ON COHERENCE IN MODAL LOGICS

Robert K. MEYER

In this paper, I define a notion of coherence for modal logics, and I develop techniques which show that a wide class of logics are coherent; included in this class are not only familiar logics like S4 but a number of logics, like my system NR, whose non-modal part is distinctly non-classical and, by extension, Anderson and Belnap's E of entailment and Ackermann's *strenge Implikation.* It will follow in particular that these logics have a number of interesting properties, including the S4 property

$$\vdash \Box A \vee \Box B \text{ iff } \vdash \Box A \text{ or } \vdash \Box B.$$

I

Roughly, a logic is coherent if it can be plausibly interpreted in its own metalogic. Specifically, we presume a sentential logic L to be formulated with a necessity operator $\Box$, non-modal connectives $\rightarrow$, $\wedge$, $\vee$, $-$ (and perhaps other connectives and constants which can be correlated with familiar truth-functions), and formulas A, B, C, etc., built up as usual from sentential variables p, q, r, etc. Henceforth we identify L with its so-called *Lindenbaum matrix* — i.e., $L = \langle F,O,T \rangle$, where F is the set of formulas of L, T is the set of theorems, and O is a set of operations corresponding to connectives of L.

Let $\mathbf{2} = \langle 2,O,\{1\} \rangle$ be the 2 element Boolean algebra (considered as a matrix), where $2 = \{O,1\}$ and with operations in 0 corresponding to all non-modal connectives in L and defined as usual. A *metavaluation* of L shall be any function $v: F \rightarrow 2$ satisfying the following conditions, for all formulas A and B:

(i) $v(\Box A) = 1$ iff $\vdash \Box A$ in L;

(ii) $v(A \rightarrow B) = v(A) \rightarrow v(B)$, $v(-A) = -v(A)$, and similarly for other non-modal connectives.

A formula A of L is *true* on a metavaluation v iff v(A) = 1; A is *metavalid* iff A is true on all metavaluations of L; L is *coherent* iff each theorem A of L is metavalid.

The following theorem is trivial, but it generalizes well-known S4 properties to all the logics that we shall prove coherent.

Theorem 1. Let L be a coherent logic. For any formula C, let C′ be a formula which results from C by replacement of truth-functionally equivalent formulas ([1]). Then

(i) $\vdash_L(\Box A \vee \Box B)'$ only if $\vdash_L \Box A$ or $\vdash_L \Box B$;

(ii) $\vdash_L(\Box A \wedge \Box B)'$ only if $\vdash_L \Box A$ and $\vdash_L \Box B$.

Proof. Ad (i). Suppose neither □A nor □B are theorems of L. Then for an arbitrary metavaluation v, v(□A) = 0 and v(□B) = 0, whence $v(\Box A \vee \Box B)' = 0$ on purely truth-functional grounds. Since L is coherent, $\Box A \vee \Box B$ and all its truth-functional equivalents are non-theorems.

Ad (ii). Similar.

II

We shall prove coherent all modal logics which can be formulated with axioms and rules of certain kinds. In order to formulate our results in as general a way as possible, while keeping in mind those cases which are interesting in practice we shall characterize the key notions rather sharply. $A[B_1, \ldots, B_n/p_1, \ldots, p_n]$ shall be the result of uniformly substituting the formulas $B_1, \ldots, B_n$ respectively for the sentential variables $p_1, \ldots, p_n$ in the formula A; s(A) shall be the class of all uniform substitutions in A. Where $\langle A_0, \ldots, A_n \rangle$, $n > 0$, is a finite sequence of formulas, a uniform substitution $\langle A_0, \ldots, A_n \rangle [B_1, \ldots, B_n/p_1, \ldots, p_n]$ shall be the sequence $\langle A_0[B_1, \ldots, B_n/p_1, \ldots, p_n], \ldots, A_n[B_1,$

([1]) A and B are truth-functionally equivalent iff they are uniform substitution instances of formulas A_0 and B_0 such that (1) the sign '□' does not occur in A_0 or B_0 and (2) $A_0 \leftrightarrow B_0$ is a classical tautology.

