(C) \vdash b_k\llbracket x \rrbracket$, we get $\beta_k^+(C) \vdash b_k\neg\varphi$. Then $M^\dagger \nvDash_D \beta_k^+(C)$, because if $M^\dagger \vDash_D \beta_k^+(C)$, then $M^\dagger \vDash_D b_k\neg\varphi$, contradicting the assumption. Thus $M^\dagger \nvDash_D \beta_k^+(C) \wedge \beta_k^-(C)$. □

Lemma 24. *Let C be bisected. Then $\vdash \beta_k^+(C) \wedge \beta_k^-(C) \equiv \lambda(C) \wedge O_k\llbracket C_k^+ \rrbracket$ for each* doc k.

Proof. We have to establish four distinct theorems of Æ.

1. $\beta_k^+(C), \beta_k^-(C) \vdash \lambda(C)$. First, let $x \in C$. Then either $x \in C_k^+$ or $x \in C_k^-$. In the first case, $b_k\llbracket x \rrbracket$ is a conjunct of $\beta_k^+(C)$; hence $\beta_k^+(C) \vdash \Diamond\llbracket x \rrbracket$. By a similar argument $\beta_k^-(C) \vdash \Diamond\llbracket x \rrbracket$ in the other case. Second, note that since $\llbracket C_k^+ \rrbracket \vdash \llbracket C \rrbracket$, $B_k\llbracket C_k^+ \rrbracket \vdash B_k\llbracket C \rrbracket$. Similarly $C_k\llbracket C_k^- \rrbracket \vdash C_k\llbracket C \rrbracket$. Hence $\beta_k^+(C), \beta_k^-(C) \vdash \Box\llbracket C \rrbracket$ (by axiom Def□).

2. $\beta_k^+(C), \beta_k^-(C) \vdash O_k\llbracket C_k^+ \rrbracket$. First note that $B_k\llbracket C_k^+ \rrbracket$ is a conjunct of $\beta_k^+(C)$. Likewise $C_k\llbracket C_k^- \rrbracket$ is a conjunct of $\beta_k^-(C)$. By Lemma 21(1) and modal logic note that $C_k\llbracket C_k^- \rrbracket \vdash C_k\neg\llbracket C_k^+ \rrbracket$.

3. $\lambda(C), O_k\llbracket C_k^+ \rrbracket \vdash \beta_k^+(C)$. First, let $x \in C_k^+$. Then $\Diamond\llbracket x \rrbracket$ is a conjunct of $\lambda(C)$. By Def□, $\lambda(C) \vdash b_k\llbracket x \rrbracket \vee c_k\llbracket x \rrbracket$. Since $\llbracket x \rrbracket \vdash \llbracket C_k^+ \rrbracket$, modal logic gives $C_k\neg\llbracket C_k^+ \rrbracket \vdash \neg c_k\llbracket x \rrbracket$. Hence $\lambda(C) \vdash b_k\llbracket x \rrbracket$. Second, trivially $O_k\llbracket C_k^+ \rrbracket \vdash B_k\llbracket C_k^+ \rrbracket$.

4. $\lambda(C), O_k\llbracket C_k^+ \rrbracket \vdash \beta_k^-(C)$. Let $x \in C_k^-$. Since $\llbracket x \rrbracket \vdash \llbracket C_k^- \rrbracket$, $B_k\neg\llbracket C_k^- \rrbracket \vdash \neg b_k\llbracket x \rrbracket$. By Lemma 21(1) and modal logic $B_k\llbracket C_k^+ \rrbracket \vdash B_k\neg\llbracket C_k^- \rrbracket$. Hence $O_k\llbracket C_k^+ \rrbracket \vdash \neg b_k\llbracket C_k^- \rrbracket$. As in the previous case $\lambda(C) \vdash b_k\llbracket x \rrbracket \vee c_k\llbracket x \rrbracket$; hence $\lambda(C) \vdash c_k\llbracket x \rrbracket$. Second, note that Lemma 21(2) and modal logic give $C_k\llbracket C \rrbracket, C_k\neg\llbracket C_k^+ \rrbracket \vdash C_k\llbracket C_k^- \rrbracket$. Clearly, $\lambda(C) \vdash C_k\llbracket C \rrbracket$ and $O_k\llbracket C_k^+ \rrbracket \vdash C_k\neg\llbracket C_k^+ \rrbracket$. Thus $\lambda(C), O_k\llbracket C_k^+ \rrbracket \vdash C_k\llbracket C_k^- \rrbracket$. □

We now complete the semantic proof of the Modal Reduction Theorem: For each logical space λ and O_I-block φ^I, for some $m \geqslant 0$, there are prime O_I-blocks $\psi_1^I, \ldots, \psi_m^I$ such that

$$\lambda \vdash \varphi^I \equiv (\psi_1^I \vee \cdots \vee \psi_m^I).$$

Proof of Theorem 20. By Lemma 4, there is a formula θ with modal depth 1 which is equivalent to φ^I. Assume first that $\theta \in s$, $s \in W^c$. As θ is a Boolean combination of formulae in Σ, and $M^\dagger$ is the filtration of M^c through Σ, we may apply the Filtration Theorem 13 to infer that $M^\dagger \vDash_{|s|} \theta$. Let C be the cluster containing $|s|$. Since $\vdash \theta \equiv \varphi^I$, the Soundness Theorem entails that $M^\dagger \vDash_{|s|} \varphi^I$. By Lemmas 2 and 9, C is bisected. Lemma 24 is then applicable, and we may infer that $M^\dagger \vDash_{|s|} \psi^I(C)$, where $\psi^I(C) = \bigwedge_{k \in I} O_k[\![C_k^+]\!]$. By the Filtration Theorem, $\psi^I(C) \in s$. This proves that $\vdash \varphi^I \supset \psi^I(C_1) \vee \cdots \vee \psi^I(C_m)$, where $C_1, \ldots, C_m$ are all the clusters satisfying φ^I.

