REPRESENTING AUTOEPISTEMIC LOGIC IN MODAL LOGIC

Frank M. Brown

***Abstract**: The nonmonotonic logic called Autoepistemic Logic is shown to be representable in a monotonic Modal Quantificational Logic whose modal laws are stronger than S5. Specifically, it is proven that a set of sentences of First Order Logic is a fixed-point of the fixed-point equation of Autoepistemic Logic with an initial set of axioms if and only if the meaning or rather disquotation of that set of sentences is logically equivalent to a particular modal functor of the meaning of that initial set of sentences. This result is important because the modal representation allows the use of powerful automatic deduction systems for Modal Logic and unlike the original Autoepistemic Logic, it is easily generalized to the case where quantified variables may be shared across the scope of modal expressions thus allowing the derivation of quantified consequences. Furthermore, this generalization properly treats such quantifiers since both the Barcan formula and its converse hold*

***Keywords**: Autoepistemic Logic, Modal Logic, Nonmonotonic Logic.*

1. Introduction

One of the most well known nonmonotonic logics [Antoniou 1997] which inherently deals with entailment conditions in addition to possibility conditions in its sentences is the so-called Autoepistemic Logic [Moore 1985][10]. The basic idea of Autoepistemic Logic is that there is a set of axioms {'Γi} and for every closed sentence χ there are two non-logical "inference rules" of the forms:

$$\frac{\chi:}{L'\chi} \qquad\qquad \frac{:\neg\chi}{\neg L'\chi}$$

where the predicate symbol L intuitively means that its argument names a sentence which is inferable. The first rule suggests that L 'χ may be inferred from χ and the second rule suggests that ¬L'χ may be inferred if χ is not inferable. When L is in Γ such "inference rules" maybe circular in that determining if they are applicable depends on the inferability or noninferability of χ which in turn depends on what else was derivable. Thus, tentatively applying such inference rules by checking whether χ has been or has not yet been inferred produces consequences which may later have to be retracted. For this reason valid inferences in a nonmonotonic logic such as Autoepistemic Logic are essentially carried out not in the original nonmonotonic language, but rather in some (monotonic) metatheory in which that nonmonotonic logic is defined. [Moore 1985; Konolige 1987; Konolige 1987b] explicated the above intuition by defining Autoepistemic Logic in terms of the set theoretic proof theory metalanguage of a First Order Logic (i.e. FOL) object language with the fixed-point equation:

$$'\kappa=(\text{ael } '\kappa\ \{'\Gamma_i\})$$

where ael is defined as:

$$(\text{ael } '\kappa\ \{'\Gamma_i\}) =_{df} (\text{fol}(\{'\Gamma_i\} \cup \{'(L\ '\chi_i) : '\chi_i \varepsilon '\kappa\} \cup \{'(\neg(L\ '\chi_i)) : '\chi_i \notin '\kappa\}))$$

where $'\chi_i$ is the ith sentence of the FOL object language and where 'κ and {'Γi} are sets of closed sentences of the FOL object language. A closed sentence is a sentence without any free variables. fol is a function which produces the set of theorems derivable in FOL from the set of sentences to which it is applied. The quotations appended to the front of these Greek letters indicate references in the metalanguage to the sentences of the FOL object language. Interpreted doxastically this fixed-point equation states:

the set of closed sentences which are believed is equal to
  the set of theorems derivable by the laws of FOL from the union of
    the set of closed sentences {'Γi},
    the set of all closed sentences of the form: '(L 'χi) for each i such that 'χi is believed,
    and the set of all closed sentences of the form: '(¬(L 'χi)) for each i such that 'χi is not believed.

The purpose of this paper is to show that all this metatheoretic machinery including the formalized syntax of FOL, the proof theory of FOL, the axioms of a strong set theory, and the set theoretic fixed-point equation is

[10]Autoepistemic Logic may be viewed as an improved version of the systems described in [McDermott 1980; McDermott 1982].

not needed and that the essence of Autoepistemic Logic is representable as a necessary equivalence in a (monotonic) Modal Quantificational Logic. Interpreted as a doxastic logic this equivalence states:

that which is believed is equivalent to: for all i Γ_i and for all i (L 'χ_i) if and only if χ_i is believed.

thereby eliminating the metatheoretic machinery.[11]

The remainder of this paper proves that this modal representation is equivalent to Autoepistemic Logic. Section 2 describes a formalized syntax for a FOL object language. Section 3 describes the part of the proof theory of FOL needed herein (i.e. theorems FOL1-FOL4). Section 4 describes the Intensional Semantics of FOL which includes laws giving the meaning of FOL sentences: M0-M8, theorems giving the meaning of sets of sentences: MS1, MS2, MS3, and laws specifying the relationship of meaning and modality to the proof theory of FOL (i.e. the laws R0, A1, A2, and A3 and the theorems: C1, C2, C3, and C4). The modal version of Autoepistemic Logic is defined in section 5 and explicated with theorems MA1-MA6 and SS1-SS2. In section 6, this modal version is shown by theorems AEL1 and AEL2 to be equivalent to the set theoretic fixed-point equation for Autoepistemic Logic. Figure 1 outlines the relationship of all these theorems in producing the final theorems AEL2, FOL4, and MA6.