$\ldots, B_n/p_1, \ldots, p_n\rangle$; $s(A_0, \ldots, A_n)$ shall be the class of all uniform substitutions in $\langle A_0, \ldots, A_n\rangle$.

A *scheme* shall be a pair $\langle A, s(A)\rangle$, where A is called the *characteristic formula* of the scheme. A *rule* shall be a pair $\langle \langle A_0, \ldots, A_n\rangle, s(A_0, \ldots, A_n)\rangle$, where the sequence of formulas $A_0, \ldots, A_n$, $n > 1$, is called the *characteristic sequence* of the rule, A_0 is called the *characteristic conclusion* of the rule, and $A_1, \ldots, A_n$ are called the *characteristic premisses* of the rule. A scheme is *tautologous* if its characteristic formula is a substitution instance of a truth-functional tautology in which the sign '$\Box$' does not occur; a rule is *truth-functional* if the sign '$\Box$' does not occur in its characteristic sequence and if the conditional whose antecedent is the conjunction of its characteristic premisses and whose consequent is its characteristic conclusion is a truth-functional tautology.

Let $L = \langle F, O, T\rangle$ be a logic, let X be a set of schemes, and let R be a set of rules. $\langle X, R\rangle$ is a *formulation* of L provided that T is the smallest set which contains s(A) whenever $\langle A, s(A)\rangle \in X$ and of which $A_0[B_1, \ldots, B_n/p_1, \ldots, p_n]$ is a member whenever $\langle A_0, \ldots, A_n\rangle$ is the characteristic sequence of a member of R and each of $A_1[B_1 \ldots, B_n/p_1, \ldots, p_n], \ldots, A_n[B_1, \ldots, B_n/p_1, \ldots, P_n]$ belongs to T. If $\langle X, R\rangle$ is a formulation of L, we call members of X *axiom schemes* and members of R *primitive rules* of the formulation. Finally, we call a rule r *admissible* for a formulation $\langle X, R\rangle$ of L iff $\langle X, R \cup \{r\}\rangle$ is a formulation of L — i.e., following Curry, if taking r as a new primitive rule does not enlarge the class of theorems.

We shall call a modal logic *regular* only if it has a formulation $\langle X, R\rangle$ satisfying the following conditions:

(1) If $\langle A, s(A)\rangle \in X$, one of the following holds:

(a) $\langle A, s(A)\rangle$ is tautologous;

(b) for some formula B, A is truth-functionally equivalent to $\Box B \rightarrow B$;

(c) for some formula B, A is truth-functionally equivalent to $\Box B \rightarrow \Box\Box B$ and $\langle \Box\Box B, \Box B\rangle$ is the characteristic sequence of an admissible rule of $\langle X, R\rangle$;

(d) for some formulas B and C, A is truth-functionally equivalent $\Box B \& \Box C \rightarrow \Box(B\&C)$ and $\langle \Box(B\&C), \Box B, \Box C\rangle$ is the characteristic sequence of an admissible rule of $\langle X,R\rangle$;

(e) for some formulas B and C, A is truth-functionally equivalent to $\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$ and $\langle \Box C, \Box(B \rightarrow C), \Box B\rangle$ is the characteristic sequence of an admissible rule of $\langle X,R\rangle$;

(f) for some formulas B and C, A is truth-functionally equivalent to $\Box(B \vee C) \rightarrow (-\Box-B \vee \Box C$ and $\langle \Box C, \Box-B, \Box(B \vee C)\rangle$ is the characteristic sequence of an admissible rule of $\langle X,R\rangle$.

(2) If $r \in R$, one of the following holds:

(a) r is truth-functional;

(b) the characteristic sequence of r is $\langle \Box B, B\rangle$ for some formula B.

(c) the characteristic sequence of r is $\langle \Box B \rightarrow \Box C, \Box B \rightarrow C\rangle$ for some formulas B and C, and $\langle \Box C, \Box B \rightarrow \Box C, \Box B\rangle$ is the characteristic sequence of an admissible rule of $\langle X,R\rangle$.