For the other direction, let C be a cluster in $M^\dagger$ such that $M^\dagger \vDash_C \theta$, and let $\psi^I(C) = \bigwedge_{k \in I} O_k[\![C_k^+]\!] \in s$. Let D be the cluster containing $|s|$. By the Filtration Theorem, $M^\dagger \vDash_{|s|} \psi^I(C)$, and then also $M^\dagger \vDash_D \psi^I(C)$. By Lemma 2, D is bisected, and Lemma 24 can be used to infer that $M^\dagger \vDash_D \beta_k^+(C) \wedge \beta_k^-(C)$. By Lemma 23, $C = D$, i.e., $|s| \in C$. By assumption, $M^\dagger \vDash_C \theta$, and so $M^\dagger \vDash_{|s|} \theta$. By the Filtration Theorem, $\theta \in s$. We have then proved that $\psi^I(C) \vdash \theta$. As θ is equivalent to φ^I, we have $\psi^I(C) \vdash \varphi^I$ for every cluster C satisfying φ^I, i.e., $\psi^I(C_1) \vee \cdots \vee \psi^I(C_m) \vdash \varphi^I$, where $C_1, \ldots, C_m$ are the clusters satisfying φ^I. □

5. Computational aspects

The size of the logical space is clearly exponential in the number of propositional variables. Explicitly specifying the logical space can, however, be avoided by using a purely Boolean formula as a basis for an *implicit* generation of the logical space. This is addressed in Section 5.1. In Sections 5.2 and 5.3 we once again address the Modal Reduction Theorem, but this time from a syntactical and constructive point of view. More precisely we show how an explicit representation can be obtained from an implicit one by a series of equivalence-preserving rewriting operations. In Section 5.4 we address the computational complexity of this procedure.

5.1. The system Æ$_\rho$

Let ρ be any consistent purely Boolean formula. The system Æ$_\rho$ contains the axioms and inference rules from the definition of Æ, as well as the following for purely Boolean φ.

$$\text{RI:} \quad \Box\rho \qquad \frac{\rho, \varphi \nvdash_{\mathrm{PL}} \bot}{\Diamond\varphi} \quad \text{(RC)}.$$

We shall use $\vdash_\rho$ to denote the deducibility relation of Æ$_\rho$. The formula ρ is called the *characteristic formula* of the system Æ$_\rho$. In the following $\lambda(\rho)$ is shorthand for $\lambda(\widehat{\widehat{\rho}})$. RI and RC were added to ensure the following result.

Lemma 25. $\vdash_\rho \lambda(\rho)$ *for any consistent purely Boolean* ρ.

Proof. $\vdash_\rho \Box \bigvee \widehat{\widehat{\rho}}$ follows by RI, the observation $\vdash \bigvee \widehat{\widehat{\rho}} \equiv \rho$, Lemma 1(4), completeness of Æ, inclusion of Æ in Æ$_\rho$ and the fact that the latter is closed under MP. $\vdash_\rho \Diamond\alpha$, for any $\alpha \in \widehat{\widehat{\rho}}$, follows in a similar manner by RC. □

Example 26. Let ρ be $p \equiv q$. Since $\vdash \rho \equiv (\neg p \wedge \neg q) \vee (p \wedge q)$ we can derive the following.

$$\frac{\rho, \neg p \wedge \neg q \nvdash_{\mathrm{PL}} \bot}{\vdash_\rho \Diamond(\neg p \wedge \neg q)} \qquad \frac{\rho, p \wedge q \nvdash_{\mathrm{PL}} \bot}{\vdash_\rho \Diamond(p \wedge q)} \qquad \vdash_\rho \Box\big((\neg p \wedge \neg q) \vee (p \wedge q)\big).$$

It is easy to see that $\lambda(\rho)$ is $\Diamond(\neg p \wedge \neg q) \wedge \Diamond(p \wedge q) \wedge \Box((\neg p \wedge \neg q) \vee (p \wedge q))$.

Theorem 27. *Let* ρ *be a consistent purely Boolean formula and* φ *any formula. Then* $\lambda(\rho) \vdash \varphi$ *iff* $\vdash_\rho \varphi$.

Proof. *Only if:* Assume $\lambda \vdash \varphi$, i.e., $\vdash \lambda \supset \varphi$ and hence $\vdash_\rho \lambda \supset \varphi$. Now $\vdash_\rho \varphi$ follows from the lemma above and the fact that MP is a rule of Æ$_\rho$. *If:* We show by induction on the proofs of Æ$_\rho$ that $\vdash_\rho \varphi$ implies $\lambda \vdash \varphi$. Now any axiom of Æ$_\rho$ except RI is an axiom of Æ, hence for the basis it suffices to establish $\lambda \vdash \Box\rho$, which follows directly from the observation $\vdash \bigvee \widehat{\widehat{\rho}} \equiv \rho$. There are three induction steps, corresponding to the rules of Æ$_\rho$. RC: If $\rho, \psi \nvdash_{\mathrm{PL}} \bot$, i.e., $\rho \nvdash_{\mathrm{PL}} \neg\psi$, there must (by trivial reasoning on truth tables) also be an atom $\alpha \in \widehat{\widehat{\rho}}$ such that $\alpha \vdash_{\mathrm{PL}} \psi$. Thus $\Diamond\alpha \vdash \Diamond\psi$ and, consequently, $\lambda \vdash \Diamond\psi$. MP: If $\vdash_\rho \psi$ was derived from $\vdash_\rho \varphi \supset \psi$ and $\vdash_\rho \varphi$, then $\lambda \vdash \varphi \supset \psi$ and $\lambda \vdash \varphi$ by the induction hypothesis, and hence $\lambda \vdash \psi$. RN: If $\vdash_\rho \Box\varphi$ was derived from $\vdash_\rho \varphi$, then $\lambda \vdash \varphi$ by the induction hypothesis, hence $\Box\lambda \vdash \Box\varphi$. As λ is completely modalized, $\lambda \vdash \Box\varphi$ follows. □

The validity of certain implications can be carried out completely within propositional logic by using the characteristic formula ρ instead of λ.

Lemma 28. *For any purely Boolean φ, ψ and ρ*:

$$\lambda(\rho) \wedge O_k\varphi \vdash B_k\psi \quad \textit{iff} \quad \rho, \varphi \vdash \psi, \tag{1}$$

$$\lambda(\rho) \wedge O_k\varphi \vdash C_k\psi \quad \textit{iff} \quad \rho, \neg\psi \vdash \varphi, \tag{2}$$

$$\lambda(\rho) \wedge O_k\varphi \vdash O_k\psi \quad \textit{iff} \quad \rho \vdash \varphi \equiv \psi, \tag{3}$$

$$\lambda(\rho) \vdash \Box\psi \quad \textit{iff} \quad \rho \vdash \psi. \tag{4}$$