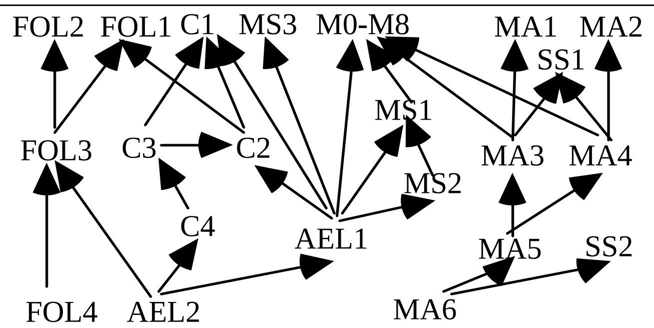

Figure 1: Dependencies among the Theorems

2. Formal Syntax of First Order Logic

We use a First Order Logic (i.e. FOL) defined as the six tuple: ($\rightarrow$, #f, $\forall$, *vars*, *predicates*, *functions*) where $\rightarrow$, #f, and $\forall$ are logical symbols, *vars* is a set of variable symbols, *predicates* is a set of predicate symbols each of which has an implicit arity specifying the number of associated terms, and *functions* is a set of function symbols each of which has an implicit arity specifying the number of associated terms. The sets of logical symbols, variables, predicate symbols, and function symbols are pairwise disjoint. Lower case Roman letters possibly indexed with digits are used as variables. Greek letters possibly indexed with digits are used as syntactic metavariables. γ, $\gamma_1,...,\gamma_n$, range over the variables, ξ, $\xi_1,...,\xi_n$ range over sequences of variables of an appropriate arity, $\pi,\pi_1...\pi_n$ range over the predicate symbols, ϕ, $\phi_1...\phi_n$ range over function symbols, δ, $\delta_1...\delta_n$, σ range over terms, and α, $\alpha_1,...,\alpha_n$, β, $\beta_1,...,\beta_n$, χ, $\chi_1,...,\chi_n$, $\Gamma_1,...,\Gamma_n$, φ range over sentences. The terms are of the forms γ and (ϕ δ1...δn), and the sentences are of the forms ($\alpha\rightarrow\beta$), #f, ($\forall\gamma$ α), and (π $\delta_1...\delta_n$). A nullary predicate π or function ϕ is written as a sentence or a term without parentheses. $\varphi\{\pi/\lambda\xi\alpha\}$ represents the replacement of all occurrences of π in φ by $\lambda\xi\alpha$ followed by lambda conversion. The primitive symbols are shown in Figure 2 with their intuitive interpretations.

Symbol	Meaning
$\alpha\rightarrow\beta$	if α then β.
#f	falsity
$\forall\gamma\ \alpha$	for all γ, α.

Figure 2: Primitive Symbols of First Order Logic

The defined symbols are listed in Figure 3 with their definitions and intuitive interpretations.

[11]The occurrence of quotation in the argument to L may be replaced by using a new symbol *L* such that (*L*) replaces (L ').

Symbol	Definition	Meaning		Symbol	Definition	Meaning
$\neg\alpha$	$\alpha \rightarrow \#f$	not α		$\alpha\wedge\beta$	$\neg(\alpha \rightarrow \neg\beta)$	α and β
#t	$\neg \#f$	truth		$\alpha\leftrightarrow \beta$	$(\alpha\rightarrow \beta) \wedge (\beta\rightarrow \alpha)$	α if and only if β
$\alpha\vee\beta$	$(\neg \alpha)\rightarrow \beta$	α or β		$\exists\gamma\, \alpha$	$\neg\forall\gamma\, \neg\alpha$	for some γ , α

Figure 3: Defined Symbols of First Order Logic

The particular FOL used herein includes the predicate symbol L and a denumerably infinite number of 0-ary function symbols representing the names (i.e. ') of the sentences (i.e.) of this First Order Logic. The FOL object language expressions are referred in the metalanguage (which also includes a FOL syntax) by inserting a quote sign in front of the object language entity thereby making a structural descriptive name of that entity. In addition to referring to object language sentences, the formalized metalanguage also needs to refer to sets of sentences of FOL. Generally, a set of sentences is represented as: {'Γ_i} which is defined as: {'Γ_i: #t} which in turn is defined as: {s: $\exists$i(s='Γ_i)} where i ranges over some range of numbers (which may be finite or non-infinite). With a slight abuse of notation we also write 'κ, 'Γ to refer to such sets.

3. Proof Theory of First Order Logic

First Order Logic (i.e. FOL) is axiomatized with a recursively enumerable set of theorems as the set of axioms is itself recursively enumerable and its inference rules are recursive. The axioms and inference rules of FOL [Mendelson 1964] are those given in Figure 4. They form a standard set of axioms and inference rules for FOL.

MA1: $\alpha \rightarrow (\beta\rightarrow \alpha)$

MA2: $(\alpha\rightarrow (\beta\rightarrow \rho)) \rightarrow ((\alpha\rightarrow \beta)\rightarrow (\alpha\rightarrow \rho))$

MA3: $((\neg \alpha)\rightarrow (\neg \beta))\rightarrow (((\neg \alpha)\rightarrow \beta)\rightarrow\alpha)$

MA4: $(\forall\gamma\, \alpha)\rightarrow \beta$ where β is the result of substituting an expression (which is free for the free positions of γ in α) for all the free occurrences of γ in α.

