41. Scriven, M., New issues in the logic of explanation, in: Hook, S. (Ed.), *Philosophy and History* (New York University Press, New York, 1963).
42. Stallman, R. M. and Sussman, G. J., Forward reasoning and dependency-directed backtracking in a system for computer-aided circuit analysis, *Artificial Intelligence* 9 (2) (1977) 135–196.
43. Sussman, G. J., Winograd, T. and Charniak, E., MICRO-PLANNER reference manual, MIT Artificial Intelligence Laboratory, AI Memo 203a (1971).
44. Takeuti, G., Formalization principle, in: Van Rootselaar, B. and Staal, J. F. (Eds), *Logic Methodology and Philosophy of Science III*, (North-Holland, Amsterdam, 1968) 105–118.
45. Van Frassen, B., Singular terms, truth-value gaps and free logic, *J. Phil.* **LXIII** (17) (1966) 481–495.
46. Weyhrauch, R. Prolegomena to a theory of formal reasoning, Stanford Artificial Intelligence Laboratory, Memo AIM-315 (1978).

Received 23 January 1979; revised version received 13 July 1979

TECHNICAL NOTE

The Mathematics of Non-Monotonic Reasoning

Martin Davis

Courant Institute of Mathematical Sciences, New York University

Recommended by Patrick J. Hayes

ABSTRACT

Part I of this memo discusses minimal entailment and minimal completion suggested by McCarthy (now included in his more general notion of circumscription) and Part II discusses the "non-monotonic logic" of McDermott and Doyle. McCarthy attempts to capture an idea inherent in Occam's razor: only those objects should be assumed to exist which are minimally required by the context. McDermott and Doyle approach the problem by discussing provability as a modality.

1. Minimal Entailment and Minimal Completion

1.1. Minimal entailment

We will work with a single sorted first-order language with equality and with no function or constant symbols. Later we'll take up needed modifications for dealing with sorts and with function and constant symbols.

Let Γ be a set of sentences. M is a *minimal model* of Γ if:

(1) $M \models A$ for each sentence $A \in \Gamma$,

(2) For every substructure N of M, $N \not\models A$ for some $A \in \Gamma$.

We say that Γ *minimally entails* C and write $\Gamma \models_m C$ to mean that the sentence C is true in all minimal models of Γ. Thus $\Gamma \models C$ implies $\Gamma \models_m C$, but (as will soon be apparent) not conversely.

Minimal entailment was suggested by John McCarthy to capture an idea inherent in Occam's razor: only those objects should be assumed to exist which are minimally required by the context.

Example 1. Let Γ consist of the sentences:

1. $\exists x.Z(x)$
2. $\forall xy.[Z(x) \wedge Z(y) \supset x = y]$
3. $\forall x.\exists y.S(xy)$
4. $\forall xy.[S(x,y) \supset \neg Z(y)]$
7. $\forall xy.[Z(y) \supset A(x,y,x)]$
8. $\forall xyzuv.[A(x,y,z) \wedge S(y,u) \wedge S(z,v) \supset A(x,u,v)]$
9. $\forall xy.[Z(y) \supset P(x,y,y)]$
10. $\forall xyzuv.[P(x,y,z) \wedge S(y,u) \wedge A(z,x,v) \supset P(x,u,v)]$

It is readily observed that there is a unique minimal model of Γ namely the standard model of arithmetic (with $Z(x)$ interpreted as "*x is zero*", $S(x,y)$ interpreted as "*y is the successor of x*", $A(x,y,z)$ interpreted as "$x+y = z$" and $P(x,y,z)$ interpreted as "$xy = z$"). Hence, $\Gamma \models_m$ C just in case C holds in the standard model of arithmetic. Since the set of true sentences of arithmetic is not recursively enumerable (indeed not even arithmetic) it follows that the relation $\models_m$ regarded as a rule of inference is infinitary.

For another example, let $\Gamma = \varnothing$. Since any structure is a model of $\varnothing$, the minimal models of $\varnothing$ are simply the structures whose domain consists of one element. The sentences minimally entailed by $\varnothing$ are then those which are true in every one-element structure, e.g. $\forall xy.x = y$. Let $\Delta = \{\exists xy.x \neq y)\}$. The minimal models of Δ are structures with two-element domains, and the sentences minimally entailed by Δ are those true in every two-element structure. Hence we see that there are sentences minimally entailed by Γ that are not minimally entailed by Δ. Minimal entailment thus lacks a basic monotonicity property common to all usual notions of deduction.

