AUTOEPISTEMIC MODAL LOGICS

Grigori Shvarts
Program Systems Institute
of the USSR Academy of Sciences
152140 Pereslavl-Zalessky, USSR

ABSTRACT

A modal approach to nonmonotonic reasoning was proposed by Drew McDermott and Jon Doyle in 1980–82. Almost immediately some disadvantages of that approach were pointed out. Robert Moore (1983) proposed his autoepistemic logic, which overcomes these difficulties. Later, some authors (Kurt Konolige, Paul Morris and others) found peculiarities of different kinds in Moore's logic and proposed rather complicated solutions to these problems. A careful mathematical analysis of Moore's and McDermott's approaches shows that Moore's logic is merely a special case of McDermott's logic, at least formally. The problems that arose in Moore's logic may find a simple and uniform solution by going back to McDermott's original concept.

INTRODUCTION

Moore [1] introduced autoepistemic logic for formalising reasoning of an agent which may contain references to the agent's own knowledge (or belief). This kind of reasoning, which Moore calls *autoepistemic*, has the nonmonotonicity property: The set of "theorems" does not increase with the set of "axioms." (Moore attributes this observation to Stalnaker's work [2] which is not available to the author.) The language of Moore's logic is the usual propositional language augmented by the modal operator L. The intended interpretation of $L\phi$ is: "the rational agent believes (or knows) ϕ". Because of nonmonotonicity, the set of autoepistemic consequences of a given premisses cannot be defined as the set of sentences obtained from the premisses by applying some axioms and inference rules. Instead, Moore [1] introduced the following fixed point construction.

Let A be any set of sentences in the modal propositional language. A set of sentences T is said to be a *stable expansion* of A iff

$$T = \{\psi : A \cup \{L\phi : \phi \in T\} \cup \{\neg L\phi : \phi \notin T\} \vdash \psi\}. \tag{1}$$

The sign $\vdash$ denotes here the usual tautological consequence relation. The stable expansion of A may be described informally as the set of beliefs of an ideal rational agent on the basis of the premisses A. Two sets of formulas added in (1) to A are produced by "positive introspection" ($\{L\phi : \phi \in T\}$) and "negative introspection" of the agent.

Moore's work was preceded by McDermott and Doyle's work [4], who attempted to formalise default reasoning, another important form of non-monotonic reasoning. The informal interpretation of $L\phi$ in [4] is "ϕ is provable." (As a primary modal operator, the dual operator $M = \neg L \neg$ is used in [4]; we use L for convenience.) The basic notion of [4] is the *fixed point* of a set A: T is a fixed point of A iff

$$T = \{\psi : A \cup \{\neg L\phi : \phi \notin T\} \vdash \psi\}. \tag{2}$$

Thus stable expansions differ from fixed points by the presence of the term $\{L\phi : \phi \in T\}$ in the right-hand side of the fixed point equation. Informally, fixed points are the possible sets of non-monotonic consequences of A. However, McDermott and Doyle's logic has some peculiarities. The most serious one is that a set of formulas may have a consistent fixed point containing both p and $\neg Lp$, which contradicts the intended interpretation of L.

McDermott [5] fixed this defect of the definition (2) by replacing the provability $\vdash$ in the classical propositional logic by the provability in some system of modal logic (with L identified with the necessity operator). He considered three well-known modal systems as possible bases for non-monotonic logic, namely the systems T, S4 and S5. But only the case of S5 was investigated in [5] in sufficient detail. It turned out that there are too many fixed points in this case; even the empty set (pure non-monotonic S5) has infinitely many fixed points. Moreover, the intersection of all fixed points of A is just the set of all monotonic S5-consequences of A, so that non-monotonic S5 collapses, in some sense, to the monotonic S5. But the non-monotonic T and S4 remained uninvestigated in [5], although, as we show in this paper, they have nice properties.

Moore [1] argued that McDermott and Doyle's logics are logics of autoepistemic reasoning, rather than of default reasoning, and considered his own logic a reconstruction of McDermott and Doyle's logic.

Later, some authors pointed out some defects of Moore's logic. Such defects are of two kinds: first, some sets of formulas have superfluous, or "ungrounded" stable expansions (Konolige [6]); second, some simple theories do not have stable expansions (Morris [7]). Solutions to these problems have been proposed. In order to get rid of superfluous expansions, Konolige [6] introduced the notions of moderately grounded and of strongly grounded extension, which are the strenghenings of the notion of stable expansion. On the other hand, Morris [7] introduced the notion of stable closure, which is a generalisation of the notion of stable expansion.

In this paper we argue that McDermott's non-monotonic modal logics may be viewed as autoepistemic logics, and Moore's logic is one of them, although the most important one. Many problems arising in Moore's logic may be solved within McDermott's logic by an appropriate choice of the underlying modal system. In particular, if we take S4 as the underlying system, then the ungrounded extensions found by Konolige disappear; and the additional extensions introduced by Morris take their place.

Accordingly, we shall call McDermott's non-monotonic modal logics the autoepistemic (modal) logics.

The paper is organized as follows. After some preliminaries in Section 1, we prove in Section 2 that Moore's logic is exactly McDermott's logic based on the modal logic known as "weak S5." In Section 3, the complete description of fixed points for McDermott's logics based on modal logics K, T, S4 and weak S5 is given. The proofs of propositions are collected in Appendix. The theorems are formulated in Appendix in more general (and more technical) form. This enables us to clarify the reasons for the differences between the nonmonotonic logics based on different modal logics; on the other hand, the reader interested only in applications to AI can avoid relatively complicated technical machinery. In Section 4 we apply the general results of Section 3 to concrete situations. In particular, we show, that using S4 as a basis for autoepistemic logic enables to avoid some difficulties appeared in Moore's logic.