MA5: $((\forall\gamma(\alpha \rightarrow \beta)) \rightarrow (\alpha\rightarrow(\forall\gamma\, \beta)))$ where γ does not occur in α.

MR1: from α and $(\alpha\rightarrow \beta)$ infer β

MR2: from α infer $(\forall\gamma\, \alpha)$

Figure 4: Inferences Rules and Axioms of FOL

In order to talk about sets of sentences we include in the metatheory set theory symbolism as developed along the lines of [Quine 1969]. This set theory includes the symbols ε, $\not\in$, $\supseteq$, $=$, $\cup$ as is defined therein.
The derivation operation (i.e. fol) of any First Order Logic obeys the Inclusion (i.e. FOL1) and Idempotence (i.e. FOL2) properties:

FOL1: (fol 'κ)$\supseteq$'κ Inclusion

FOL2: (fol 'κ)$\supseteq$(fol(fol 'κ)) Idempotence

From these two properties we prove:

FOL3: (ael 'κ 'Γ 'α_i:'β_{ij}/'χ_i))=(fol(ael 'κ 'Γ 'α_i:'β_{ij}/'χ_i))

proof: FOL1 and FOL2 imply that (fol(fol 'κ))=(fol 'κ). Since ael begins with fol this implies: 'κ=(fol(ael 'κ)) QED.

FOL4: ('κ=(ael 'κ 'Γ 'α_i:'β_{ij}/'χ_i))$\rightarrow$('κ=(fol 'κ))

proof: From the hypothesis and FOL3: 'κ=(fol(ael 'κ 'Γ 'α_i:'β_{ij}/'χ_i)) is derived. Using the hypothesis to replace (ael 'κ 'Γ 'α_i:'β_{ij}/'χ_i) by 'κ in this result gives: 'κ=(fol 'κ). QED.

4. Intensional Semantics of FOL

The meaning (i.e. mg) [Brown 1978, Boyer&Moore 1981] or rather disquotation of a sentence of First Order Logic (i.e. FOL) is defined to satisfy the laws given in Figure 5 below mg is defined in terms of mgs which maps each FOL object language sentence and an association list into a meaning. Likewise, mgn maps a FOL object language term and an association list into a meaning. An association list is simply a list of pairs consisting of an object language variable and the meaning to which it is bound.

M0: (mg 'α) =df (mgs '($\forall\gamma 1 \ldots \gamma n\ \alpha$)'()) where '$\gamma 1 \ldots$'$\gamma n$ are all the free variables in 'α

M1: (mgs '($\alpha \rightarrow \beta$)a) $\leftrightarrow$ ((mgs 'α a)$\rightarrow$(mgs 'β a))

M2: (mgs '#f a) $\leftrightarrow$ #f

M3: (mgs '($\forall\ \gamma\ \alpha$)a) $\leftrightarrow$ $\forall$x(mgs 'α(cons(cons 'γ x)a))

M4: (mgs '($\pi\ \delta_1 \ldots \delta_n$)a) $\leftrightarrow$ (π(mgn 'δ_1 a)...(mgn 'δ_n a)) for each predicate symbol 'π.

M5: (mgn '($\phi\ \ \delta_1 \ldots \delta_n$)a) = ($\phi$(mgn '$\delta_1$ a)...(mgn 'δ_n a)) for each function symbol 'ϕ.

M6: (mgn 'γ a) = (cdr(assoc 'γ a))

M7: (assoc v L) = (if(eq? v(car(car L))) (car L) (assoc v(cdr L)))
where: cons, car, cdr, eq?, if are axiomatized as they are axiomatized in Scheme.

M8: (mgn ''α a) = 'α

Figure 5: The Meaning of FOL Sentences

The meaning of a set of sentences is defined in terms of the meanings of the sentences in the set as:

(ms 'κ) =df $\forall$s((sε'κ)$\rightarrow$(mg s))

MS1: (ms{'α: Γ}) $\leftrightarrow$ $\forall\xi(\Gamma\rightarrow\alpha)$ where ξ is the sequence of all the free variables in 'α and where Γ is any sentence of the intensional semantics.

proof: (ms{'α:Γ}) Unfolding ms and the set pattern abstraction symbol gives: $\forall$s((sε{s: $\exists\xi$((s='α)$\wedge\Gamma$)})$\rightarrow$(mg s))

where ξ is a sequence of the free variables in 'a. This is equivalent to: $\forall$s(($\exists\xi$((s='α)$\wedge\Gamma$)))$\rightarrow$(mg s))

which is logically equivalent to: $\forall$s$\forall\xi$ (((s='κ)$\wedge\Gamma$)$\rightarrow$(mg s)) which is equivalent to: $\forall\xi$($\Gamma\rightarrow$(mg 'α))

Unfolding mg using M0-M8 then gives: $\forall\xi(\Gamma\rightarrow\alpha)$ QED

The meaning of the union of two sets of FOL sentences is the conjunction of their meanings (i.e. MS3) and the meaning of a set is the meaning of all the sentences in the set (i.e. MS2):

MS2: (ms{'Γ_i}) $\leftrightarrow$ $\forall$i$\forall\xi_i\Gamma_i$

proof: (ms{'Γ_i}) Unfolding the set notation gives: (ms{'Γ_i: #t})

By MS1 this is equivalent to: $\forall_i\forall\xi_i$(#t$\rightarrow\Gamma_i$) which is equivalent to: $\forall$i$\forall\xi_i$(Γ_i QED.