Example 2. Let Γ consist of the sentences:

1. $\forall x.\exists y.S(x,y)$
2. $\exists y.\forall x.\neg S(x,y)$
3. $\forall xyz.[S(x,y) \wedge S(x,z) \supset y = z]$
4. $\forall xyz.[S(y,x) \wedge S(z,x) \supset y = z]$

Then every model of Γ contains a submodel isomorphic to the natural numbers. But this submodel contains an infinite chain of sub-submodels corresponding to the natural numbers $\geqslant k$ for each k. Hence Γ *has no minimal model.*

1.2. Minimal completions

For any formula Ψ, we write Ψ° for the *closure* of Ψ, that is, the sentence obtained from Ψ by prefixing it by universal quantifiers with respect to all of the variables which are free in Ψ. If A is a sentence and $\Phi = \Phi(x)$ is a formula in which the variable x (perhaps among others) is free, we write A^Φ for the sentence obtained from A by relativizing all quantifiers in A to Φ (that is each universal quantifier

$\forall t \cdots$ is replaced by $\forall t.[\Phi(t) \supset \cdots]$ and each existential quantifier $\exists t. \cdots$ is replaced by $\exists t.[\Phi(t) \wedge \cdots]$).

Let A be some sentence of a given first order language. We write

$$\Omega(A) = \{[A^\Phi \supset \forall x.\Phi(x)]^\circ\}$$

where Φ ranges over all formulas of the given language in which the variable x is free. The theory whose axioms are the elements of the set $\Omega(A) \cup \{A\}$ is called the *minimal completion* of A, written MC(A). We write $A \vdash_m B$ to mean that $\vdash_{MC(A)} B$, and say that B is *minimally inferable* from A.

Example 3. Let our language contain a single unary predicate symbol Z. We consider $T = MC(\exists x.Z(x))$. Here $\Omega(\exists x.Z(x)) = \{\exists x.[\Phi(x) \wedge Z(x)] \supset \forall x.\Phi(x)\}$. Taking $\Phi(x) = Z(x)$, we see that $\vdash_T \exists x.Z(x) \supset \forall x.Z(x)$, and hence $\vdash_T \forall x.Z(x)$. Thus we have shown that

$$\exists x.Z(x) \vdash_m \forall x.Z(x)$$

It is obvious that also:

$$\exists x.Z(x) \models_m \forall x.Z(x)$$

The main theorem (conjectured by John McCarthy) is:

Theorem. *If* $A \vdash_m B$ *then* $A \models_m B$.

Proof. Let M be a minimal model of A. Let $C \in \Omega(A)$.

Lemma. $M \models C$.

Proof of Lemma. Let D be the domain of the model M. Let $\Phi = \Phi(x,u_1,u_2, \ldots u_n)$, where the exhibited variables are a complete list of those free in Φ. Let $a_1,a_2, \ldots, a_n \in D$.

Let

$$\Psi(x) = \Phi(x,a_1,a_2, \ldots, a_n)$$

Then it suffices to verify that

$$M \models A^\Psi \supset \forall x.\Psi(x).$$

Thus suppose that $M \models A^\Psi$. Let

$$D_0 = \{a \in D \mid M \models \Psi(a)\}.$$

Let N be the structure obtained from M by restricting all relations to D_0. Then (as can be shown by an easy induction on the length of A), $N \models A$. Since N is a substructure of M and M is a minimal model of A, M = N. Then, $D = D_0$. Hence finally, $M \models \forall x.\Psi(x)$.

Proof of theorem (concluded). Let $A \vdash_m B$, i.e.

$$\{A\} \cup \Omega(A) \vdash B.$$

By the lemma and the fact that M is a model of A,

$$M \models \{A\} \cup \Omega(A),$$

so that $M \models B$. This completes the proof.