1 PRELIMINARIES

We consider the usual propositional modal language with the logical connectives $\vee, \wedge, \supset, \neg$ and with the modal operator L (necessity). All modal logics in question have two inference rules: modus ponens $(\phi, \phi \supset \psi / \psi)$ and necessitation $(\phi / L\phi)$. Their axioms include all instances of propositional tautologies and some axiom schemata from the following list:

K $$L(\phi \supset \psi) \supset (L\phi \supset L\psi)$$

T $$L\phi \supset \phi$$

4 $$L\phi \supset LL\phi$$

5 $$\neg L\phi \supset L\neg L\phi.$$

K is the modal logic based on the single axiom schema K. T is K together with T, S4 is T together with 4, S5 is S4 together with 5, K45 (also called weak S5) is S5 without T.

If S is a logic and Γ is a set of sentences, then $\Gamma \vdash_S \phi$ means that ϕ is deducible from Γ by means of the axioms and inference rules of S. If S is classical propositional logic then we omit the subscript S.

By a *modal logic* we mean any logic in the modal propositional language which contains all instances of propositional tautologies, and whose inference rules are just modus ponens and necessitation.

NB. Some authors (e.g. Chellas [8], Konolige [6]) use $\Gamma \vdash_S \phi$ in a stronger sense: they mean that for some finite subset $\{A_1, \ldots, A_n\}$ of Γ, $\vdash_S (A_1 \wedge \ldots \wedge A_n) \supset \phi$. It is important to distinguish our understanding of $\vdash_S$ from the stronger one. For instance, we have $p \vdash Lp$, and in the stronger sense this this is not true. Our definition follows McDermott [5].

A *Kripke model* is a triple $\mathcal{M} = \langle M, R, V \rangle$, where M is a nonempty set (called the *set of worlds)*, R is a binary relation on M, and for each $\alpha \in M$, $V(\alpha)$ is a set of propositional variables, which are said to be *true* in the world α. The forcing relation $\langle \mathcal{M}, \alpha \rangle \models \phi$ (the formula ϕ is true in the world α of the model M) is defined inductively as follows: $\langle \mathcal{M}, \alpha \rangle \models p$ iff $p \in V(\alpha)$; $\langle \mathcal{M}, \alpha \rangle \models \phi \wedge \psi$ $(\phi \vee \psi)$ iff $\langle \mathcal{M}, \alpha \rangle \models \phi$ and (or) $\langle \mathcal{M}, \alpha \rangle \models \psi$; $\langle \mathcal{M}, \alpha \rangle \models \neg\phi$ iff not $\langle \mathcal{M}, \alpha \rangle \models \phi$; $\langle \mathcal{M}, \alpha \rangle \models L\phi$ iff for each β with $\alpha R \beta$, $\langle \mathcal{M}, \beta \rangle \models \phi$.

We write $\alpha \models \phi$ if it is clear from the context which $\mathcal{M}$ is meant.

We write $\mathcal{M} \models \phi$ iff for each $\alpha \in M$, $\langle \mathcal{M}, \alpha \rangle \models \phi$.

Any Kripke model is also called a K-*model*. If R is reflexive, then the K-model is called a T-*model*; if R is reflexive and transitive, then $\mathcal{M}$ is called an S4-*model*; if R is universal on M (i.e., $R = M \times M$), then $\mathcal{M}$ is called an S5-*model*. If R is transitive and Euclidean (i.e., $\alpha R \beta$ and $\alpha R \gamma$ implies $\beta R \gamma$), then $\mathcal{M}$ is called a K45-*model*. $\Gamma \models_S \phi$ means that for each S-model $\mathcal{M}$, if $\mathcal{M} \models \Gamma$ then $\mathcal{M} \models \phi$. It is well known that, if S is one of K, T, S4, S5 and K45, then, for each Γ and ϕ, $\Gamma \models_S \phi$ iff $\Gamma \vdash_S \phi$ (completeness theorems, see e.g. Chellas [8], McDermott [5]). (In monographs on modal logic, the completeness theorem is usually proved in a weaker form, for the empty Γ only. McDermott deduced the form we need from this weaker form. His method is applicable to any normal mod

A set T of formulas is called an *S-extension* of a set A iff

$$T = \{\psi : A \cup \{\neg L\phi : \phi \notin T\} \vdash_S \psi\}$$

Proposition 1.1. (McDermott [5]). *Let S, T be any normal modal logics contained in S5, and $S \subseteq T$. Then each S-extension of A is also a T-extension of A.*

T is called *stable* iff: (i) T is closed under tautological consequence, (ii) for each ϕ, if $\phi \in T$ then $L\phi \in T$, and (iii) for each ϕ, if $\phi \notin T$ then $\neg L\phi \in T$. Clearly, for each normal logic S, each S-extension is a stable set. T is called an S5-*set* iff for some S5-model $\mathcal{M}$, $T = \{\phi : \mathcal{M} \models \phi\}$. Moore [9] has proved that a consistent set is stable iff it is an S5-set.

An *objective formula* is a formula not containing occurrences of L.

2 MOORE'S LOGIC AND K45

Moore [1, p. 89] established the following connection between his logic and K45: T is a stable expansion of A iff

$$T = \{\psi : A \cup \{L\phi : \phi \in T\} \cup \{\neg L\phi : \phi \notin T\} \vdash_{K45} \psi\}.$$

Konolige [6] strengthened this result: T is a stable expansion of A iff

$$T = \{\psi : A \cup \{L\phi : \phi \in T_0\} \cup \{\neg L\phi : \phi \in \overline{T}_0\} \vdash_{K45} \psi\},$$

where T_0 is the set of all objective formulas in T, and $\overline{T}_0$ is the set of all objective formulas not in T.

Remark. Konolige writes $A \cup LA$ instead of A since he uses $\vdash$ in the stronger sense, see NB in Section 1. (3) is equivalent to Konolige's result since $\vdash_{K45} \phi \supset L\phi$ for each modalized ϕ.

Proposition 2.1. *If T is consistent, then T is a stable expansion of A iff T is a K45-extension of A.*

So Moore's logic may be considered a special case of McDermott's logic.