MS3: (ms('$\kappa\cup$'Γ)) $\leftrightarrow$ ((ms 'κ)$\wedge$(ms 'Γ))

proof: Unfolding ms and union in: (ms('$\kappa\cup$'Γ)) gives: $\forall$s((sε{s: (sε'κ)$\vee$(sε'Γ)})$\rightarrow$(mg s)) or rather:

$\forall$s(((sε'κ)$\vee$(sε'Γ))$\rightarrow$(mg s)) which is logically equivalent to: ($\forall\alpha$((sε'κ)$\rightarrow$(mg s)))$\wedge$($\forall$s((sε'Γ)$\rightarrow$(mg s)))

Folding ms twice then gives:((ms 'κ)$\wedge$(ms 'Γ)) QED.

The meaning operation may be used to develop an Intensional Semantics for a FOL object language by axiomatizing the modal concept of necessity so that it satisfies the theorem:

C1: ('$\alpha\varepsilon$(fol 'κ)) $\leftrightarrow$ ([] ((ms 'κ)$\rightarrow$(mg 'α)))

for every sentence 'α and every set of sentences 'κ of that FOL object language. The necessity symbol is represented by a box: []. C1 states that a sentence of FOL is a FOL-theorem (i.e. fol) of a set of sentences of FOL if and only if the meaning of that set of sentences necessarily implies the meaning of that sentence. One modal logic which satisfies C1 for FOL is the Z Modal Quantificational Logic described in [Brown 1987; Brown 1989] whose theorems are recursively enumerable and which extends the weaker possibility axioms used in [Lewis 1936; Bressan 1972; Hendry & Pokriefka 1985]. We note that Z includes all the laws of S5 modal Logic [Hughes & Cresswell 1968] whose modal axioms and inference rules are given in Figure 6. Therein, κ and Γ represent arbitrary sentences of the intentional semantics.

R0: from α infer ([] κ)	A2: ([]($\kappa\rightarrow\Gamma$)) $\rightarrow$ (([]κ)$\rightarrow$ ([]Γ))
A1: ([]κ) $\rightarrow$ κ	A3: ([]κ) $\vee$ ([]$\neg$[]κ)

Figure 6: The Laws of S5 Modal Logic

These S5 modal laws and the laws of FOL given in Figure 4 constitute an S5 Modal Quantificational Logic similar to [Carnap 1946; Carnap 1956], and a FOL version [Parks 1976] of [Bressan 1972] in which the Barcan formula: ($\forall\gamma$([]κ))$\rightarrow$([]$\forall\gamma\kappa$) and its converse hold. The R0 inference rule implies that anything derivable in the metatheory is necessary. Thus, in any logic with R0, contingent facts would never be asserted as additional axioms of the metatheory. For example, we would not assert $([](\kappa\leftrightarrow\Gamma))$ as an axiom and then try to prove $([](\kappa\rightarrow\alpha))$. Instead we would try to prove that $([](\kappa\leftrightarrow\Gamma))\rightarrow([](\kappa\rightarrow\alpha))$.

The defined Modal symbols used herein are listed in Figure 7 with their definitions and interpretations.

Symbol	Definition	Meaning		Symbol	Definition	Meaning
$<>\kappa$	$\neg [] \neg\kappa$	α is logically possible		$[\kappa]\,\Gamma$	$[]\,(\kappa\rightarrow\Gamma)$	β entails α
$\kappa\equiv\Gamma$	$[]\,(\kappa\leftrightarrow\Gamma)$	α is logically equivalent to β		$<\kappa>\,\Gamma$	$<>\,(\kappa\wedge\Gamma)$	α and β is logically possible

Figure 7: Defined Symbols of Modal Logic

For example, folding the definition of entailment, C1 may be rewritten more compactly as:

C1': $('\alpha\varepsilon(\text{fol } '\kappa)) \leftrightarrow ([(\text{ms } '\kappa)](\text{mg } '\alpha))$

This compact notation for entailment is used hereafter.

From the laws of the Intensional Semantics we prove that the meaning of the set of FOL consequences of a set of sentences is the meaning of that set of sentences (C2), the FOL consequences of a set of sentences contain the FOL consequences of another set if and only if the meaning of the first set entails the meaning of the second set (C3), and the sets of FOL consequences of two sets of sentences are equal if and only if the meanings of the two sets are logically equivalent (C4):

C2: $(\text{ms}(\text{fol } '\kappa))\equiv(\text{ms } '\kappa)$

proof: The proof divides into two cases:

(1) $[(\text{ms } '\kappa)](\text{ms}(\text{fol } '\kappa)))$ Unfolding the second ms gives: $[(\text{ms } '\kappa)]\forall s((s\varepsilon(\text{fol } '\kappa))\rightarrow(\text{mg } s))$

By the soundness part of C1 this is equivalent to: $[(\text{ms } '\kappa)]\forall s(([(\text{ms } '\kappa)](\text{mg } s))\rightarrow(\text{mg } s))$

By the S5 laws this is equivalent to: $\forall s(([(\text{ms } '\kappa)](\text{mg } s))\rightarrow [(\text{ms } '\kappa)](\text{mg } s))$ which is a tautology.