Example 1 (continued). Let A be the conjunction of 1.–10. Then as we already have seen, $A \vDash_m C$ if and only if C is a true sentence of arithmetic. We proceed to calculate MC(A). To do so, we first note that:

$$A^{\Phi} = 1^{\Phi} \wedge 2^{\Phi} \wedge \cdots \wedge 10^{\Phi}.$$

Furthermore we observe that if P is a purely universal prenex sentence, then $P \vdash P^{\Phi}$. Thus we have $\vdash_{MC(A)} C$ if and only if

$$\{A\} \cup \{[1^{\Phi} \wedge 3^{\Phi} \supset \forall x.\Phi(x)]^{\circ}\} \vdash C.$$

In other words what we have added to Axioms 1–10 is the scheme:

$$[\exists x.(Z(x) \wedge \Phi(x)) \wedge \forall x.(\Phi(x) \supset \exists y.(S(x,y) \wedge \Phi(y)))] \supset \forall x.\Phi(x).$$

But this is just the scheme of mathematical induction. Thus, MC(A) is just the familiar theory: *Peano arithmetic*. In fact the idea underlying the notion of *minimal completion* is to 'complete' any finite set of axioms in the way that the induction scheme completes the Peano postulates.

Example 2 (continued). Let A be the conjunction of the four sentences of this example. By our remark about purely universal prenex sentences, we have $3 \vdash 3^{\Phi}$; $4 \vdash 4^{\Phi}$. We let

$$\Phi(z) = \exists y.S(y,z),$$

so that we may take the axioms of MC(A) to be A

$$\forall x.[\exists z.S(z,x) \supset \exists y.(\exists z.S(z,y) \wedge S(x,y))]$$

and

$$\exists x.[\exists z.S(z,x) \wedge \forall y.(\exists z.S(z,y) \supset \neg S(y,x))].$$

But it is easy to check that these sentences are all consequences of A. Hence $\vdash_{MC(A)} \forall x.\exists y.S(y,x)$. But this contradicts sentence 2. We conclude that *a consistent set of sentences can possess an inconsistent minimal completion.*

Example 1 shows that the converse of the main theorem is false. Below we give a very weak partial converse:

Theorem. *Suppose that* (1) *every model of A has a minimal submodel, and* (2) *whenever B is true in such a minimal submodel it is also true in the orignal model. Then, $A \vDash_m B$ implies that $A \vdash_m B$ (and indeed that $A \vdash B$).*

Proof. Let K be a model of A. Let M be a minimal submodel of K. By hypothesis we have $M \vDash B$, and using the hypothesis again, $K \vDash B$. We have shown that $A \vdash B$.

Note. Hypothesis (2) is automatically satisfied if B is a purely existential prenex sentence.

1.3. Sorts

It would have been easy to develop the above material in a many-sorted context. But this is not necessary. We have only to construe a many-sorted logic as a single-

sorted one in the usual way. The additional sort axioms need not be considered explicitly in forming the minimal completion since they are purely universal prenex sentences.

1.4. Function and constant symbols

It is well known how to eliminate function and constant symbols by introducing new predicate symbols. Thus to eliminate the constant symbol a we introduce the predicate symbol Z and the axioms:

$$\exists x.Z(x)$$
$$\forall x\, y.[Z(x) \wedge Z(y) \supset x = y]$$

In forming $\Omega(A)$ this will contribute:

$$\exists x.[\Phi(x) \wedge Z(x)]$$

to the antecedent. In the original formulation, this is equivalent to $\Phi(a)$. Similarly if f is a function symbol of (say) 2 arguments, elimination of f leads to a purely universal axiom plus the axiom

$$\forall x\, y.\exists z.F(x,y,z)$$

whose contribution to the antecedent in forming the minimal completion is

$$\forall x\, y.[\Phi(x) \wedge \Phi(y) \supset \exists z.(\Phi(z) \wedge F(x,y,z))].$$

In the original formulation this would have been simply the closure condition

$$\forall x\, y.[\Phi(x) \wedge \Phi(y) \supset \Phi(f(x,y))].$$

In spite of this equivalence, it can be advantageous to avoid function symbols. Thus if Peano's postulates are formulated using function symbols for successor, addition, and multiplication, when we form the minimal completion, the antecedent will contain the closure conditions:

$$\forall x\, y.\Phi(x) \wedge \Phi(y) \supset \Phi(x+y),$$
$$\forall x\, y.\Phi(x) \wedge \Phi(y) \supset \Phi(xy),$$

and this is a weakening of mathematical induction. The fact that these closure conditions are not necessary in the induction scheme, i.e. that a set containing 0 and closed under successor is automatically also closed under addition and multiplication, is expressed by the fact that the recursions for addition and multiplication can be expressed in the predicate symbol formulation as *purely universal prenex sentences*. (Cf. Example 1.)