Remark. Proposition 2.1 fails for an inconsistent T. For instance, the theory $\{\neg Lp\}$ has two stable expansions: a consistent one, which does not contain p, and an inconsistent one (the set of all formulas). Only the consistent expansion is a K45-extension. In the rest of the paper we shall write "stable expansion" for "consistent stable expansion".

3 FORMAL PROPERTIES OF AUTOEPISTEMIC MODAL LOGICS

Konolige [6, p. 355] described a construction which enables us to construct a stable set W_A containing a given set A of objective formulas, such that W_A is a stable expansion of A, and the objective formulas in W_A are exactly the tautulogical consequences of A. W_A is described as an S5-set. We generalize this construction to arbitrary A and investigate in which cases the resulting stable set is an S-extension, and when all S-extensions can be constructed in this way.

By A^L we denote the set of all subformulas of (the elements of) A that begin with L. *A-formulas* are the formulas constructed from the elements of A^L and propositional variables by means of propositional connectives. If Φ is a set of formulas then $\neg\Phi$ denotes $\{\neg\phi : \phi \in \Phi\}$.

Let W be a consistent stable set containing A. Let $\Psi = A^L \cap W$, and $\Phi = A^L \setminus \Psi$. Obviously, if $L\psi \in W$ then $\psi \in W$, and if $L\phi \in \Phi$ then $\phi \notin \Phi$. Hence $A \cup \Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$ is consistent,contained in W and does not imply ϕ for any $L\phi \in \Phi$ in the propositional calculus. This motivates the following definition.

Definition 3.1. Let $\Phi \subseteq A^L, \Psi = A^L \setminus \Phi$. Φ is said to be *admissible for A* iff $A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$ is propositionally consistent and for each $L\phi \in \Phi$, ϕ is not a tautological consequence of $A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$.

Let V be any valuation assigning the truth values 1 (true) and 0 (false) to propositional letters, and let Φ be any set of formulas beginning with L. Define $V_\Phi(p) = V(p)$ if p is a propositional variable, $V_\Phi(L\phi)$ =(0 if $L\phi \in \Phi$, 1 otherwise), and extend V_Φ to all formulas of modal language by means of the usual truth-tables.

Definition 3.2. Let Φ be admissible for A, $\Psi = A^L \setminus \Phi$. Define the *$\langle A, \Phi\rangle$-generated S5-model* $\mathcal{N}_{A,\Phi} = \langle N_{A,\Phi}, R, U_{A,\Phi}\rangle$ as follows. Set $N_{A,\Phi}$ to be the set of all valuations V such that $V_\Phi(A) = 1$ and for each $L\psi \in \Psi$, $V_\Phi(\psi) = 1$. Set R to be universal on $N_{A,\Phi}$, and for $V \in N_{A,\Phi}$ set $U_{A,\Phi}(V) = \{p : V(p) = 1\}$. Denote $W_{A,\Phi} = \{\phi : \mathcal{N}_{A,\Phi} \models \phi\}$.

$N_{A,\Phi}$ is nonempty, since $A \cup \{\neg L\phi : L\phi \in \Phi\} \cup \Psi \cup \{\psi : L\psi \in \Psi\}$ is propositionally consistent. Consequently, the model $\mathcal{N}_{A,\Phi}$ is well-defined.

Lemma 3.1. *If Φ is admissible for A and ϕ is an A-formula, then for each $V \in N_{A,\Phi}$, $\langle \mathcal{N}_{A,\Phi}, V\rangle \models \phi$ iff $V_\Phi(\phi) = 1$.*

Corollary 3.1. *$W_{A,\Phi}$ is a stable set containing $A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$.*

Corollary 3.2. *$A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$ is consistent with S5.*

Each Φ admissible for A defines a theory $W_{A,\Phi}$ which is stable and contains A. First we shall try to investigate in which cases $W_{A,\Phi}$ is an S-extension of A.

Definition 3.3. Let $\Phi \subseteq A^L$, $\Psi = A^L \setminus \Phi$. Φ is said to be *S-admissible for A* iff it is admissible for A and, in addition, for each $L\psi \in \Psi$, $A \cup \neg\Phi \vdash_S \psi$.

Theorem 3.1. *Let S be any modal logic contained in S5, and let Φ be S-admissible for A. Then $W_{A,\Phi}$ is an S-extension of A.*

Does a set Φ, S-admissible for A, uniquely determine the S-extension of A containing $\neg\Phi$? If S is S5 then this this not the case: Even the empty set (which is S5-admissible for itself) has infinitely many S5-extensions. But for many other logics this is the case.

Theorem 3.2. *Let S be any of K, T, S4, K45, and let Φ be S-admissible for A. Then $W_{A,\Phi}$ is the unique S-extension of A containing $A \cup \neg\Phi$.*

Note on the proof. The proof (see Appendix) is rather complicated and uses Kripke models for S. The key property of S which enables us to obtain the result is the following one. Let $\mathcal{M} = \langle M, R, V\rangle$ be an S-model, and let $\alpha \notin M$. Then for some R^* such that

$$R \cup (\{\alpha\} \times Rg(R)) \subseteq R^* \subseteq (\{\alpha\} \times (\{\alpha\} \cup M)) \cup R,$$

$< M \cup \{\alpha\}, R^*, V^* >$ is an S-model for each V^*, too. S5-models do not possess this property, K,T and S4 possess it trivially, and for K45 we can achieve it by a slight modification of the notion of K45-model (preserving, of course, the completeness theorem). The theorem holds for each modal logic possessing the above-mentioned property.□

Another question is whether every S-extension is $W_{A,\Phi}$ for some S-admissible for A set Φ.