(2) $[(\text{ms}(\text{fol } '\kappa))](\text{ms } '\kappa)$ Unfolding ms twice gives: $[\forall s((s\varepsilon(\text{fol } '\kappa))\rightarrow(\text{mg } s))]\forall s((s\varepsilon '\kappa)\rightarrow(\text{mg } s))$

which is: $[\forall s((s\varepsilon(\text{fol } '\kappa))\rightarrow(\text{mg } s))]((s\varepsilon '\kappa)\rightarrow(\text{mg } s))$ Backchaining on the hypothesis and then dropping it gives: $(s\varepsilon '\kappa)\rightarrow(s\varepsilon(\text{fol } '\kappa))$. Folding $\supseteq$ gives an instance of FOL1. QED.

C3: $(\text{fol } '\kappa)\supseteq(\text{fol } '\Gamma) \leftrightarrow ([(\text{ms } '\kappa)](\text{ms } '\Gamma))$

proof: Unfolding $\supseteq$ gives: $\forall s((s\varepsilon(\text{fol } '\Gamma))\rightarrow(s\varepsilon(\text{fol } '\kappa)))$

By C1 twice this is equivalent to: $\forall s(([(\text{ms } '\Gamma)](\text{mg } s))\rightarrow([(\text{ms } '\kappa)](\text{mg } s)))$

By the laws of S5 modal logic this is equivalent to: $([(\text{ms } '\kappa)]\forall s(([(\text{ms } '\Gamma)](\text{mg } s)))\rightarrow(\text{mg } s)))$

By C1 this is equivalent to: $[(\text{ms } '\kappa)]\forall s((s\varepsilon(\text{fol } '\Gamma))\rightarrow(\text{mg } s))$. Folding ms then gives: $[(\text{ms } '\kappa)](\text{ms}(\text{fol } '\Gamma))$

By C2 this is equivalent to: $[(\text{ms } '\kappa)](\text{ms } '\Gamma)$. QED.

C4: $((\text{fol } '\kappa)=(\text{fol } '\Gamma)) \leftrightarrow ((\text{ms } '\kappa)\equiv(\text{ms } '\Gamma))$

proof: This is equivalent to $(((\text{fol } '\kappa)\supseteq(\text{fol } '\Gamma))\wedge((\text{fol } '\Gamma)\supseteq(\text{fol } '\kappa))) \leftrightarrow ([(\text{ms } '\kappa)](\text{ms } '\Gamma))\wedge([(\text{ms } '\Gamma)](\text{ms } '\kappa))$ which follows by using C3 twice.

5. Autoepistemic Logic Represented in Modal Logic

The fixed-point equation for Autoepistemic Logic may be expressed in S5 Modal Quantificational Logic by the necessary equivalence:

$$\kappa\equiv(\text{AEL } \kappa\ \Gamma)$$

where AEL is defined as follows: $(\text{AEL } \kappa\ \Gamma) =_{df} \Gamma\wedge\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$

where χ_i is the ith sentence of the FOL object language.

Given below are some simple properties of AEL used to prove the equivalence of the proof theoretic and modal representations of Autoepistemic Logic. The first two theorems state that AEL entails Γ and that AEL entails for all i, $(L\ '\chi_i)$ if and only if χ_i holds in κ.

MA1: $[(AEL\ \kappa\ \Gamma)]\Gamma$

proof: By R0 it suffices to prove: $(AEL\ \kappa\ \Gamma)\rightarrow\Gamma$. Unfolding AEL gives: $(\Gamma\wedge\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))\rightarrow\Gamma$ which is a tautology. QED.

MA2: $[(AEL\ \kappa\ \Gamma)]\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$

proof: By R0 it suffices to prove: $(AEL\ \kappa\ \Gamma)\rightarrow\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$

Unfolding AEL gives: $[\Gamma\wedge\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))\rightarrow\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))$ which is a tautology. QED.

The concept (i.e. ss) of the combined meaning of all the sentences of the FOL object language whose meanings are entailed by a proposition is defined as follows: $(ss\ \kappa)$ =df $\forall s(([\kappa](mg\ s))\rightarrow(mg\ s))$. SS1 shows that a proposition entails the combined meaning of the FOL object language sentences that it entails. SS2 shows that if a proposition is necessarily equivalent to the combined meaning of all the FOL object language sentences that it entails, then there exists a set of FOL object language sentences whose meaning is necessarily equivalent to that proposition:

SS1: $[\kappa](ss\ \kappa)$

proof: By R0 it suffices to prove: $\kappa\rightarrow(ss\ \kappa)$. Unfolding ss gives: $\kappa\rightarrow\forall s(([\kappa](mg\ s))\rightarrow(mg\ s))$

which is equivalent to: $\forall s(([\kappa](mg\ s))\rightarrow(\kappa\rightarrow(mg\ s)))$ which is an instance of A1. QED.

SS2: $(\kappa\equiv(ss\ \kappa))\rightarrow\exists s(\kappa\equiv(ms\ s))$

proof: Letting s be {s: $([\kappa](mg\ s))$ gives: $(\kappa\equiv(ss\ \kappa))\rightarrow(\kappa\equiv(ms\{s: ([\kappa](mg\ s))))$

Unfolding ms and lambda conversion gives: $(\kappa\equiv(ss\ \kappa))\leftrightarrow(\kappa\equiv\forall s(([\kappa](mg\ s))\rightarrow(mg\ s)))$

Folding ss gives a tautology. QED.