It is readily observed that many familiar modal, deontic, and epistemic logics are regular, including the Lewis systems S2, S3, and S4, the Feys-Gödel-von Wright system M, ,the Lemmon system SO.5, and others. Of particular interest for present purposes is the fact that no conditions are placed on non-modal axioms and rules, save that they be classically valid; thus the Y-systems of Curry's [4] and the relevant modal logic NR of [6] are regular.

We shall show that all regular modal logics are coherent by associating with each of them a special kind of structure. Let L be a regular modal logic. The *weak canonical matrix* W for L is the triple $\langle 2\times F, O, D\rangle$, where $2\times F$ is the set of pairs $\langle x,A\rangle$ such that $x = 0$ or $x = 1$ and A is a formula of L, $\langle x,A\rangle$ belongs to the set D of designated elements of $2\times F$ iff $x = 1$, and O is a set of operations corresponding to the connectives of L and defined as follows on all $\langle x,A\rangle$ and $\langle y,B\rangle$ in $2\times F$:

(I) $\langle x,A\rangle \rightarrow \langle y,B\rangle = \langle x\rightarrow y, A\rightarrow B\rangle$, $-\langle x,A\rangle = \langle -x,-A\rangle$, and similarily for other non-modal connectives and constants;

(II) $\Box\langle x,A\rangle = \langle 1,\Box A\rangle$ iff $x = 1$ and $\Box A$ is a theorem of L; $\Box\langle x,A\rangle = \langle 0,\Box A\rangle$ otherwise.

A canonical interpretation of L in its weak canonical matrix W is any function $f: F \rightarrow 2 \times F$ satisfying the following conditions:

(a) If p is a sentential variable, $f(p) = \langle 0,p\rangle$ or $f(p) = \langle 1,p\rangle$;

(b) $f(A \rightarrow B) = f(A) \rightarrow f(B)$, $f(\Box A) = \Box f(A)$, and similarly for other connectives; if the sentential constant t occurs in L, $f(t) = \langle 1,t\rangle$. A formula A of L is *weakly valid* in W iff $f(A) \in D$ for all canonical interpretations f of L in W. We now prove the key theorem.

Theorem 2. Let L be a regular modal logic, and let W be its weak canonical matrix as defined above. Then for all formulas A of L, the following conditions hold.

(i) If A is a theorem of L, A is weakly valid in W;

(ii) $\Box A$ is a theorem of L iff $f(\Box A) = \langle 1,\Box A\rangle$ for all canonical interpretations f of L in W;

(iii) $\Box A$ is a non-theorem of L iff $f(\Box A) = \langle 0,\Box A\rangle$ for all canonical interpretations f of L in W.

Proof. (iii) follows directly from the definitions of f and W. (ii) follows from (i) and the fact that if $f(\Box A) = \Box f(A) = \langle 1,\Box A\rangle$ for any canonical interpretation f, then by (II) $\Box A$ is a theorem of L.We finish the proof of the theorem by proving (i).

Since L is regular, it has a formulation $\langle X,R\rangle$ satisfying the conditions on p. 660. Hence if A is a theorem of L, there is a sequence of formulas $A_1, \ldots, A_n$ such that A_n is A and such that each A_i, $1 \leqslant i \leqslant n$, is either a substitution instance of the characteristic formula of a member of X or follows from predecessors by virtue of a rule in R. Given such a sequence, we assume on inductive hypothesis that A_h is weakly valid for all h less than arbitrary i, and we show that $f(A_i) = \langle 1,A_i\rangle$ for an arbitrary canonical interpretation f, and hence that A_i is weakly valid. There are two cases, with subcases corresponding to the conditions on regularity of p. 660.

Case 1. $A_i \in s(B)$, where $\langle B, s(B) \rangle$ is an axiom scheme.

(a) B is a truth-functional tautology. Then B, and hence A_i, is a substitution instance of a classical tautology C in which '□' does not occur. But C is weakly valid on purely truth-functional considerations, whence so is A_i.