Theorem 3.3. *Let S be K,T or S4. Then each S-extension of A equals $W_{A,\Phi}$ for some S-admissible Φ.*

Note on the proof. In addition to the property mentioned above, we need here the following closure property: If $< M, R, U >$ and $< N, Q, V >$ with $M \cap N = \emptyset$ are S-models, then

$$\langle M \cup N, R \cup (M \times N) \cup Q, U \cup V \rangle$$

is an S-model too. K, T and S4 possess this property trivially, but K45 does not. Again the theorem holds for each modal logic possessing these two properties.□

Theorem 3.3 is wrong for K45—for example, $A = \{Lp \supset p\}$ has a stable expansion $W_{A,\emptyset}$, but $\emptyset$ is not K45-admissible for A.

Definition 3.4. Let Φ be admissible for A, $\Psi = A^L \setminus \Phi$. Φ is said to be *propositionally admissible for A* iff for each $L\psi \in \Psi$, ψ is a tautological consequence of $A \cup \neg\Phi \cup \Psi$.

Theorem 3.4. *T is a stable expansion (or K45-extension) of A iff T is $W_{A,\Phi}$ for some Φ which is propositionally admissible for A.*

Theorem 3.4 can be obtained as a corollary to Theorems 1.1 and 1.2, see Appendix. In [10] we presented a simpler proof without any use of modal logic.

4 APPLICATIONS

Proposition 4.1. *Let S be any modal logic for which Theorem 3.2 is true, and let A be any set of objective sentences consistent with S. Then A has $W_{A,\emptyset}$ as its unique S-extension.*

Proposition 4.1 for the empty A was proved in [11] (directly, not as a corollary to the general theorems from the previous section).

For Moore's logic, this fact was established by Konolige [6]. Thus, for objective axioms there is no difference between autoepistemic K, T, S4, K45. But for arbitrary axioms all these logics are different.

The following two examples are taken from [11] too, but here we explain them more simply using the general results of Section 2.

Example 4.1. Let $A = \{L(Lp \supset LLp) \supset p\}$. In S4 this formula is equivalent to p. So, by Proposition 4.1, A has a unique S4-extension, namely $T = W_{p,\emptyset}$. Let A have a T-extension. Then, by Proposition 1.1, it must be T. Hence, by Proposition 3.2, $\emptyset$ is T-admissible for A, i.e., $A \vdash_{\mathrm{T}} p$, which is wrong. So A has no T-extensions.

Example 4.2. $B = \{\neg Lp \supset p\}$ is known to have no stable expansions. But $B \vdash_{\mathrm{T}} p$, hence $\emptyset$ is T-admissible for B, and B has a T-extension.

Example 4.3.. Consider the theory $C = \{Lp \wedge Lq \supset p \wedge q, \neg Lp \supset p\}$. Since $C \cup \{Lp, Lq\} \vdash p \wedge q$, $\emptyset$ is propositionally admissible for C, so $W_{C,\emptyset}$ is a stable expansion of C. It may be easily shown that there is no other set propositionally admissible for C. On the other hand, $\{Lq\}$ is the only set S4-admissible for C, so that $W_{C,Lq}$ is the only S4- extension of C. Thus, C has one S4-extension and one stable expansion, and the two are different.

$A = \{Lp \supset p\}$ has two K45-extensions, namely $T_1 = W_{A,\{Lp\}}$ and $T_2 = W_{A,\emptyset}$, but only the former is an S4-extension. These facts may be established using the results of Section 3.

Konolige [6] considered two examples:

$$A = \{Lp \supset p\},\ B = \{\neg Lp \supset q, Lp \supset p\}.$$

A has two stable expansions, T_1 and T_2, described above. Konolige considered T_2 an "anomalous" extension, since then "the agent's belief in P is grounded in her assumption that she believes P." In order to avoid this situation, Konolige introduced the notion of moderately grounded extension, and T_2 turned out to be not moderately grounded.

B has two extensions: $S_1 = W_{B,\{Lp\}}$ and $S_2 = W_{B,\emptyset}$. Konolige regards S_2 as anomalous too, but both S_1 and S_2 are moderately grounded. To eliminate S_2, he introduced the concept of strongly grounded extension—a rather complicated definition possessing some undesirable properties.

We suggest another possibility. If we consider, e.g., S4-extensions instead of K45-extensions, then T_2 and S_2 fail to be extensions, but T_1 and S_1 remain.

Let us consider, on the other hand, some examples of the lack of extensions. Morris [7] gives two such examples (we have simplified his notation).

The axiom set

$$C = \{\neg LB \supset A,\ \neg LA \supset \neg F,\ \neg LB \supset F, \neg F\}$$

is a "simplified taxonomy example". Here A should be understood as "Tweety is an abnormal animal", B should be understood as "Tweety is an abnormal bird", F should be understood as "Tweety can fly".

The second example is

$$D = \{\neg Lq \supset p, \neg Lr \supset \neg p\}$$

(the "Nixon paradox"; q stands for "Nixon is a Quaker", r stands for "Nixon is a Republican", p stands for "Nixon is a pacifist").

Morris considers it anomalous that C and D do not have stable expansions. He introduced the notion of stable closure. C has one stable closure, and D has two. Using the results of Section 3, we conclude that C has the unique S4-extension $W_{C,\{LA\}}$, and D has two S4-extensions, $W_{D,\{Lq\}}$ and $W_{D,\{Lr\}}$. All these extensions exactly coincide with Morris's stable closures.

Lifschitz [personal communication] considered it unsatisfactory that $\{Lp\}$ has no stable expansions. Again, $\{Lp\}$ has the unique S4-extension, which contains Lp and p.

Thus the use of S4-extensions instead of stable expansions allows us to avoid some "ungrounded" extensions and, on the other hand, to add some new extensions required in applications. But we do not assert that autoepistemic S4 is always "better" than Moore's logic. We only wanted to demostrate that the problems of different kind may get a uniform solution. We think that different aspects of autoepistemic reasoning should be reflected by different formalisations.