Theorems MA3 and MA4 are analogous to MA1 and MA2 except that AEL is replaced by the combined meaning of all of the sentences entailed by AEL.

MA3: $[ss(AEL\ \kappa\ \forall i\Gamma_i)]\forall i\Gamma_i$

proof: By R0 it suffices to prove: $(ss(AEL\ \kappa\ \forall i\Gamma_i))\rightarrow\forall i\Gamma_i$

Unfolding ss gives: $(\forall s(([(AEL\ \kappa\ \forall i\Gamma_i)](mg\ s))\rightarrow(mg\ s)))\rightarrow\forall i\Gamma_i$

which is equivalent to: $(\forall s(([(AEL\ \kappa\ \forall i\Gamma_i)](mg\ s))\rightarrow(mg\ s)))\rightarrow\Gamma_i$

which by the meaning laws is equivalent to: $(\forall s(([(AEL\ \kappa\ \forall i\Gamma_i)](mg\ s))\rightarrow(mg\ s)))\rightarrow(mg\ '\Gamma_i)$

Backchaining on $(mg\ '\Gamma_i)$ with s in the hypothesis assigned to be $'\Gamma_i$ in the conclusion shows that it suffices to prove: $([(AEL\ \kappa\ \forall i\Gamma_i)](mg\ '\Gamma_i))$ which by the meaning laws is equivalent to: $([(AEL\ \kappa\ \forall i\Gamma_i)]\Gamma_i)$

which by the laws of S5 Modal Logic is equivalent to: $([(AEL\ \kappa\ \forall i\Gamma_i)]\forall i\Gamma_i)$ which is an instance of MA1. QED.

MA4: $[(ss(AEL\ \kappa\ \Gamma))]\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$

proof: By R0 it suffices to prove: $(ss(AEL\ \kappa\ \Gamma))\rightarrow\forall i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$

which is equivalent to: $(ss(AEL\ \kappa\ \Gamma))\rightarrow((([\kappa]\chi_i)\rightarrow(L\ '\chi_i))\wedge((\neg([\kappa]\chi_i))\rightarrow(\neg(L\ '\chi_i))))$

Unfolding ss gives: $(\forall s(([(AEL\ \kappa\ \Gamma)](mg\ s))\rightarrow(mg\ s)))\rightarrow((([\kappa]\chi_i)\rightarrow(L\ '\chi_i))\wedge((\neg([\kappa]\chi_i))\rightarrow(\neg(L\ '\chi_i))))$

Letting the quantified s in the hypothesis have the two instances: $'(L\ '\chi_i)$ and $'(\neg(L\ '\chi_i))$ and then dropping that hypothesis gives:

$((([(AEL\ \kappa\ \Gamma)](mg\ '(L\ '\chi_i)))\rightarrow(mg\ '(L\ '\chi_i)))\wedge(([(AEL\ \kappa\ \Gamma)](mg\ '(\neg(L\ '\chi_i))))\rightarrow(mg\ '(\neg(L\ '\chi_i)))))$

$\rightarrow((([\kappa]\chi_i)\rightarrow(L\ '\chi_i))\wedge((\neg([\kappa]\chi_i))\rightarrow(\neg(L\ '\chi_i))))$

By the meaning laws M0-M8 this is equivalent to:

$((([(AEL\ \kappa\ \Gamma)](L\ '\chi_i))\rightarrow(L\ '\chi_i))\wedge(([(AEL\ \kappa\ \Gamma)](\neg(L\ '\chi_i)))\rightarrow(\neg(L\ '\chi_i))))\rightarrow((([\kappa]\chi_i)\rightarrow(L\ '\chi_i))\wedge((\neg([\kappa]\chi_i))\rightarrow(\neg(L\ '\chi_i))))$

Using these instances of the hypothesis to backchain on $(L\ '\chi_i)$ and $(\neg(L\ '\chi_i))$ in the conclusion, and then dropping these instances gives:

$((([\kappa]\chi_i)\rightarrow([(\mathrm{AEL}\ \kappa\ \Gamma)](\mathrm{L}\ '\chi_i)))\wedge((\neg([\kappa]\chi_i)))\rightarrow([(\mathrm{AEL}\ \kappa\ \Gamma)](\neg(\mathrm{L}\ '\chi_i))))$

Using the laws of S5 Modal Logic then gives: $([(\mathrm{AEL}\ \kappa\ \Gamma)]((([\kappa]\chi_i)\rightarrow(\mathrm{L}\ '\chi_i))\wedge((\neg([\kappa]\chi_i))\rightarrow(\neg(\mathrm{L}\ '\chi_i))))$

which is equivalent to: $[(\mathrm{AEL}\ \kappa\ \Gamma)]((\mathrm{L}\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$ which holds by MA2. QED.

Finally MA5 and MA6 show that talking about the meanings of sets of FOL sentences in the modal representation of Autoepistemic Logic is equivalent to talking about propositions in general.