1.5. Relativization

We write $A,B \vDash_m C$ to mean that C is true in every model of $A \cup B$ which is a minimal model of A.

This corresponds to the intuitive idea of minimizing only with respect to A.

This relativized notion reduces to the original one according to the following theorem:

Theorem. A,B $\models_m$ C *if and only if* A $\models_m$ (B $\supset$ C).

Proof. Let M be a minimal model of A. Then M $\models$ (B $\supset$ C) if and only if M $\models$ B implies M $\models$ C.

We can similarly relativize the notion of minimal completion. Thus we write MC(A,B) for the theory whose axioms are the elements of the set $\{A\} \cup \{B\} \cup \Omega(A)$. And, we write A,B $\vdash_m$ C to mean that $\vdash_{MC(A,B)}$ C. Then we have:

Theorem. A,B $\vdash_m$ C *if and only if* A $\vdash_m$ (B $\supset$ C).

Proof. We have at once that

$$\{A\} \cup \{B\} \cup \Omega(A) \vdash C$$

if and only if

$$\{A\} \cup \Omega(A) \vdash (B \supset C).$$

The above relativizations yield at once a relativized form of our main theorem:

Theorem. A,B $\vdash_m$ C *implies* A,B $\models_m$ C.

2. 'Non-Monotonic' Logic

In this brief memo, I won't be especially concerned with the AI reasons motivating the non-monotonic logic of McDermott and Doyle, but only with some technical clarification. The memo is based on their MIT AI Memo 486 which has also appeared in abbreviated form in the Proceedings of the Fourth Workshop on Automated Deduction, Austin, Texas, February 1979.

2.1. A reformulation

One begins with a system of logic possessing the usual propositional connectives, which may or may not have provision for quantifiers, but which does have a "modal" operator M. Thus for any formula Ψ the system will contain a formula $M\Psi$ whose intended interpretation is the assertion that Ψ is consistent. The provability relation in this system between a set of premises Γ and a conclusion Ψ is as usual written

$$\Gamma \vdash \Psi.$$

The provability relation $\vdash$ is to be one of the standard monotonic relations in the literature (e.g. propositional or predicate calculus) and is not required to take account of the modal operator M (although it may). The relation $\vdash$ will be used to define a non-monotonic relation $\vdash_c$. As usual the formula Ψ is said to be *consistent* with the set of formulas Γ if it is not the case that $\Gamma \vdash \neg\Psi$.

Let $\{\lambda_i \mid i = 1,2,3,\cdots\} = L$ be some enumeration of all of the formulas of the given system. Let Γ be a given set of premises and set:

$$\Gamma_0 = \Gamma$$

$$\Gamma_{i+1} = \text{if } \textit{for some } \beta \in L \quad M\beta \in \Gamma_i \textit{ and } \Gamma_i \vdash \neg\beta, \quad \text{then } L; \\ \text{else} \quad \text{if } \Gamma_i \cup \{\lambda_i\} \textit{ is consistent}, \quad \text{then } \Gamma_i \cup \{M\lambda_i\}; \\ \text{else } \Gamma_i.$$

Finally let $\Gamma_\infty = \cup\Gamma_i$. It is not difficult to see that Γ_∞ may well depend on the particular enumeration of L. Thus (using an example from McDermott and Doyle) let $\Gamma = \{MC \supset \neg D, MD \supset \neg C\}$. Then either MC or MD but not both will be in Γ_∞ according as C or D is encountered first in the enumeration. Now the McDermott and Doyle definition of 'non-monotonic' deducibility amounts to:

$$\Gamma \vdash_c \Psi \textit{ if } \Gamma_\infty \vdash \Psi,$$

for every enumeration of L. Thus in the above example:

$$\Gamma \vdash_c MC \vee MD.$$

For another example note that $\{MC \supset \neg C\} \vdash_c \beta$ for all $\beta \in L$. This is because when C is encountered in the enumeration MC will be adjoined, so that at the next step the first clause in the conditional expression will be activated and we will have $\Gamma_\infty = L$. The non-monotonic behavior is then vividly demonstrated by the extension $\{MC \supset \neg C, C\}$ from which MC cannot be derived (because the presence of C blocks the adjunction of MC). (Both examples are again from McDermott and Doyle.)