(b) A_i is truth-functionally equivalent to $\Box C \rightarrow C$, for some formula C. Then $f(A_i)$ is $\Box f(C) \rightarrow f(C)$, which is designated on truth-functional grounds if $f(C) = \langle 1, C \rangle$; if $f(C) = \langle 0, C \rangle$, $\Box f(C) = \langle 0, \Box C \rangle$ by (II) on p. 662, and so truth-functionally $f(A_i) = \langle 1, A_i \rangle$.

(c) A_i is truth-functionally equivalent to $\Box C \rightarrow \Box\Box C$ for some formula C, and if $\Box C$ is a theorem of L so is $\Box\Box C$. By (II) unless it is the case that both $f(C) = \langle 1, C \rangle$ and $\Box C$ is a theorem of L, $f(A_i)$ is designated by falsity of antecedent; in the remaining case, it is designated by truth of consequent.

(d) A_i is truth-functionally equivalent to $\Box C \& \Box D \rightarrow \Box(C\&D)$, where if both $\Box C$ and $\Box D$ are theorems of L so also is $\Box(C\&D)$. By (II) unless it is the case that $f(C) = \langle 1, C \rangle$, $f(D) = \langle 1, D \rangle$, $\Box C$ is a theorem of L, and $\Box D$ is a theorem of L, $f(A_i)$ is designated by falsity of antecedent; in the remaining case, it is designated by truth of consequent.

(e) (f). Similar.

Case 2. A_i follows from predecessors in virtue of a rule $r \in R$, where we may assume all predecessors weakly valid.

(a) r is truth-functional. Then on purely truth-functional grounds, A_i is weakly valid.

(b) A_i is $\Box C$, and for some $h < i$, A_h is C. On inductive hypothesis, $f(C) = \langle 1, C \rangle$ for an arbitrary canonical interpretation f, whence, since $\Box C$ is a theorem of L, $f(\Box C) = \langle 1, \Box C \rangle$.

(c) A_i is $\Box C \rightarrow \Box D$, and for some $h < i$, A_h is $\Box C \rightarrow D$; furthermore, if $\Box C$ and $\Box C \rightarrow \Box D$ are both theorems, so is $\Box D$. We may assume that $f(C) = \langle 1,C \rangle$ and that $\Box C$ is a theorem of L (else $f(A_i)$ is designated by falsity of antecedent). Then $f(\Box C) = \langle 1, \Box C \rangle$; furthermore, since A_h is weakly valid $f(D) = \langle 1,D \rangle$ and, since $\Box D$ is a theorem of L on our assumptions, $f(A_i) = \langle 1,A_i \rangle$ by truth of consequent. This completes the inductive argument and the proof of theorem 2.

Theorem 2 has some interesting applications in addition to those with which we are primarily concerned here. If, for example, we define for a regular modal logic L and L-theory to be any set of formulas of L which contains all theorems of L and which is closed under the truth-functionally valid rules of L, then for each such L there is a consistent and complete L-theory T such that $\Box A \in T$ iff $\Box A$ is a theorem of L, and hence, by consistency and completeness, such that $-\Box A \in T$ iff $\Box A$ is a non-theorem of L. For by theorem 2, it is clear that the set of formulas which take designated values on any canonical interpretation in the weak canonical matrix will constitute such a theory. This suggests in particular a mode of attack on the decision problem for the class of formulas of the form $\Box A$ for any regular modal logic; find a recursive set of axioms for T satisfying the above conditions. Since many modal logics, including NR and E, have the property that A is a theorem iff $\Box A$ is a theorem, the construction of suitable T would solve the decision problem for all formulas, closing long open problems for the systems mentioned (Cf. Anderson's [1]). We return to our main business with a corollary.

Corollary 2.1. Every regular modal logic L is coherent.

Proof. We must show that each theorem A of L is true on an arbitrary metavaluation v. Define a canonical interpretation f of L in the weak canonical matrix W by letting $f(p) = \langle 0,p \rangle$ if $v(p) = 0$ and $f(p) = \langle 1,p \rangle$ if $v(p) = 1$ for each sentential variable p; clearly this suffices to determine the value of f on each formula of L.