Proof. We prove the first and leave the others to the reader. The "if" direction follows by standard modal logic. Conversely, assume that $\lambda(\rho) \wedge O_k\varphi \vdash B_k\psi$. If ρ is inconsistent, $\rho, \varphi \vdash \psi$ follows trivially. Otherwise, $\lambda(\rho) \wedge O_k\varphi$ has a model M in which $\|\varphi\| \subseteq \|\psi\|$. By Lemma 17(2), $\widehat{\widehat{\varphi}} \cap \widehat{\widehat{\rho}} \subseteq \widehat{\widehat{\psi}}$. By Lemma 15(2), $\rho, \varphi \vdash \psi$. □

Lemma 29. *Let $i \preceq k$. For any purely Boolean φ, ψ and ρ*:

$$\lambda(\rho) \wedge O_i\varphi \vdash \neg O_k\psi \quad \textit{if } \rho, \psi \nvdash \varphi.$$

Proof. Assume that $\lambda(\rho) \wedge O_i\varphi \wedge O_k\psi$ is consistent. By Lemma 19(2), $\widehat{\widehat{\rho}} \cap \widehat{\widehat{\psi}} \subseteq \widehat{\widehat{\varphi}}$. By Lemma 15(2) $\rho, \psi \vdash \varphi$. □

5.2. Rewriting rules

Generating the λ-expansions of an O_I-block can be done by rewriting it using provable equivalences within the logic. The formula is first *expanded*, then *collapsed*.

To apply the *expand rule* one must select a modal atom β of modal depth 1 and substitute it with $\top$ and $\bot$ in the following way:

$$O_i\varphi \to_\beta \bigl(O_i\varphi[\beta/\top] \wedge \beta\bigr) \vee \bigl(O_i\varphi[\beta/\bot] \wedge \neg\beta\bigr). \tag{E}$$

The soundness of this rule follows immediately from Lemma 3. After using the expand rule, one may apply the *collapse rules*. In the rules below, all occurrences of φ and ψ are propositional and $i \preceq k$. The collapse rules pertaining to B-formulae are:

$$O_i\varphi \wedge B_k\psi \to_\rho O_i\varphi \quad \text{if } \rho \vdash (\varphi \supset \psi), \tag{B^1}$$

$$O_i\varphi \wedge \neg B_k\psi \to_\rho \bot \quad \text{if } \rho \vdash (\varphi \supset \psi), \tag{B^2}$$

$$O_k\varphi \wedge B_i\psi \to_\rho \bot \quad \text{if } \rho \nvdash (\varphi \supset \psi), \tag{B^3}$$

$$O_k\varphi \wedge \neg B_i\psi \to_\rho O_k\varphi \quad \text{if } \rho \nvdash (\varphi \supset \psi) \tag{B^4}$$

to C-formulae:

$$O_k\varphi \wedge C_i\psi \to_\rho O_k\varphi \quad \text{if } \rho \vdash (\neg\psi \supset \varphi), \tag{C^1}$$

$$O_k\varphi \wedge \neg C_i\psi \to_\rho \bot \quad \text{if } \rho \vdash (\neg\psi \supset \varphi), \tag{C^2}$$

$$O_i\varphi \wedge C_k\psi \to_\rho \bot \quad \text{if } \rho \nvdash (\neg\psi \supset \varphi), \tag{C^3}$$

$$O_i\varphi \wedge \neg C_k\psi \to_\rho O_i\varphi \quad \text{if } \rho \nvdash (\neg\psi \supset \varphi) \tag{C^4}$$

and to $\Box$-formulae:

$$O_i\varphi \wedge \Box\psi \to_\rho O_i\varphi \quad \text{if } \rho \vdash \psi, \tag{$\Box^1$}$$

$$O_i\varphi \wedge \neg\Box\psi \to_\rho \bot \quad \text{if } \rho \vdash \psi, \tag{$\Box^2$}$$

$$O_i\varphi \wedge \Box\psi \to_\rho \bot \qquad \text{if } \rho \nvdash \psi, \tag{$\Box^3$}$$

$$O_i\varphi \wedge \neg\Box\psi \to_\rho O_i\varphi \qquad \text{if } \rho \nvdash \psi. \tag{$\Box^4$}$$

As shown below, the rules are sound. They are not complete, as distribution and simplification rules are required. This, however, is outside the scope of the present article.

Lemma 30. *The collapse rules are sound.*

Proof. Let $\lambda = \lambda(\rho)$. For B^1 and B^2 we need to show that

$$\lambda \wedge O_i\varphi \vdash B_k\psi \quad \text{if } \rho \vdash (\varphi \supset \psi),$$

which by Lemma 28(1) is equivalent to "$\lambda \wedge O_i\varphi \vdash B_k\psi$ if $\lambda \wedge O_i\varphi \vdash B_i\psi$", which follows from persistence. For B^3 and B^4 we need to show that

$$\lambda \wedge O_k\varphi \vdash \neg B_i\psi \quad \text{if } \rho \nvdash (\varphi \supset \psi),$$

which follows by Lemmas 19(3) and 28(1) and persistence ($\rho \nvdash \varphi \supset \psi$ yields $\lambda \wedge O_k\varphi \nvdash B_k\psi$ by Lemma 28(1), and as k is the only *doc* involved, Lemma 19(3) now yields $\lambda \wedge O_k\varphi \vdash \neg B_k\psi$, which implies $\lambda \wedge O_k\varphi \vdash \neg B_i\psi$ by persistence). Soundness of the C-rules is proved similarly. For $\Box^1$ and $\Box^2$ we need to show that

$$\lambda \wedge O_i\varphi \vdash \Box\psi \quad \text{if } \rho \vdash \psi,$$

which follows from Lemma 28(4). For $\Box^3$ and $\Box^4$ we need to show that

$$\lambda \wedge O_i\varphi \vdash \neg\Box\psi \quad \text{if } \rho \nvdash \psi,$$

which follows by Lemmas 19 and 28 ($\rho \nvdash \psi$ is equivalent to $\lambda \nvdash \Box\psi$ by Lemma 28(4), and now by a trivial, degenerate special case of Lemma 19(3) this yields $\lambda \vdash \neg\Box\psi$, and hence also $\lambda \wedge O_i\varphi \vdash \neg\Box\psi$). □

5.3. A syntactic proof of the Modal Reduction Theorem

Let β be a modal atom of depth 1, and let v be either $\bot$ or $\top$. Then β/v is said to be a *Boolean binding*. A set of Boolean bindings is a *modal valuation* if it contains no subset of the form $\{\beta/\bot, \beta/\top\}$, i.e., if it never "binds" the same modal atom to conflicting values.