2.2. Compactness

McDermott and Doyle state:

> "The most striking result shows that the analogue of the compactness theorem of classical logic does not hold for non-monotonic theories. This has important repercussions on the methods useful in constructing 'models' of theories incrementally."

Now the compactness property for a system of logic can be formulated in various ways which are all equivalent for monotonic logics. But since the key element (from which the notion derives its name) is that in an inference from an infinite set of premises the infinite set can be replaced by a finite subset, it seems clear that the notion of compactness is simply not relevant for a non-monotonic provability relation. The statement on 'important repercussions' is entirely obscure. (See Section 2.6 below.)

2.3. The non-modal case

Call a formula *non-modal* if it can be built up using only propositional connectives from formulas none of which is of the form $M\beta$. Thus non-modal formulas may contain 'M' but only in the scope of quantifiers (so the occurrences are opaque with respect to propositional calculus). Then it is easy to see that if Γ is a set of non-modal formulas and Ψ is a non-modal formula then

$$\Gamma \vdash_c \Psi \quad \textit{if and only if}$$
$$\textit{either } \Gamma \vdash \Psi \textit{ or } \Psi = \neg\beta, \textit{ where } \beta \textit{ is consistent with } \Gamma.$$

2.4. The propositional calculus

It is easy to see that when the original logical system is the classical propositional calculus, there is an algorithm for testing whether $\Gamma \vdash_c \Psi$, where Γ is a finite set of premises. Namely one simply tests all finite sets of subformulas for truth functional consistency with Γ, and is thus presented with a finite list of possible adjunctions to Γ. McDermott and Doyle give a detailed procedure for this case.

2.5. The predicate calculus; degrees of unsolvability

In the general case, the decision problem is of the same degree of unsolvability as the decision problem for the classical predicate calculus, which of course has the same degree of unsolvability as the halting problem. However, the provability relation $\vdash_c$ is not recursively enumerable so there can be no semi-decision procedure. To see this, note first that the procedure outlined in Section 2.4 can be used in the general case if we possess an "oracle" for predicate calculus to be used in carrying out the needed consistency tests. Hence the degree of unsolvability is no greater than that for predicate calculus. On the other hand the set of provable formulas as well as the set of non-provable formulas are each many-one reducible to $\vdash_c$.

2.6. Model theory

In McDermott and Doyle's discussion of model theory, no way is given of associating a truth value (in the interpretation) with $M\beta$ given a value for β. Hence their definition is incomplete. What is worse, the problem seems fatal: there is no obvious way to complete it.

Received 16 September 1979

A Logic for Default Reasoning

R. Reiter

Department of Computer Science, University of British Columbia, Vancouver, B.C., Canada

Recommended by Patrick J. Hayes

ABSTRACT

The need to make default assumptions is frequently encountered in reasoning about incompletely specified worlds. Inferences sanctioned by default are best viewed as beliefs which may well be modified or rejected by subsequent observations. It is this property which leads to the non-monotonicity of any logic of defaults.

In this paper we propose a logic for default reasoning. We then specialize our treatment to a very large class of commonly occurring defaults. For this class we develop a complete proof theory and show how to interface it with a top down resolution theorem prover. Finally, we provide criteria under which the revision of derived beliefs must be effected.

The gods did not reveal, from the beginning,
All things to us, but in the course of time
Through seeking we may learn and know things better.
But as for certain truth, no man has known it,
Nor shall he know it, neither of the gods
Nor yet of all the things of which I speak.
For even if by chance he were to utter
The final truth, he would himself not know it:
For all is but a woven web of guesses.

Xenophanes

1. Introduction and Motivation

Various forms of default reasoning commonly arise in Artificial Intelligence. Such reasoning corresponds to the process of deriving conclusions based upon patterns of inference of the form "in the absence of any information to the contrary, assume . . .". Reasoning patterns of this kind represent a form of plausible inference and are typically required whenever conclusions must be drawn despite the absence of total knowledge about a world.

Artificial Intelligence 13 (1980), 81–132