We now show that $f(B) = \langle 1,B \rangle$ if $v(B) = 1$ and $f(B) = \langle 0,B \rangle$ if $v(B) = 0$, by induction on the length of B. This is true by specification for sentential variables, and it is trivial on inductive hypothesis if the principal connective of B is non-modal. Suppose finally that B is of the form $\Box C$. If $\Box C$ is a theorem of L, $v(\Box C) = 1$ by definition of a metavaluation and $f(\Box C) = \langle 1, \Box C\rangle$ by (ii) of the theorem; if $\Box C$ is a non-theorem of L, $v(\Box C) = 0$ by definition and $f(\Box C) = \langle 0, \Box C\rangle$ by (iii) of the theorem. This completes the inductive argument, and shows that f(B) agrees with v(B) for arbitrary B.

We complete the proof of the corollary by noting that since by the theorem each theorem A of L is weakly valid, $v(A) = 1$ for all metavaluations v. Hence if A is a theorem, A is metavalid, and so L is coherent.

III

We now apply theorems 1 and 2 to the relevant logics NR and E. That NR is regular is simply a matter of checking the axioms and rules of [6] to see that they meet the conditions of p. 660. This proves that NR has by theorem 1 the S4 disjunction property; it is also establishes that one cannot prove that two apodictic formulas of NR are consistent unless one can prove both formulas. For introducing a consistency operator o into NR via the definition

$$\text{DO.} \quad A \circ B =\text{df}\ \overline{A \rightarrow \overline{B}}$$

then since A&B is truth-functionally equivalent to AoB, if one can prove in NR

$$\Box A \circ \Box B$$

then by theorem 1 one can prove both $\Box A$ and $\Box B$. (The converse is trivial — if one can prove both $\Box A$ and $\Box B$, one can prove in NR that they are consistent). This solves for NR a problem analogous to one raised in [1] by Anderson for E (2).

(2) The problem is not quite analogous, for what Anderson was asking was whether one could prove $\Box A$ consistent with $\Diamond B$ without being able to prove both in E, essentially. In this case it turns out there are formulas A and B of E such that one can, thus refuting Anderson's apparent conjecture.

NR was introduced in [5] because it putatively contained the system E of entailment, and hence the equivalent Ackermann system Π', exactly on the definition

D1. $A \Rightarrow B =\text{df} \Box(A \rightarrow B)$.

This has remained conjecture, however, and so the results we have obtained for NR do not automatically apply to E. Furthermore, in the Anderson-Belnap formulation of E, '$\Box$' is defined by

D2. $\Box A =\text{df} (A \rightarrow A) \rightarrow A$,

wich, were it turned into a definitional axiom for a version of E with '$\Box$' primitive, would not meet our conditions for regularity.

The way out, in order to apply our results, is to embed E is a version NE of itself; for elegance, we suppose E formulated with a sentential constant t (no actual inflation, in view of the elimination procedure of [2]) and the axioms schemes and rules given by E2-E3, E5-E16, and E56-E57 of [5]. To form NE, we take '$\Box$' as an additional primitive and add to the above axiom schemes those whose characteristic formulas are $\Box p \rightarrow p$, $\Box p \rightarrow \Box\Box p$, $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$, $(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$, and $\Box p \rightarrow (t \rightarrow \Box p)$; as a new rule, we add one whose characteristic sequence is $\langle \Box p, p \rangle$. We now have,

Theorem 3. NE is regular. Furthermore, if A^* is the formula of NE got by replacing in a formula A of E each subformula $B \rightarrow C$ with $\Box(B \rightarrow C)$, beginning with innermost parts, then A is a theorem of E iff A^* is a theorem of NE.

Proof. That NE is regular follows from the definition of regularity on p. 4. To show that if A is a theorem of E, A^* is a theorem of NE, it suffices to show that if B is an axiom of E, B^* is a theorem of NE and that *modus ponens* holds for $\Rightarrow$, as defined by D1, in NE; it follows that for each step $A_1, \ldots, A_n$ in a derivation of A in E, A_i^* is a theorem of NE. Actual verification of the axioms of E in NE poses no problems and is left to the reader.