APPENDIX. PROOFS OF PROPOSITIONS

Proposition 2.1. *If T is consistent, then T is a stable expansion of A iff T is a K45-extension of A.*

Proof. Let T be a consistent stable expansion of A. Moore [1, p.89] proved that stable expansions contain all instances of all K45-axioms. Since T is closed under the necessitation rule, we have for each ψ,

$$\psi \in T \text{ iff } A \cup \{L\phi : \phi \in T\} \cup \{\neg L\phi : \phi \notin T\} \vdash_{K45} \psi.$$

If $\phi \in T$, then $L\phi \in T$, hence $\neg L\phi \notin T$. But $\neg L \neg L\phi \supset L\phi$ is a theorem of K45, hence for each $\phi \in T$, $\{\neg L\phi : \phi \notin T\} \vdash_{K45} L\phi$. Hence T is a K45-extension.

Let T be a K45-extension. Let $\mathcal{P}$ be propositional calculus augmented by the necessitation rule. Let Ax be the set of all instances of the modal axiom schemes of K45. Then for each ψ,

$$\psi \in T \text{ iff } A \cup \{\neg L\phi : \phi \notin T\} \cup Ax \vdash_{\mathcal{P}} \psi.$$

Hence

$$\psi \in T \text{ iff } A \cup \{\neg L\phi : \phi \notin T\} \cup \{L\phi : \phi \in T\} \cup Ax \vdash \psi$$

Moore [1, p. 89, the last paragraph of Section 4] proved, that this is equivalent to

$$\psi \in T \text{ iff } A \cup \{\neg L\phi : \phi \notin T\} \cup \{L\phi : \phi \in T\} \vdash \psi,$$

which proves the proposition.□

Lemma 3.1. *If Φ is admissible for A and ϕ is an A-formula, then for each $V \in N_{A,\Phi}$, $\langle \mathcal{N}_{A,\Phi}, V\rangle \models \phi$ iff $V_\Phi(\phi) = 1$.*

Proof. By induction on the complexity of ϕ. The only nontrivial case is when ϕ has the form $L\eta$ in the induction step. Since $L\eta \in A^L$, we have $L\eta \in \Phi$ or $L\eta \in \Psi$.

Let $V_\Phi(L\eta) = 0$. Then $L\eta \in \Phi$ and

$$A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\} \not\vdash \eta.$$

Since $\Phi \cup \Psi = A^L$, there exists a valuation W such that

$$W_\Phi(A) = W_\Phi(\{\psi : L\psi \in \Psi\}) = 1 \text{ and } W_\Phi(\eta) = 0.$$

This W is an element of $N_{A,\Phi}$; hence, by the induction hypothesis, $\langle \mathcal{N}, W\rangle \not\models \eta$, so that $\langle \mathcal{N}, V\rangle \models \neg L\eta$.

Let $V_\Phi(L\eta) = 1$. Then $L\eta \in \Psi$, hence for each $W \in N_{A,\Phi}$, $W_\Phi(\eta) = 1$. Hence, by the induction hypothesis, $W \models \eta$ for each $W \in N$, and $V \models L\eta$.□

Corollary 3.1. *$W_{A,\Phi}$ is a stable set containing $A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$.*

Proof. Immediately follows from Lemma 3.1. (Recall that $W_{A,\Phi}$ is $\{\phi : \langle \mathcal{N}_{A,\Phi}, V\rangle \models \phi.\}$□

Corollary 3.2. *$A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$ is consistent with S5.*

Proof. All formulas in this set are true in the S5-model $\mathcal{N}_{A,\Phi}$. □

Untill the end of Appendix, we shall write $\mathcal{N}$ for $\mathcal{N}_{A,\Phi}$ and N for $N_{A,\Phi}$.

Theorem 3.1. *Let S be any modal logic contained in S5, and let Φ be S-admissible for A. Then $W_{A,\Phi}$ is an S-extension of A.*

Proof. It is sufficient to prove that for each ϕ and for $B = \{\neg L\eta : \text{for some } \alpha, \langle \mathcal{N}, \alpha\rangle \not\models \eta\}$,

$$\mathcal{N} \models \phi \text{ iff } A \cup B \vdash_S \phi \tag{A1}$$

The "if" part follows from Corollary 3.1, because S5 contains S.

Let us prove the "only if" part by induction on the maximal nesting depth of L in ϕ, $m(\phi)$. Assume that the "only if" part of (A1) is valid for each ϕ with $m(\phi) < n$, and assume $m(\phi) = n$ and $\mathcal{N} \models \phi$. We may assume that ϕ is in the conjunctive normal form, i.e., ϕ is $\phi_1 \wedge \ldots \wedge \phi_k$, and each ϕ_i has the form

$$\psi \vee \neg L\eta_1 \vee \ldots \vee \neg L\eta_l \wedge L\zeta_1 \vee \ldots \vee L\zeta_m \tag{A2}$$

with $m(\psi) = 0$, $l, m \geq 0$ and $\mathcal{N} \models \phi_i$.

If for some j and for some α, $\langle \mathcal{N}, \alpha\rangle \models \neg L\eta_j$, then for some β, $\langle \mathcal{N}, \beta\rangle \not\models \eta_j$. Hence $\neg L\eta_j \in B$ and $A \cup B \vdash_S \phi_i$.

If for some α and for some j, $\langle \mathcal{N}, \alpha\rangle \models L\zeta_j$, then $\mathcal{N} \models \zeta_j$. But $m(\zeta_j) < n$, hence, by the induction hypothesis, $A \cup B \vdash_S \zeta_j$. Applying the necessitation rule and the propositional logic, we obtain the derivability of ϕ_i.

The only remaining case is that of

$$\mathcal{N} \models L\eta_1 \wedge \ldots \wedge L\eta_l \wedge \neg L\zeta_1 \wedge \ldots \wedge \neg L\zeta_m.$$

Since $\mathcal{N} \models \phi_i$, we conclude $\mathcal{N} \models \psi$ from (A2). By Lemma 3.1, this means that ψ is a tautological consequence of $A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$. Since Φ is S-admissible for A, and S contains the necessitation rule, we get $A \cup \neg\Phi \vdash_S \psi$. Hence $A \cup B \vdash_S \phi_i$. Thus, $A \cup B \vdash_S \phi_i$ in each case for each i, so that $A \cup B \vdash \phi$.□

A *frame* is a pair $\langle M, R\rangle$, where R is a binary relation on M.