MA5: $(\mathrm{ss}(\mathrm{AEL}\ \kappa\ \forall i\Gamma_i)) \leftrightarrow (\mathrm{AEL}\ \kappa\ \forall i\Gamma_i)$

proof: In view of SS1, it suffices to prove: $(\mathrm{ss}(\mathrm{AEL}\ \kappa\ \forall i\Gamma_i))\rightarrow(\mathrm{AEL}\ \kappa\ \forall i\Gamma_i)$

Unfolding the second occurrence of AEL gives: $(\mathrm{ss}(\mathrm{AEL}\ \kappa\ \forall i\Gamma_i))\rightarrow(\forall i\Gamma_i\wedge\forall i((\mathrm{L}\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))$

which holds by theorems MA3 and MA4. QED.

MA6: $(\kappa\equiv(\mathrm{AEL}\ \kappa\ \forall i\Gamma_i)))\rightarrow\exists s(\kappa\equiv(\mathrm{ms}\ s))$

proof: $(\kappa\equiv(\mathrm{ss}(\mathrm{AEL}\ \kappa\ \forall i(\mathrm{mg}\ '\Gamma_i))))$ is derived from the hypothesis and MA5. Using the hypothesis to replace $(\mathrm{AEL}\ \kappa\ \forall i(\mathrm{mg}\ '\Gamma_i)))$ by κ in this result gives: $(\kappa\equiv(\mathrm{ss}\ \kappa))$. By SS2 this implies the conclusion. QED.

6. Conclusion: Autoepistemic Logic represented in Modal Logic

The relationship between the proof theoretic definition of Autoepistemic Logic [Moore 1985] and the modal representation is proven in two steps. First theorem AEL1 shows that the meaning of the set ael is the proposition AEL and then theorem AEL2 shows that a set of FOL sentences which contains its FOL theorems is a fixed-point of the fixed-point equation of Autoepistemic Logic with an initial set of axioms if and only if the meaning (or rather disquotation) of that set of sentences is logically equivalent to AEL of the meanings of that initial set of sentences.

AEL1: $(\mathrm{ms}(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\}))\equiv(\mathrm{AEL}(\mathrm{ms}\ '\kappa)(\forall i\Gamma_i))$

proof: By R0 it suffices to prove: $(\mathrm{ms}(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\}))\leftrightarrow(\mathrm{AEL}(\mathrm{ms}\ '\kappa)\Gamma)$. The left side is: $\mathrm{ms}(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\})$

Unfolding the definition of ael gives: $\mathrm{ms}(\mathrm{fol}(\{'\Gamma_i\}\cup\{'(\mathrm{L}\ '\chi_i)\text{: }'\chi_i\varepsilon(\mathrm{fol}\ '\kappa)\}\cup\{'(\neg(\mathrm{L}\ '\chi_i))\text{: }'\chi_i\not\in(\mathrm{fol}\ '\kappa)\}))$

By C2 this is equivalent to: $\mathrm{ms}(\{'\Gamma_i\}\cup\{'(\mathrm{L}\ '\chi_i)\text{: }'\chi_i\varepsilon(\mathrm{fol}\ '\kappa)\}\cup\{'(\neg(\mathrm{L}\ '\chi_i))\text{: }'\chi_i\not\in(\mathrm{fol}\ '\kappa)\})$

Using C1 twice gives: $\mathrm{ms}(\{'\Gamma_i\}\cup\{'(\mathrm{L}\ '\chi_i)\text{: }([\mathrm{ms}\ '\kappa]\chi_i)\}\cup\{'(\neg(\mathrm{L}\ '\chi_i))\text{: }\neg([\mathrm{ms}\ '\kappa]\chi_i)\})$

Using MS3 twice gives: $(\mathrm{ms}\{'\Gamma_i\})\wedge(\mathrm{ms}\{'(\mathrm{L}\ '\chi_i)\text{: }([\mathrm{ms}\ '\kappa]\chi_i)\})\wedge(\mathrm{ms}\{'(\neg(\mathrm{L}\ '\chi_i))\text{: }\neg([\mathrm{ms}\ '\kappa]\chi_i)\})$

Using MS2 gives: $(\forall i\Gamma_i)\wedge(\mathrm{ms}\{'(\mathrm{L}\ '\chi_i)\text{: }([\mathrm{ms}\ '\kappa]\chi_i)\})\wedge(\mathrm{ms}\{'(\neg(\mathrm{L}\ '\chi_i))\text{: }\neg([\mathrm{ms}\ '\kappa]\chi_i)\})$

Applying MS1 twice gives: $(\forall i\Gamma_i)\wedge\forall i(([\mathrm{ms}\ '\kappa]\chi_i)\rightarrow(\mathrm{L}\ '\chi_i))\wedge\forall i((\neg([\mathrm{ms}\ '\kappa]\chi_i))\rightarrow(\neg(\mathrm{L}\ '\chi_i)))$

which is logically equivalent to: $(\forall i\Gamma_i)\wedge\forall i((\mathrm{L}\ '\chi_i)\leftrightarrow([\mathrm{ms}\ '\kappa]\chi_i))$

Folding the definition of AEL gives: $(\mathrm{AEL}(\mathrm{ms}\ '\kappa)(\forall i\Gamma_i))$ QED.