Conversely, suppose A* is a theorem of NE. Replace each occurrence of the primitive sign '□' in its proof with '□' as defined by D2; it is easily seen that each step of the transformed derivation is a theorem of E; hence A*, thus transformed, is a theorem of E. Finish the proof by showing that A*, with '□' defined by D2, entails A in E.

Theorem 3 suggests a new definition of coherence for a system in which entailment is taken as primitive. For E in particular, formulated with t, $\wedge$, $\vee$, —, and $\rightarrow$ primitive, we define a *metavaluation* to be any function defined on the set of formulas of E with values in $\{1,0\}$ satisfying the following conditions, for all formulas A and B:

(i) $v(A \rightarrow B) = 1$ iff $A \rightarrow B$ is a theorem of E;
(ii) $v(t) = 1$;
(iii) $v(A \vee B) = 1$ iff $v(A) = 1$ or $v(B) = 1$;
(iv) $v(A \wedge B) = 1$ iff $v(A) = 1$ and $v(B) = 1$;
(v) $v(—A) = 1$ iff $v(A) = 0$.

As before, we call a formula of E *metavalid* if it is true on all metavaluations; E is *coherent* if all its theorems are metavalid. We then have

Corollary 3.1. E is coherent. Furthermore a formula $A \rightarrow B$ is a theorem of E iff it is true on an arbitrary metavaluation. Accordingly, if the sign '$\rightarrow$' does not occur in C, a formula $(A_1 \rightarrow B_1) \vee \ldots \vee (A_n \rightarrow B_n) \vee C$ is a theorem of E iff either $(A_i \rightarrow B_i)$ is a theorem of E for some i or C is a truth-functional tautology.

Proof. Let A be a formula of E, and let A* be the translation of A into NE given by the theorem. Let v be any metavaluation of E, and let v* be the metavaluation of NE which agrees with v on sentential variables. Use the theorem to show, for each subformula B of A, $v(B) = v^*(B^*)$. But if A is a theorem of E, A* is a theorem of NE and is hence, by the coherence of NE, true on all v*; so A is true on v. But v was arbitrary; hence E is coherent. This proves the first statement; the second is immediate from the definition of a metavaluation.

For that final part of the theorem, Anderson and Belnap noted in [3] that all tautologies in which '→' does not occur are theorems of E. The suffiency part of the last statement then follows by elementary properties of disjunction.

On the other hand, assume that none of $A_1 \to B_1, \ldots, A_n \to B_n$ are theorems of E and that C is not a tautology. Since '→' does not occur in C, there is an assignment of 0 or 1 to sentential variables which falsifies it. The extension of v to a metavaluation will falsify the disjunction, which is accordingly a non-theorem of E.

I remark in conclusion that of course theorem 3 and its corollary are straightforwardly applicable to Ackermann's *strenge Implikation*, in view of the fact that it has the same theorems as E. They are also applicable, *mutatis mutandis*, to related systems straightforwardly translatable into regular modal logics (3).

Indiana University Robert K. MEYER

BIBûLIOGRAPHY

[1] ANDERSON, A. R., "Some open problems concerning the system E of entailment", *Acta Philosophica Fennica, 16* (1963).
[2] ANDERSON, A. R., and N. D. BELNAP, Jr., "Modalities in Ackermann's 'rigorous implication'," *The Journal of Symbolic Logic, 24* (1969), 107-111.
[3] ANDERSON, A. R., and N. D. BELNAP, Jr., "A simple treatment of truth functions", *The Journal of Symbolic Logic, 24* (1959), 301-2.
[4] CURRY, H. B., *Foundations of Mathematical Logic*, N. Y., 1963.
[5] MEYER, R. K., "E and S4", *Notre Dame Journal of Formal Logic, 11* (1970), 181-199.
[6] MEYER, R. K., "Entailment and relevant implication," *Logique et Analyse, n. s.*

(3) My thanks are due to the National Science Foundation, which has partially supported this research through grant GS-2648.