

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

Copyright © 1980 by North-Holland Publishing Company

In order to fix some of these ideas, we begin by surveying a number of instances of default reasoning as they often occur in the Artificial Intelligence literature. Most of the examples are adapted from those in Reiter (1978a).

1.1. Some examples of default reasoning

1.1.1. Defaults and exceptions

A good deal of what we know about a world is ‘almost always’ true, with a few exceptions. Such facts usually assume the form “Most P ’s are Q ’s” or “Most P ’s have property Q ”. For example most birds fly except for penguins, ostriches, the Maltese falcon etc. Given a particular bird, we will conclude that it flies unless we happen to know that it satisfies one of these exceptions. How is the fact that most birds fly to be represented? The natural first order representation explicitly lists the exceptions to flying:

$$(x). \text{BIRD}(x) \wedge \neg \text{PENGUIN}(x) \wedge \neg \text{OSTRICH}(x) \wedge \dots \supset \text{FLY}(x).$$

But with this representation one cannot conclude of a ‘general’ bird that it can fly. To see why, consider an attempt to prove $\text{FLY}(\text{tweety})$ where all we know of tweety is that it is a bird. Then we must establish the subgoal

$$\neg \text{PENGUIN}(\text{tweety}) \wedge \neg \text{OSTRICH}(\text{tweety}) \wedge \dots$$

which is impossible given that there is no further information about tweety. We are blocked from concluding that tweety can fly even though intuitively we want to deduce just that.

What is required is somehow to allow tweety to fly *by default*. How is this default to be interpreted? We take it to mean something like “If x is a bird, then in the absence of any information to the contrary, infer that x can fly”. The problem then is to interpret the phrase “in the absence of any information to the contrary”. The interpretation we adopt is “It is consistent to assume that x can fly”. Thus “If x is a bird and it is consistent to assume that x can fly, then infer that x can fly”. We represent this more formally as the following *default rule*:

$$\frac{\text{BIRD}(x) : \text{MFLY}(x)}{\text{FLY}(x)} \quad ^1 \quad (1.1)$$

Here M is to be read as “it is consistent to assume”. The exceptions to flight are then given a standard first order representation.

$$(x). \text{PENGUIN}(x) \supset \neg \text{FLY}(x)$$

$$(x). \text{OSTRICH}(x) \supset \neg \text{FLY}(x)$$

etc.

Notice that if $\text{FLY}(\text{tweety})$ is inferred by default then the assertion $\text{FLY}(\text{tweety})$ has the status of a *belief*; it is subject to change, say by the subsequent discovery

¹ In plain text we use the notation $\text{BIRD}(x) : \text{MFLY}(x) / \text{FLY}(x)$.

that tweety is a penguin. We can then reinterpret the default rule (1.1) as saying “If x is a bird and it is consistent to believe that x can fly then one may believe that x can fly”.

There still remains a problem of interpretation. For exactly what is meant by the consistency requirement associated with a default? Consistent with what? Providing an appropriate formal definition of this consistency requirement is perhaps the thorniest issue in defining a logic for default reasoning, and we defer this question to Section 2. For the time being a good intuitive interpretation is to view this consistency requirement with respect to all of the first order facts about the world, together with all of the other beliefs sanctioned by all of the other default rules in force.

Notice that we have provided a representation for the ‘fuzzy’ quantifier ‘most’ or ‘almost all’ in terms of defaults, without appealing to frequency distributions or fuzzy logics.

Notice also that the dual ‘fuzzy’ quantifier ‘few’ has a dual representation as a default. For example “Few Americans are socialists” has the representation

$$\frac{\text{AMERICAN}(x) : M \neg \text{SOCIALIST}(x)}{\neg \text{SOCIALIST}(x)}$$

1.1.2. *Frames, knowledge representation languages and defaults*

Default reasoning plays a prominent role in the influential frames proposal of Minsky (1975). See also Hayes (1977). At least two knowledge representation languages, FRL (Roberts and Goldstein (1977)), and KRL (Bobrow and Winograd (1977)) have been designed as attempts, in part, to computationally represent some of Minsky’s proposals. Not surprisingly, both FRL and KRL provide for default assignments. For example, in KRL the unit for a person in an airline travel system has the form:

```
[Person UNIT Basic
  ⋮
  <hometown{(a City) Palo Alto; DEFAULT}>
  ⋮
  :].
```

This declaration can be regarded as an instruction to the KRL interpreter to proceed as follows:

Whenever x is a person, then in the absence of any information to the contrary assume $\text{hometown}(x) = \text{Palo Alto}$.