When φ is any formula and σ is the sequence $\langle \beta_1/v_1, \ldots, \beta_m/v_m\rangle$ of Boolean bindings, we will write $\varphi[\sigma]$ for $((\varphi[\beta_1/v_1])\ldots)[\beta_m/v_m]$. Note that a modal valuation only contains bindings for modal atoms of depth 1; thus the set $\{B_k\top/\top, B_kB_k\top/\bot\}$ is not a modal valuation even though it does not (directly) assign conflicting values to the same modal atom. It is disallowed for good reasons, as the two sequences $\langle B_k\top/\top, B_k\top/\top\rangle$ and $\langle B_kB_k\top/\bot\rangle$—which both contain only bindings from this set—yield conflicting values when applied to the formula $B_kB_k\top$. Proper modal valuations are better behaved, as seen from the next lemma. Since their bindings only apply to modal atoms of depth 1, a sequence of such bindings only removes one modal operator at a time, going outwards in the formula. Hence whenever $\varphi[\sigma]$ is purely Boolean, then so is $\psi[\sigma]$ for any subformula ψ of φ. This observation is crucial in the proof of the next lemma.

Lemma 31. *Let V be a modal valuation and let σ and τ be two sequences of bindings from V. Then $\varphi[\sigma] = \varphi[\tau]$ if $\varphi[\sigma]$ and $\varphi[\tau]$ are both purely Boolean.*

Proof. Suppose $\varphi[\sigma]$ and $\varphi[\tau]$ are both purely Boolean, and that σ and τ only contain bindings from V. We show, by induction on ψ, the more general result that $\psi[\sigma] = \psi[\tau]$ for any subformula ψ of φ.

This is trivial for purely Boolean ψ, hence the result holds in the basis. The induction steps for Boolean connectives are also immediate. The induction steps for modal operators are all similar; we consider B_k.

Hence suppose (for the induction hypothesis) that $\psi[\sigma] = \psi[\tau]$, and that $B_k\psi$ is a subformula of φ. By the above observation, $\psi[\sigma]$ and $\psi[\tau]$ are purely Boolean as well. Now let σ_0 and τ_0 be the shortest initial segments of σ and τ, respectively, such that $\psi[\sigma_0] = \psi[\sigma] = \psi[\tau] = \psi[\tau_0]$. Then $(B_k\psi)[\sigma_0] = B_k(\psi[\sigma_0])$ and $(B_k\psi)[\tau_0] = B_k(\psi[\tau_0])$.

Hence both σ and τ contain a binding for $B_k(\psi[\sigma_0])$ to the right of the respective initial segments σ_0 and τ_0. As both bindings occur in V and V is a modal valuation, the two bindings are identical. Thus $(B_k\psi)[\sigma] = (B_k\psi)[\tau]$. □

A modal valuation V is said to be a *modal valuation of* φ if $\varphi[\sigma]$ is purely Boolean for some sequence σ of bindings from V. The above lemma says that this value is independent of the particular σ, provided σ contains sufficiently many bindings to make the result purely Boolean. Hence we shall be permitted to write $\varphi[V]$ for this unique purely Boolean formula when V is a modal valuation of φ.

Note that two distinct modal valuations may produce the same formula, i.e., both $\{B_k p/\top, B_k\top/\top\}$ and $\{B_k p/\bot, B_k\bot/\top\}$ evaluate $B_kB_k p$ to $\top$. Let V be a modal valuation; the following function is useful.

$$\phi(V) = \bigwedge_{\beta/v \in V} (\beta \equiv v).$$

Thus $\phi(V)$ is equivalent to a conjunction of the modal atoms in the bindings in V, negated if bound to $\bot$. A modal valuation of φ is *minimal* if no proper subset is a modal valuation of φ; we write $M(\varphi)$ for the set of minimal modal valuations of φ. By $mod(\varphi)$ we denote the set of modal atoms (of any depth) occurring in φ.

Lemma 32. *If* $|mod(\varphi)| = m$, *then* $|M(\varphi)| \leqslant 2^m$. *Moreover,* $\bigvee_{V\in M(\varphi)} \phi(V)$ *is a tautology.*

Proof. By induction on the number of modal atoms in φ. The basis is obvious, as $M(\varphi) = \{\emptyset\}$ for any purely Boolean φ, and $\phi(\emptyset) = \top$. The induction step follows from the observation that if β is a modal atom of depth 1 occurring in φ, then something is a minimal modal valuation of φ iff for $v = \top$ or $v = \bot$ it is of the form $\{\beta/v\} \cup V$, where V is a minimal modal valuation of $\varphi[\beta/v]$ not containing the binding $\beta/\tilde{v}$, where $\tilde{v}$ is the opposite Boolean value of v. □

Lemma 33. *For any formula* φ *we have the following*:

$$\vdash O_i\varphi \equiv \bigvee_{V\in M(\varphi)} \bigl(O_i\bigl(\varphi[V]\bigr) \wedge \phi(V)\bigr).$$

Proof. By the preceding lemma we have $\vdash O_i\varphi \equiv \bigvee_{V\in M(\varphi)}(O_i\varphi \wedge \phi(V))$, and by repeated applications of Lemma 1(4), each disjunct $O_i\varphi \wedge \phi(V)$ is equivalent to $O_i(\varphi[V]) \wedge \phi(V)$. □

Let $\varphi^I = \bigwedge_{k\in I} O_k\varphi_k$ be an O_I-block and let V be a modal valuation of $\bigwedge_{k\in I}\varphi_k$. Then $\bigwedge_{k\in I} O_k(\varphi_k[V])$ is said to be the *expansion candidate* of φ^I wrt. V. Since V is a modal valuation of each φ_k, the expansion candidates are all prime O_I-blocks. Hence the next lemma takes us almost to the Modal Reduction Theorem:

Lemma 34. *Let* $\varphi^I = \bigwedge_{k\in I} O_k\varphi_k$ *be an* O_I*-block and* $m = |mod(\bigwedge_{k\in I}\varphi_k)|$. *Let* $V_1, \ldots, V_{2^m}$ *be the minimal modal valuations of* $\bigwedge_{k\in I}\varphi_k$ *and* $\psi_1^I, \ldots, \psi_{2^m}^I$ *be the expansion candidates of* φ^I *wrt.* $V_1, \ldots, V_{2^m}$, *respectively. Then*

$$\vdash \varphi^I \equiv \bigvee_{i=1}^{2^m} \bigl(\psi_i^I \wedge \phi(V_i)\bigr).$$

Proof. This is a straightforward generalization of Lemma 33: by Lemma 32 we have $\vdash \varphi^I \equiv \bigvee_{i=1}^{2^m}(\varphi^I \wedge \phi(V_i))$, and by repeated applications of Lemma 1(4), each disjunct $\varphi^I \wedge \phi(V_i)$ is equivalent to $\psi_i^I \wedge \phi(V_i)$. □

From the proofs of Lemmas 32–34 it can be seen that the DNF formulae at the right-hand sides in Lemmas 33 and 34 are derivable from the respective left-hand sides by repeated applications of the *expand* rewriting rule in Lemma 3, together with rules that distribute disjunction over conjunction, and remove any conjunctions containing pairs of opposite literals.