AEL2: $((\mathrm{fol}\ '\kappa)=(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\}))\leftrightarrow((\mathrm{ms}\ '\kappa)\equiv(\mathrm{AEL}(\mathrm{ms}\ '\kappa)(\forall i\Gamma_i)))$

proof: $(\mathrm{fol}\ '\kappa)=(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\})$. By FOL3 this is equivalent to: $(\mathrm{fol}\ '\kappa)=(\mathrm{fol}(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\}))$

By C4 this is equivalent to: $(\mathrm{ms}\ '\kappa)\equiv(\mathrm{ms}(\mathrm{ael}(\mathrm{fol}\ '\kappa)\{'\Gamma_i\}))$.

By AEL1 this is equivalent to: $(\mathrm{ms}\ '\kappa)\equiv(\mathrm{AEL}(\mathrm{ms}\ '\kappa)(\forall i\Gamma_i))$ QED.

Theorem AEL2 shows that the set of theorems: $(\mathrm{fol}\ '\kappa)$ of a set $'\kappa$ is a fixed-point of Autoepistemic Logic if and only if the meaning $(\mathrm{ms}\ '\kappa)$ of $'\kappa$ is a solution to the necessary equivalence. Furthermore, by FOL4 there are no other fixed-points (such as a set not containing all its theorems) and by MA6 there are no other solutions (such as a proposition not representable as a sentence in the First Order Logic object language). Therefore the Modal representation of Autoepistemic Logic (i.e. AEL), faithfully represents the original set theoretic description of Autoepistemic Logic (i.e. ael). Finally, we note that $(\mathrm{ms}\ '\kappa)$ and $\forall i\Gamma_i$ may be generalized to be arbitrary propositions κ and Γ giving the more general modal representation: $\kappa\equiv(\mathrm{AEL}\ \kappa\ \Gamma)$.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Numbers: MIP-9526532, EIA-9818341, and EIA-9972843.

Bibliography

[Antoniou 1997] Antoniou, Grigoris 1997. *NonMonotonic Reasoning*, MIT Press.

[Boyer&Moore 1981] R. S. Boyer and J. Strother Moore, "Metafunctions: proving them correct and using them efficiently as new proof procedures," *The Correctness Problem in Computer Science,* R. S. Boyer and J. Strother Moore, eds., Academic Press, New York, 1981.

[Bressan 1972] Bressan, Aldo 1972. *A General Interpreted Modal Calculus,* Yale University Press.

[Brown 1978] Brown,F.M., "A Semantic Theory for Logic Programming", *Colloquia Mathematica Societatis Janos Bolyai 26*, *Mathematical Logic in Computer Science*, Salgotarjan, Hungry, 1978.

[Brown 1987] Brown, Frank M. 1987. "The Modal Logic Z", In *The Frame Problem in AI*"*; Proc. of the 1987 AAAI Workshop*, Morgan Kaufmann, Los Altos, CA .

[Brown 1989] Brown, Frank M. 1989. "The Modal Quantificational Logic Z Applied to the Frame Problem", advanced paper *First International Workshop on Human & Machine Cognition*, May 1989 Pensacola, Florida. Abreviated version published in *International Journal of Expert Systems Research and Applications, Special Issue: The Frame Problem. Part A.* eds. Keneth Ford and Pattrick Hayes, vol. 3 number 3, pp169-206 JAI Press 1990. Reprinted in *Reasoning Agents in a Dynamic World: The Frame problem*, editors: Kenneth M. Ford, Patrick J. Hayes, JAI Press 1991.

[Carnap 1946] Carnap, Rudolf 1946. "Modalities and Quantification" *Journal of Symbolic Logic*, vol. 11, number 2, 1946.

[Carnap 1956] Carnap, Rudolf 1956. *Meaning and Necessity: A Study in the Semantics of Modal Logic*, The University of Chicago Press.

[Hendry & Pokriefka 1985] Hendry, Herbert E. and Pokriefka, M. L. 1985. "Carnapian Extensions of S5", *Journal of Phil. Logic 14*.

[Hughes & Cresswell 1968] Hughes, G. E. and Creswell, M. J., 1968. *An Introduction to Modal Logic*, Methuen & Co. Ltd., London

[Konolige 1987] Konolige, Kurt 1987. "On the Relation Between Default Theories and Autoepistemic Logic", *IJCAI87 Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, 1987.

[Konolige 1987b] Konolige, Kurt 1987. "On the Relation Between Default Logic and Autoepistemic Theories", Artificial Intelligence 35(3):343-382.

[Lewis 1936] Lewis, C. I. 1936. Strict Implication, *Journal of Symbolic Logic*, vol I.

[McDermott 1980] McDermott, D. and Doyle, J. 1980. "Nonmonotonic Logic I" *Artificial Intelligence*, 13.

[McDermott 1982] McDermott, D. 1982. "Nonmonotonic Logic II: Nonmonotonic Modal Theories", JACM, vol. 29, No. 1.

[Mendelson 1964] Mendelson, E. 1964. *Introduction to Mathematical Logic*, Van Norstrand, Reinhold Co., New York.

[Moore 1985] Moore, R. C. 1985. "Semantical Considerations on Nonmonotonic Logic" *Artificial Intelligence*, 25.

[Parks 1976] Parks, Z. 1976. "An Investigation into Quantified Modal Logic", Studia Logica 35, p109-125.

[Quine 1969] Quine, W.V.O., Set Theory and Its Logic, revised edition, Oxford University Press, London, 1969.

Author information

Frank M. Brown- Artificial Intelligence Laboratory, University of Kansas, Lawrence, Kansas, 66045, e-mail: brown@ku.edu.