Definition A1. Let S be a modal logic, and let $\mathcal{K}$ be a class of frames. We say that S *is characterized* by $\mathcal{K}$ if, for each set of formulas Γ and each formula ϕ, $\Gamma \vdash \phi$ if and only if the following condition is satisfied: For each Kripke model $\mathcal{M} = \langle M, R, V\rangle$ with $\langle M, R\rangle \in \mathcal{K}$, $\mathcal{M} \models \Gamma$ implies $\mathcal{M} \models \phi$.

We can say, for instance, that K is characterized by the class of all frames, T is characterized by the class of all reflexive frames, S4 is characterized by the class of all reflexive transitive frames, and S5 is characterized by the class of all universal frames (i.e., frames of the form $\langle M, M \times M\rangle$). K45 is characterized by the class of all transitive Euclidean frames.

Definition A2. Let $\mathcal{K}$ be a class of frames. We say that $\mathcal{K}$ *admits a quasi-amalgamation* if, for each $\langle M, R\rangle \in \mathcal{K}$ and for $\alpha_0 \notin M$, there exists a frame $\langle M \cup \{\alpha_0\}, R^*\rangle \in \mathcal{K}$ such that

$$(\{\alpha_0\} \times Rg(R)) \cup R \subseteq R^* \subseteq [\{\alpha_0\} \times (\{\alpha_0\} \cup M)] \cup R.$$

Such a frame is called *a quasi-amalgamation of $\langle M, R\rangle$ in $\mathcal{K}$.*

The notion of quasi-amalgamation is a generalisation of the well-known notion of amalgamation (see, for instance, Hughes and Cresswell [12]).

The classes of all frames, of all reflexive frames and of all reflexive and transitive frames admit a quasi-amalgamation. In each of these cases, a quasi-amalgamation of a frame $\langle M, R\rangle$ is given by the expression $\langle M \cup \{\alpha_0\}, R \cup \{\alpha_0\} \times (\{\alpha_0\} \cup M)\rangle$.

Theorem A1. *Let $\mathcal{K}$ be a class of frames admitting a quasi-amalgamation. Let S be characterized by $\mathcal{K}$, and let Φ be S-admissible for A. Then $W_{A,\Phi}$ is the only consistent S-extension of A containing $\neg\Phi$.*

Proof. $W_{A,\Phi}$ is an S-extension of A by Theorem 3.1, and contains $\neg\Phi$ by Corollary 3.1.

Let T be any consistent S-extension of A, and let $\neg\Phi \subseteq T$. T is stable; on the other hand, Moore [3] proved that stable sets coincide iff they contain the same objective formulas. Thus it is sufficient to prove that, for each ϕ with $m(\phi) = 0$,

$$\phi \in S \text{ if and only if } \mathcal{N} \models \phi.$$

Let us prove the "if" part first. If for each $\alpha \in N$, $\langle \mathcal{N}, \alpha\rangle \models \phi$, then, by Lemma 3.1, ϕ is a tautological consequence of $A \cup \neg\Phi \cup \Psi \cup \{\psi : L\psi \in \Psi\}$. Since Φ is S-admissible and S contains the necessitation rule, we have $A \cup \neg\Phi \vdash_S \phi$. But $A \cup \neg\Phi \subseteq T$ and T is deductively closed under S; hence $\phi \in S$.

"Only if": Assume that $\phi \in T$ but, for some $V \in N_{A,\Phi}$, $\langle \mathcal{N}, V\rangle \models \neg\phi$. Fix these V and ϕ. Since T is consistent with S, there is a Kripke model $\mathcal{M} = \langle M, R, W\rangle$ with $\langle M, R\rangle \in \mathcal{K}$ and $\mathcal{M} \models T$.

Consider the Kpipke model $\mathcal{M}^* = \langle M \cup \{\alpha_0\}, R^*, W^*\rangle$, where $\langle M \cup \{\alpha_0\}, R^*\rangle$ is a quasi-amalgamation of $\langle M, R\rangle$ in $\mathcal{K}$, $W^*(\alpha) = W(\alpha)$ for $\alpha \in M$, and $W^*(\alpha_0) = U_{A,\Phi}(V)$. Let us prove by induction on the complexity of ψ that, for each A-formula ψ,

$$\langle \mathcal{M}^*, \alpha_0\rangle \models \psi \text{ if and only if} \langle \mathcal{N}, V\rangle \models \psi. \tag{A3}$$

The only nontrivial step in the proof is the case when ψ has the form $L\eta$ in the induction step. Since $L\eta \in A^L$, we have $L\eta \in \Psi$ or $L\eta \in \Phi$.

If $\langle \mathcal{N}, V\rangle \not\models L\eta$, then $L\eta \in \Phi$ by Corollary 3.1. Hence $\neg L\eta \in T$, and, for some $\beta \in Rg(R), \langle \mathcal{M}, \beta\rangle \not\models \eta$. Hence $\langle \mathcal{M}^*, \beta\rangle \not\models \eta$. By Definition 2.3, $\alpha_0 R^* \beta$. Thus $\langle \mathcal{M}^*, \alpha_0\rangle \models \neg L\eta$.

If $\langle \mathcal{N}, V\rangle \models L\eta$, then $L\eta \in \Psi$. Hence, by Corollary 3.1, $\langle \mathcal{N}, V\rangle \models \eta$, and, by the induction hypothesis,

$$\langle \mathcal{M}^*, \alpha_0\rangle \models \eta. \tag{A4}$$

Since $A \cup \neg\Phi \subseteq T$, and Φ is admissible, we have $\eta \in T$. Hence, for all $\beta \in M$, $\langle \mathcal{M}^*, \beta\rangle \models \eta$. Then, from (A4), $\langle \mathcal{M}^*, \alpha_0\rangle \models L\eta$. (A3) is proven.