Lemma 35. *Let* λ *be a logical space,* ψ^I *a prime* O_I*-block and* ϕ *any conjunction of modal literals of depth* 1. *Then* $\lambda \vdash (\psi^I \wedge \phi) \equiv \psi^I$ *or* $\lambda \vdash (\psi^I \wedge \phi) \equiv \bot$.

Proof. ψ^I contains some $O_k\psi$, ψ propositional, for every $k \in I$. Hence the result follows by repeated applications of the collapse rules, in conjunction with such propositional rules as associativity and commutativity of conjunction. □

The Modal Reduction Theorem follows directly from the two preceding lemmas.

5.4. Complexity

If a problem is at least as hard as the hardest problem in a complexity class C, it is C-*hard*. A problem in C that is also C-hard, is C-*complete*. The *polynomial hierarchy* of complexity classes is defined as follows: $\Delta_0^p = \Sigma_0^p = \Pi_0^p = \mathrm{P}$, and for all $i \geqslant 0$, $\Delta_{i+1}^p = \mathrm{P}^{\Sigma_i^p}$, $\Sigma_{i+1}^p = \mathrm{NP}^{\Sigma_i^p}$ and $\Pi_{i+1}^p = \mathrm{coNP}^{\Sigma_i^p}$. On the first level of the polynomial hierarchy we find the familiar classes $\Delta_1^p = \mathrm{P}$, $\Sigma_1^p = \mathrm{NP}$ and $\Pi_1^p = \mathrm{coNP}$. Propositional satisfiability and validity are NP- and coNP-complete respectively. On level 2 we find the versions of the level 1 classes that have access to an NP oracle (they can solve any problem in NP in constant time), most notably $\Delta_2^p = \mathrm{P}^{\mathrm{NP}}$, $\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}} = \mathrm{NP}^{\mathrm{coNP}}$, and $\Pi_2^p = \mathrm{coNP}^{\mathrm{NP}}$. A property of the polynomial hierarchy is that determining whether an instance of a problem is in Σ_i^p, is in Π_{i-1}^p. Thus determining whether an instance of a problem is in Σ_2^p, is in coNP.

In order to prove Σ_2^p-membership we need an algorithm which nondeterministically generates a possible expansion, and then with a linear (in the size of the input formula) number of coNP-complete calls, determines whether it really is an expansion.

Algorithm 1. Does the O_I-block φ^I have a $\lambda(\rho)$-expansion in the logic $Æ_\rho$? Nondeterministically generate an expansion candidate ψ^I of φ^I wrt. some modal valuation V. Then determine whether (1) ψ^I is $Æ_\rho$-consistent, and if it is, determine whether (2) $\psi^I \wedge \phi(V)$ is $Æ_\rho$-consistent. If (2) is true, ψ^I is a $\lambda(\rho)$-expansion, otherwise it is not.

Theorem 36. *The problem of determining whether the O_I-block φ^I has a $\lambda(\rho)$-expansion in the logic $Æ_\rho$ is Σ_2^p-complete.*

Proof. *Membership*: Assume that φ^I is of the form $O_1\varphi_1 \wedge \cdots \wedge O_n\varphi_n$. By Lemma 29, condition (1) of Algorithm 1 can be checked with $n - 1$ propositional validity tests: for each $1 \leqslant i < n$, determine whether $\lambda(\rho) \wedge O_i(\varphi_i[V]) \vdash \neg O_{i+1}(\varphi_{i+1}[V])$. By Lemmas 29 and 34, condition (2) can be checked with $|mod(\varphi_1) \cup \cdots \cup mod(\varphi_n)|$ propositional validity tests: for each modal literal β that is a conjunct in $\phi(V)$, determine whether $\lambda(\rho) \wedge O_i(\varphi_i[V]) \vdash \beta$ if β is of the form $B_i\chi$ or $C_i\chi$, and whether $\lambda(\rho) \vdash \beta$ if β is of the form $\Box\chi$.

Hardness: Since $Æ_\top$ is equivalent to the propositional fragment of Levesque's system, determining whether a formula of the form $O_k\varphi$ is satisfiable in $Æ_\top$ is equivalent to determining whether $\{\varphi\}$ has a stable expansion in autoepistemic logic [8], a problem which is Σ_2^p-hard [5]. □

6. Example: Supernormal defaults

The purpose of this section is to demonstrate how the procedure introduced in Section 5 can be applied to the formalization of default theories. We restrict ourselves to the class of *supernormal* defaults, i.e., statements of the form "if φ is consistent, add it to the belief set". Defaults of this form have been studied in traditional Reiter-style default logics, and it is well known that they are particularly well behaved. In particular, a supernormal default theory is guaranteed to have extensions.

In Section 6.1 we use Æ with a single confidence level to formalize an example due to Reiter [12]. This illustrates features of Æ and motivates Section 6.2. In that section we enrich the default representation with priorities, the function of which is to constrain the order in which the defaults are tested. Priorities are implemented by means of confidence levels and the underlying persistence property. The proof of the adequacy of the representation employs the procedure provided in Section 5.

In the literature formalizations of prioritized default theories have been proposed in which the order relation has been given either a descriptive interpretation [1] or a prescriptive interpretation [2]. For supernormal defaults, the two proposals coincide with the formalization addressed in Section 6.2. An encoding of prescriptively ordered default theories into Æ is given in [3] along with further discussion of the subject.

6.1. Default conditionals without priorities

Assume a finite set D of *default names* and a function φ which assigns a purely Boolean formula to each element in D. As usual we write φ_a for $\varphi(a)$.

We shall first represent the default theory which consists of supernormal defaults with consequent $\varphi(a)$, for all $a \in D$. To this end we let the index set I consist of a single *doc* and let $\prec$ be empty. The property corresponding to the statement "the proposition φ holds by default" is formalized by the formula $b\varphi \supset \varphi$ within the scope of the O-modality. We will refer to this formula as a *default conditional* when it occurs within the modal O-context. The purpose of the following function is to individuate default conditionals:

$$\delta(a) = b\varphi_a \supset \varphi_a.$$

Note that $\delta(a)$ is equivalent to $\neg\varphi_a \supset B\neg\varphi_a$, i.e., should φ be false, the subject will believe that it is.

We shall illustrate this default representation by means of an example inspired by Reiter [12, Example 2.1].

Example 37. Let $D = \{a, b, c\}$, $\varphi_a = p$, $\varphi_b = q$, and $\varphi_c = r$. Let $\psi = \kappa \wedge \delta(a) \wedge \delta(b) \wedge \delta(c)$. A novel point in Æ is that we can select different, and illuminating, characteristic formulae to span the logical space, and thereby characterize the joint impact that the logical space and the "knowledge base" κ have on the evaluation of defaults. Below we define four distinct logical spaces and draw some consequences of relevance for the default representation. In each case, κ is satisfiable and without occurrences of p, q, r.

$$\begin{aligned}
&\rho_1 = \kappa && \lambda(\rho_1) \vdash \Diamond(\kappa \wedge p \wedge q \wedge r)\\
&\rho_2 = \kappa \supset (\neg p \vee \neg q) && \lambda(\rho_2) \vdash \neg\Diamond(\kappa \wedge p \wedge q) \wedge \Diamond(\kappa \wedge p \wedge r) \wedge \Diamond(\kappa \wedge q \wedge r)\\
&\rho_3 = \kappa \supset (\neg p \vee \neg q \vee \neg r) && \lambda(\rho_3) \vdash \neg\Diamond(\kappa \wedge p \wedge q \wedge r)\\
&\rho_4 = \kappa \supset \bigl(q \supset (\neg p \wedge \neg r)\bigr) && \lambda(\rho_4) \vdash \neg\Diamond(\kappa \wedge q \wedge p) \wedge \neg\Diamond(\kappa \wedge q \wedge r).
\end{aligned}$$

Noticing the pattern for ρ_1 and ρ_2 the reader will easily figure out entailed possibilities in the latter two cases as well (and not mere impossibilities). A crucial point in the Æ approach to default representation is that the $\Diamond$ operator serves the function of an operator for logical possibility. It can serve this function precisely because the logical space formalizes the notion of logical necessity. The b operator similarly formalizes the notion of *being consistent with what is believed*, which lies at the heart of the formal rendition of defaults by means of default conditionals.

Another central point in Æ is that the evaluation of defaults is carried out within the logic itself in the form of provable equivalents. In this example the reduction to explicit belief representations are reflected in the following theorems of Æ.

$$\begin{aligned}
&\lambda(\rho_1) \vdash O\psi \equiv O(\kappa \wedge p \wedge q \wedge r),\\
&\lambda(\rho_2) \vdash O\psi \equiv O(\kappa \wedge p \wedge r) \vee O(\kappa \wedge q \wedge r),\\
&\lambda(\rho_3) \vdash O\psi \equiv O(\kappa \wedge p \wedge q) \vee O(\kappa \wedge p \wedge r) \vee O(\kappa \wedge q \wedge r),\\
&\lambda(\rho_4) \vdash O\psi \equiv O(\kappa \wedge q) \vee O(\kappa \wedge p \wedge r).
\end{aligned}$$

Recalling the definition of λ-expansion from Section 4.2 we see that there are exactly two distinct expansions for $\lambda(\rho_4)$. For the maximal logical space this notion corresponds exactly to the notion of a stable expansion in autoepistemic logic, cf. [8]. The fact that the O operator permits us to characterize the belief state precisely is essential for the provability of the equivalences.

Conceptually there is an important difference in the way that defaults are treated in Æ (and in autoepistemic approaches in general) and in Reiter-style default logics. To use a conditional for the representation of a default means to state an *invariant* of a belief set. It is equivalences like those above which demonstrate that adoption of such conditionals gives the belief set a behavior which matches intuitions behind default rules. The Æ representation provides a denotational representation of defaults while Reiter provides an operational account.

The belief set in the example above is characterized by a simultaneous belief in different default conditionals. In consequence there is no way of resolving conflicts. This is the subject of the next section.

6.2. Prioritizing supernormal defaults

Let $<$ be a strict partial order on D. The intuition is that if $a < b$, then default a has to be applied before default b. To capture this let an *application sequence* be a finite string over D without repetition of symbols. The concatenation of s and t is denoted st. The *empty sequence* is denoted ε. It has length 0 and is the identity element wrt. concatenation. The *prefix order* $\preceq$ is defined as $s \preceq st$, where s and t are any application sequences (including ε). The *precedence set over* $P = (D, <)$ is defined as the least set T_P such that

(1) $\varepsilon \in T_P$, and
(2) $tb \in T_P$ if $t \in T_P$, b does not occur in t, and every a such that $a < b$ occurs in t.

The following observation is easy to prove.

Lemma 38. *Let $P = (D, <)$. Then $(T_P, \prec)$ is a tree rooted in ε whose nodes are application sequences over D. Each leaf corresponds to a topological sorting of P and every topological sorting corresponds to a leaf in the tree.*

We want to show that these simple concepts provide us with a constructive way of analyzing the generation process of extensions. The following example depicts the graphical structure we obtain from the topologically sorted strings.

Example 39. Let D be as in Example 37, and assume that the extension of the $<$ order is $b < a$ and $b < c$. Let P be $(D, <)$. The corresponding precedence tree is depicted as solid lines in Fig. 5. Dotted lines indicate branches in the tree with an empty order relation; application sequences along dotted lines clash with the $<$-relation of this example.

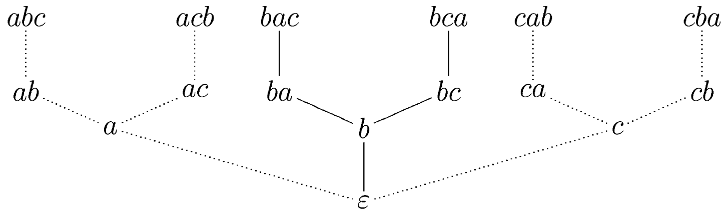

Fig. 5. Solid lines mark $(T_{(D,<)}, \prec)$. For an empty priority relation, include the dotted lines.

Let us now address the assignment of propositional content to the nodes in the tree. A subtle technical detail to this end is that we shall have to alter the definition of a default conditional and say that a default conditional is a formula $b_i\varphi \supset \varphi$ which occurs within the scope of an O_k-modality *such that $i \preceq k$*. The key lemma follows.