We have

$$T = Th_S(A \cup \{\neg L\psi : \psi \notin T\}). \tag{A5}$$

From (A3) and Corollary 3.1 we get:

$$\langle \mathcal{M}^*, \alpha_0\rangle \models A. \tag{A6}$$

For each $\psi \notin T$, for some $\beta \in Rg(R)$ we have $\langle \mathcal{M}^*, \beta\rangle \not\models \psi$; hence $\langle \mathcal{M}^*, \alpha_0\rangle \models \neg L\psi$. Since $\mathcal{M} \models T$, using (A5) and (A6) we get $\mathcal{M}^* \models T$. Hence $\langle \mathcal{M}^*, \alpha_0\rangle \models \phi$. But from (A3) and our assumption we get $\langle \mathcal{M}^*, \alpha_0\rangle \models \neg\phi$, contradiction.□

Theorem 3.2. *Let S be any of K, T, S4, K45, and let Φ be S-admissible for A. Then $W_{A,\Phi}$ is the unique S-extension of A containing $A \cup \neg\Phi$.*

Proof. The classes of all frames, of all reflexive frames, of all reflexive transitive frames characterize, respectively, the logics K, T and S4, so that we can apply Theorem A1. The class of all transitive Euclidean frames, which is known to characterize K45, does not admit a quasi-amalgamation. Let us call a frame $\langle M, R\rangle$ *strongly Euclidean* if it is transitive and, for each α and β from $Rg(R)$, $\alpha R\beta$. $\langle M \cup \{\alpha\}, \{\alpha\} \times Rg(R)\rangle$ is a quasi-amalgamation of this frame, and it is strongly Euclidean if the initial frame is strongly Euclidean. If $\langle M, R\rangle$ is transitive and Euclidean, then for $\alpha \in M$ the frame $\langle\{\beta \in M : \alpha = \beta \text{ or } \alpha R\beta\}, R|\{\beta \in M : \alpha = \beta \text{ or } \alpha R\beta\}\rangle$ is strongly Euclidean. From this fact we can easy deduce that K45 is characterized by class of all strongly Euclidean frames, and apply Theorem A1.□

Definition A3.. A class $\mathcal{K}$ of frames is said to be *closed* iff, for all $\langle M_1, R_1\rangle \in \mathcal{K}$ and $\langle M_2, R_2\rangle \in \mathcal{K}$ with $M_1 \cap M_2 = \emptyset$, the frame $\langle M_1 \cup M_2, R_1 \cup (M_1 \times M_2) \cup R_2\rangle$ belongs to $\mathcal{K}$, and $\mathcal{K}$ contains, along with each frame, all frames isomorphic to it.

Lemma A1. *Let $\Phi \subseteq A^L$, $\Psi = A^L \setminus \Phi$. Let $\mathcal{M}_1 = \langle M_1, R_1, V_1\rangle$ and $\mathcal{M}_2 = \langle M_2, R_2, V_2\rangle$ be Kripke models such that $M_1 \cap M_2 = \emptyset$, $\mathcal{M}_1 \models \neg\Phi$ and $\mathcal{M}_2 \models \psi$ for all $L\psi \in \Psi$. Let $\mathcal{M}$ be $\langle M_1 \cup M_2, R_1 \cup (M_1 \times M_2) \cup R_2\rangle$. Then, for each A-formula ϕ and for each $\alpha \in M_1$,*

$$\langle \mathcal{M}_1, \alpha\rangle \models \phi \text{ if and only if} \langle \mathcal{M}, \alpha\rangle \models \phi.$$

Proof. By induction on the complexity of ϕ. The only nontrivial step in the proof is the case when ϕ has the form $L\psi$.

Let $\langle \mathcal{M}, \alpha\rangle \models L\psi$. Then, for each $\beta \in M_1$ with $\alpha R_1 \beta$, $\langle \mathcal{M}, \beta\rangle \models \psi$. Hence, by the induction hypothesis, for all such β, $\langle \mathcal{M}_1, \beta\rangle \models \psi$, which means that $\langle \mathcal{M}_1, \alpha\rangle \models \phi$.

Let $\langle \mathcal{M}, \alpha\rangle \not\models L\psi$. Then, for some β with $\alpha R\beta$, $\langle \mathcal{M}, \beta\rangle \not\models \psi$. If $\beta \in M_1$, then, using the induction hypothesis, we get $\langle \mathcal{M}_1, \alpha\rangle \not\models L\psi$. Let $\beta \in M_2$. If $L\psi \in \Psi$, then $\mathcal{M}_2 \not\models \psi$, which contradicts the assumptions of the lemma. Thus, $L\psi \in \Phi$. Hence $\mathcal{M}_1 \models \neg L\psi$, and $\langle \mathcal{M}_1, \alpha\rangle \not\models L\psi$.
□

Theorem A2. *Let $\mathcal{K}$ be a closed class of frames, and let S be characterized by $\mathcal{K}$. Let T be any consistent S-extension of A. Let $\Psi = T \cap A^L$, and let $\Phi = A^L \setminus \Psi$. Then Φ is S-admissible for A.*

Proof. Since T is consistent with S, there is a model $\mathcal{M}_2 = \langle M_2, R_2, V_2\rangle$ with $\langle M_2, R_2\rangle \in \mathcal{K}$, $\mathcal{M}_2 \models T$. Since $A \cup \neg\Phi \subseteq T$, $A \cup \neg\Phi$ is consistent with S.

Let $L\psi \in \Psi$; since T is stable, $\psi \in T$.