Lemma 40. *Let λ be a logical space, and let κ and $\bigwedge_{k\in I}\varphi_k$ be purely Boolean. Then the O_I-block $\bigwedge_{k\in I} O_k(\kappa \wedge \bigwedge_{i\preceq k}(b_i\varphi_i \supset \varphi_i))$ has a unique λ-expansion.*

Proof. Say that $J \subseteq I$ is *closed* if $i \in J$ whenever $i \preceq k$ and $k \in J$. Now writing ζ_k for $(\kappa \wedge \bigwedge_{i\preceq k}(b_i\varphi_i \supset \varphi_i))$, we show that for each closed $J \subseteq I$ there will be some modal valuation V of $\bigwedge_{k\in J} B_k\neg\varphi_k$ such that

$$\lambda \vdash \bigwedge_{k\in J} O_k\zeta_k \equiv \bigwedge_{k\in J} O_k\big(\zeta_k[V]\big) \wedge \phi(V) \equiv \bigwedge_{k\in J} O_k\big(\zeta_k[V]\big).$$

The lemma follows from this, as the special case $J = I$ yields the prime O_I-block $\bigwedge_{k\in I} O_k(\zeta_k[V])$ which by Lemma 19(2) is consistent with λ as each $\zeta_k[V]$ is equivalent to the conjunction $\kappa \wedge \bigwedge\{\varphi_i \mid i \preceq k,\ B_i\neg\varphi_i/\bot \in V\}$, i.e., with non-increasing sets of conjuncts for "greater" doc's k.

The proof is by induction on J. It is trivial for empty J, as $\bigwedge_{k\in\emptyset}\zeta_k \equiv \phi(\emptyset) \equiv \top$. For the induction step, suppose J is non-empty; then $J = J_0 \cup \{j\}$ for some j, J_0 such that $j \preceq k$ for no $k \in J_0$. J_0 is closed since J is, hence by the induction hypothesis there is a modal valuation V_0 of $\bigwedge_{k\in J_0} B_k\neg\varphi_k$ such that

$$\lambda\vdash \bigwedge_{k\in J_0} O_k\zeta_k \equiv \bigwedge_{k\in J_0} O_k\big(\zeta_k[V_0]\big)\wedge\phi(V_0)\equiv \bigwedge_{k\in J_0} O_k\big(\zeta_k[V_0]\big).$$

Now $O_j(\zeta_j[V_0])$ is of the form $O_j(\psi\wedge(b_j\varphi_j\supset\varphi_j))$ for a purely Boolean ψ. By Lemma 3 this is equivalent to the disjunction of

$$O_j\big(\psi\wedge(b_j\varphi_j\supset\varphi_j)\big)[B_j\neg\varphi_j/\top]\wedge B_j\neg\varphi_j,\quad \text{i.e.,}\quad O_j\psi\wedge B_j\neg\varphi_j,\quad \text{and}$$
$$O_j\big(\psi\wedge(b_j\varphi_j\supset\varphi_j)\big)[B_j\neg\varphi_j/\bot]\wedge \neg B_j\neg\varphi_j,\quad \text{i.e.,}\quad O_j(\psi\wedge\varphi_j)\wedge\neg B_j\neg\varphi_j.$$

Now (if we let ρ be such that $\lambda=\lambda(\rho)$) $\rho\vdash\psi\supset\neg\varphi_j$ iff $\rho\vdash(\psi\wedge\varphi_j)\supset\neg\varphi_j$, hence by soundness of the collapse rules exactly one of the two disjuncts reduces to its first conjunct, while the other reduces to $\bot$. Hence either for $v=\top$ or $v=\bot$ we have

$$\lambda\vdash O_j\big(\zeta_j[V_0]\big)\equiv O_j\big(\zeta_j[V_0][B_j\neg\varphi_j/v]\big)\wedge(B_j\neg\varphi_j\equiv v)\equiv O_j\big(\zeta_j[V_0][B_j\neg\varphi_j/v]\big).$$

Putting this together, we see that the following are all equivalent given λ.

$$\bigwedge_{k\in J_0} O_k\zeta_k\wedge O_j\zeta_j,$$

$$\bigwedge_{k\in J_0} O_k\big(\zeta_k[V_0]\big)\wedge\phi(V_0)\wedge O_j\zeta_j,$$

$$\bigwedge_{k\in J_0} O_k\big(\zeta_k[V_0]\big)\wedge\phi(V_0)\wedge O_j\big(\zeta_j[V_0]\big),$$

$$\bigwedge_{k\in J_0} O_k\big(\zeta_k[V_0]\big)\wedge\phi(V_0)\wedge O_j\big(\zeta_j[V_0][B_j\neg\varphi_j/v]\big)\wedge(B_j\neg\varphi_j\equiv v),$$

$$\bigwedge_{k\in J_0} O_k\big(\zeta_k[V_0]\big)\wedge O_j\big(\zeta_j[V_0][B_j\neg\varphi_j/v]\big).$$

Setting $V=V_0\cup\{B_j\neg\varphi_j/v\}$, we obtain $\zeta_k[V_0]=\zeta_k[V]$ for any $k\in J_0$, while $\zeta_j[V_0][B_j\neg\varphi_j/v]=\zeta_j[V]$. Thus the first and two last of the above can be identified as the three formulae that were to be shown equivalent given λ. □

Let us now address the encoding of the prioritized default theory in Æ. The idea is to use T_P as the index set I in the signature of Æ and use the prefix ordering $\preceq$ to distinguish degrees of confidence. When we interpret modalities as application sequences we employ them as devices for protecting information. We extend the δ function in an interesting way:

$$\delta(sa)=b_{sa}\varphi_a\supset\varphi_a.$$

Note that it is the last term in the application sequence which selects the particular φ, while the modal context of the consistency check is given by the whole term.

On the basis of Lemma 40 we propose the following representation of the prioritized default theory $P=(D,<)$ with assignment function φ.

$$[\![P,\kappa]\!]_\varepsilon=O_\varepsilon\kappa,$$
$$[\![P,\kappa]\!]_{sa}=O_{sa}\Bigg(\kappa\wedge\bigwedge_{tb\preceq sa}\delta(tb)\Bigg),$$
$$[\![P,\kappa]\!]=\bigwedge_{t\in T_P}[\![P,\kappa]\!]_t.$$

Theorem 41. *$[\![P,\kappa]\!]$ has a unique λ-expansion.*

Proof. Follows immediately from Lemma 40 when we use T_P as index set I, put $\varphi_{sa} = \varphi_{ta}$ if sa and ta are both in T_P, and put $\varphi_\epsilon = \top$. □

The maximal nodes in the unique λ-expansion of $[\![(D, <), \kappa]\!]$ play a central role as they correspond to the "final" beliefs up to $<$. We close the article by addressing our example once again.

Example 42. Continuing Example 39 we note that $[\![P, \kappa]\!]$ is

$$\begin{aligned}&O_\varepsilon\kappa \wedge O_b\big(\kappa \wedge \delta(b)\big)\\ &\quad \wedge O_{ba}\big(\kappa \wedge \delta(b) \wedge \delta(ba)\big) \wedge O_{bac}\big(\kappa \wedge \delta(b) \wedge \delta(ba) \wedge \delta(bac)\big)\\ &\quad \wedge O_{bc}\big(\kappa \wedge \delta(b) \wedge \delta(bc)\big) \wedge O_{bca}\big(\kappa \wedge \delta(b) \wedge \delta(bc) \wedge \delta(bca)\big),\end{aligned}$$

which has the unique $\lambda(\rho_4)$-expansion

$$O_\varepsilon\kappa \wedge O_b(\kappa \wedge q) \wedge O_{ba}(\kappa \wedge q) \wedge O_{bac}(\kappa \wedge q) \wedge O_{bc}(\kappa \wedge q) \wedge O_{bca}(\kappa \wedge q).$$

7. History and related work

The present work extends the conference paper [10]. It was initiated by the first author's doctoral thesis [17], in which the Modal Reduction Theorem for the system $Æ_\top$ was first established. The theorem for Levesque's system has later been discovered independently by Levesque and Lakemeyer and appears as Corollary 9.5.6 in [9]; their proof is similar in style to the proof in Section 5 but with a less general transformation strategy. The construction in the semantical proof of the Modal Reduction Theorem is inspired by a note of Segerberg [15] written in response to [17].

The present work has been extended to a multi-modal language. In [19] the model theory of $Æ_\top$ is generalized to the multi-modal case, while a proof theory for the multi-modal extension of $Æ_\top$ is given in [18]. This includes cut-elimination results for a sequent calculus formulation of the logic. A sequent calculus for the logic addressed in this article is trivially obtained by restricting the language in [18] to a single agent.

In the language of Æ formulated in this paper, it is not possible to express properties about indices in I or about the preference relation $\prec$ within the language. It would be interesting to see whether the techniques of term-modal logics [4] can be applied also to Æ, possibly with a cautious introduction of quantifiers. If the language is extended to decidable fragments of first-order logic, it can presumably still be used to represent defaults along the lines sketched in this paper. The point is that the term universe must be finite. We must also restrict the language to formulae in Σ_0^1, like $\Diamond\exists x\,\mathrm{bird}(x)$, and formulae in Π_0^1 like $\Box\forall x(\mathrm{penguin}(x) \supset \mathrm{bird}(x))$. Such formulae do not generate new terms and hence reduce to purely Boolean logic. In this way the system can be extended to restricted fragments of first-order logic which nevertheless are sufficient for describing a number of common-sense situations. Of course, if we extend the language to full first-order logic, the system can no longer be used to represent default reasoning along the lines sketched in this paper. In general it will then impossible be to represent the space of conceivability with a finite formula. In fact such systems suffer from a fundamental incompleteness property [6].

References

[1] G. Brewka, T. Eiter, Prioritizing default logic, in: S. Hölldobler (Ed.), Intellectics and Computational Logic, Papers in Honor of Wolfgang Bibel, in: Applied Logic Series, vol. 19, Kluwer Academic Publishers, Dordrecht, 2000, pp. 27–45.

[2] J.P. Delgrande, T. Schaub, Expressing preferences in default logic, Artificial Intelligence 123 (2000) 41–87.

[3] I. Engan, T. Langholm, E.H. Lian, A. Waaler, Default reasoning with preference within only knowing logic, in: Proceedings of LPNMR'05, Lecture Notes in Artificial Intelligence, vol. 3662, 2005, pp. 304–316.

[4] M. Fitting, L. Thalmann, A. Voronkov, Term-modal logics, Studia Logica 69 (2001) 133–169.

[5] G. Gottlob, Complexity results for nonmonotonic logics, Journal of Logic and Computation 2 (3) (1992) 397–425.

[6] J.Y. Halpern, G. Lakemeyer, Levesque's axiomatization of only knowing is incomplete, Artificial Intelligence 74 (1995) 381–387.

[7] W. Lenzen, Recent Work in Epistemic Logic, Acta Philosophica Fennica, vol. 30, North-Holland, Amsterdam, 1978.

[8] H.J. Levesque, All I know: A study in autoepistemic logic, Artificial Intelligence 42 (1990) 263–309.

[9] H.J. Levesque, G. Lakemeyer, The Logic of Knowledge Bases, The MIT Press, Cambridge, MA, 2000.

[10] E.H. Lian, T. Langholm, A. Waaler, Only knowing with confidence levels: Reductions and complexity, in: J.J. Alferes, J. Leite (Eds.), Proceedings of JELIA'04, in: Lecture Notes in Artificial Intelligence, vol. 3225, Springer, Berlin, 2004, pp. 500–512.

[11] I. Pörn, On the nature of emotions, in: P. Needham, J. Odelstad (Eds.), Changing Positions: Essays dedicated to Lars Lindahl on the occasion of his fiftieth birthday, in: Philosophical Studies, vol. 38, Uppsala Universitet, Uppsala, 1986, pp. 205–214.

[12] R. Reiter, A logic for default reasoning, Artificial Intelligence 13 (1980) 81–132.

[13] R. Rosati, On the decidability and complexity reasoning about only knowing, Artificial Intelligence 116 (2000) 193–215.

[14] R. Rosati, A sound and complete tableau calculus for reasoning about only knowing and knowing at most, Studia Logica 69 (2001) 171–191.

[15] K. Segerberg, Some modal reduction theorems in autoepistemic logic, Uppsala Prints and Preprints in Philosophy, Uppsala University, 1995.

[16] M. Sergot, Normative positions, in: P. McNamara, H. Prakken (Eds.), Norms, Logics and Information Systems, IOS Press, Amsterdam, 1999, pp. 289–308.

[17] A. Waaler, Logical studies in complementary weak S5, Doctoral thesis, University of Oslo, 1994.

[18] A. Waaler, Consistency proofs for systems of multi-agent only knowing, Advances in Modal Logic 5 (2005) 347–366.

[19] A. Waaler, B. Solhaug, Semantics for multi-agent only knowing (extended abstract), in: R. van der Meyden (Ed.), Proceedings of TARK X, ACM Digital Library, 2005, pp. 109–125.