Assume, on the contrary, that

$$A \cup \neg\Phi \not\vdash_S \psi. \tag{A7}$$

Then there is a model $\mathcal{M}_1 = \langle M_1, R_1, V_1\rangle$ with $\langle M_1, R_1\rangle \in \mathcal{K}$, $\mathcal{M}_1 \models A \cup \neg\Phi$, such that, for some $\alpha \in M_1$,

$$\langle M_1, \alpha\rangle \not\models \psi. \tag{A9}$$

We may assume $M_1 \cap M_2 = \emptyset$. Consider the model

$$\mathcal{M} = \langle M_1 \cup M_2, R_1 \cup (M_1 \times M_2) \cup R_2, V_1 \cup V_2\rangle.$$

By Lemma A1, $\mathcal{M} \models A$; clearly, for each $\neg L\psi \in T$, $\mathcal{M} \models \neg L\psi$. Since T is an S-extension of A, and $\mathcal{M}$ is a model of S, we have $\mathcal{M} \models T$. Hence $\mathcal{M} \models \psi$. But by (A8) and Lemma A1, $\langle \mathcal{M}, \alpha \rangle \not\models \psi$, contradiction. Thus (A7) is false, and Φ is S-admissible for A. □

Theorem 3.3. *Let S be K,T or S4. Then each S-extension of A equals $W_{A,\Phi}$ for some S-admissible Φ.*

Proof. The classes of all frames, of all reflexive frames and of all reflexive transitive frames all are closed and admit a quasi-amalgamation. The theorem follows from Theorems A2 and A1.□

Theorem 3.4. *T is a stable expansion (or K45-extension) of A iff T is $W_{A,\Phi}$ for some Φ which is propositionally admissible for A.*

Proof. Let $\Phi \subseteq A^L$, let Φ be propositionally admissible for A, and let $\Psi = A^L \setminus \Phi$. Let $B = A \cup \{\psi : L\psi \in \Psi\}$. We have $B \cup \neg\Phi \vdash \psi$ for each $\psi \in \Psi$. Hence Φ is S-admissible for B for each S contained in S5, and, in particular for K45. By Theorem 3.1 and Proposition 2.1, B has a stable expansion $T = W_{B,\Phi}$. Using the definition of $W_{B,\Phi}$, we conclude that $W_{B,\Phi} = W_{A,\Phi}$. Since, for each $L\psi \in \Psi$,

$$A \cup \{L\psi : \psi \in T\} \cup \{\neg L\psi : \psi \notin T\} \vdash \psi,$$

T is a stable expansion of A too.

Conversely, let T be a stable expansion of A. Let $\Psi = A^L \cup T$, $\Phi = A^L \setminus \Psi$. Let $L\psi \in \Psi$. Then ψ is a tautological consequence of $A \cup \{L\psi : \psi \in T\} \cup \{\neg L\psi : \psi \notin T\}$. Since $\psi \in A^L$, ψ is a tautological consequence of $A \cup \Psi \cup \neg\Phi$ also. Thus Φ is propositionally admissible for A.

Consider $B = A \cup \{\psi : L\psi \in \Psi\}$. $B \subseteq T$, so that T is a $K45$-extension of B, too. We have $B^L = A^L$ and, trivially, for $L\psi \in \Psi$, $B \cup \neg\Phi \vdash \psi$. Hence Φ is $K45$-admissible for B. Hence, by Theorem 3.2, T equals to $W_{B,\Phi}$. Let us recall the construction of the model $\mathcal{N}_{A,\Phi}$: We see that $W_{A,\Phi}$ coincides with $W_{B,\Phi}$.□

Proposition 4.1. *Let S be any modal logic for which Theorem 3.2 is true, and let A be any set of objective sentences consistent with S. Then A has $W_{A,\emptyset}$ as its unique S-extension.*

Proof. $A^L = \emptyset$, so that $\emptyset$ is trivially admissible for A. Apply Theorem 3.2.□

Acknowledgments

I am greatly indebted to Vladimir Lifschitz for the help in different ways. Discussions with Vladimir were very usefull and stimulating; he sent me newest papers on nonmonotonic logic, which are hardly available in the USSR, especially in such a little provincial town as Pereslavl-Zalessky. Vladimir polished my terrible English and helped to prepare camera-ready empy of the paper by using TEX.

The final version was prepared when I was in Stanford University as a visiting scholar. I am thankful to Professors Patrick Suppes and John McCarthy for the invitation to visit Stanford and for the help in arranging my visit.

References

[1] Moore, R.C., Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, **25** (1985), 75–94.

[2] Stalnaker, R.A., *A Note on Non-Monotonic Modal Logic.* Department of Philosophy, Cornell University (1980).

[3] Lifschitz, V., The mathematics of nonmonotonic reasoning. *Proc. LICS-89.*

[4] McDermott, D. and Doyle, J., Nonmonotonic logic I. *Artificial Intelligence*, **13** (1980), 41–72.

[5] McDermott, D., Nonmonotonic logic II: Nonmonotonic modal theories. *Journal of the ACM*, **29**, 33–57.

[6] Konolige, K., On the relation between default and autoepistemic logic. *Artificial Intelligence*, **35** (1988), 342–382.

[7] Morris, P., Autoepistemic stable closures and contradiction resolution. In: Reinfrank e.a. (Eds.), *Non-Monotonic Reasoning (Lecture Notes in Artificial Intelligence, v.346)*, Springer, 1988, 60–73.

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

[9] Moore, R.C., Possible world semantics for autoepistemic logic. In: M.Ginsberg (Ed.), *Readings on Non-Monotonic Reasoning.* Morgan Kaufmann, 1987, 137–142.

[10] Shvarts, G.F., Stable extensions of autoepistemic theories. In: *All-Union Conference on Artificial Intelligence (Proceedings)*, v. 1, 1988, 483–487 (in Russian).

[11] Shvarts, G.F., Nonmonotonic logics based on different modal logics. In:*COLOG-88. Papers Presented at the International Conference on Computer Logic. Part 2.* Tallinn, 1988.

[12] Hughes, G.E., Cresswell, M.J., *A Companion to Modal Logic.* Methuen, London, 1984.