In view of the previous discussion, this can be represented by the following default rule:

$$\frac{\text{PERSON}(x) : M \text{hometown}(x) = \text{Palo Alto}}{\text{hometown}(x) = \text{Palo Alto}}$$

1.1.3. *The closed world assumption*

Deductive question–answering systems often implicitly appeal to a particular form of default reasoning in the process of answering queries. As an illustration of the default involved, consider a data base representing an airline flight schedule, and the query “Does Air Canada flight 113 connect Vancouver with New York?” A deductive question–answering system will typically treat the data base together with some general knowledge about the flight domain as a set of premises from which it will attempt to prove $\text{CONNECT}(\text{AC113}, \text{Van}, \text{NY})$. If this proof succeeds, then the system responds “yes”. The interesting case is when this proof fails, for then such a system will typically respond “no”, i.e. it will conclude $\neg \text{CONNECT}(\text{AC113}, \text{Van}, \text{NY})$. Failure to find a proof has sanctioned an inference. Now $\text{CONNECT}(\text{AC113}, \text{Van}, \text{NY})$ is not provable iff $\neg \text{CONNECT}(\text{AC113}, \text{Van}, \text{NY})$ is consistent with the data base so that the derivation of $\neg \text{CONNECT}(\text{AC113}, \text{Van}, \text{NY})$ by failing to prove $\text{CONNECT}(\text{AC113}, \text{Van}, \text{NY})$ is justified by the following closed world default rule:

$$\frac{: M \neg \text{CONNECT}(x, y, z)}{\neg \text{CONNECT}(x, y, z)} .$$

In general, if R is an n -ary relation, then the following is a *closed world default* for R :

$$\frac{: M \neg R(x_1, \dots, x_n)}{\neg R(x_1, \dots, x_n)} .$$

If this default is in force for all relations R of some domain, then reasoning is being done under the *closed world assumption* (Reiter (1978b)). In effect, the closed world assumption says that for any relation R , and any individuals x_1, \dots, x_n one can assume $\neg R(x_1, \dots, x_n)$ whenever it is consistent to do so.

Reasoning about a world under the closed world assumption considerably simplifies the representation of that world. In effect, only positive information about the world need be explicitly represented in the data base. Negative information is not so represented, but is inferred by default. Since in general the amount of negative knowledge about a world vastly exceeds the positive things known about it, there is a considerable computational and representational advantage to reasoning under the closed world assumption. For example, in the airline data base we would not have to represent the facts that AC113 does not connect London and Bombay, or Toronto and Boston, or New York and Paris, etc.

There are a number of other settings in Artificial Intelligence where the closed world assumption is made. Reiter (1978a) discusses two of these in some detail, specifically blocks world problem-solving, and reasoning about taxonomies (IS-A hierarchies).

1.1.4. *The frame default*

The frame problem (Raphael (1971)) arises in the representation of dynamic worlds. The problem stems from the need to represent those aspects of the world which remain invariant under certain state changes. For example, moving a particular object or switching on a light will not change the colours of any objects in the world. In a first order representation of such worlds, it is necessary to explicitly represent all of the invariants under all state changes. These are referred to as the frame axioms for the world being modelled. Thus, to represent the fact that painting an object does not alter the locations of objects would require, in the situational calculus of McCarthy and Hayes (1969) a frame axiom something like

$$(x y z s c)\text{LOCATION}(x,y,s) \supset \text{LOCATION}(x,y,\text{paint}(z,c,s)),$$

where s is a state variable, x and z are objects, c is a colour, and y is some location.

The problem is that in general a vast number of such axioms will be required, e.g. object locations also remain invariant when lights are switched on, when it thunders, when someone speaks, etc. so there is a major difficulty in even articulating a deductively adequate set of frame axioms for a given world.

A solution to the frame problem is a representation of the world coupled with appropriate rules of inference such that the frame axioms are neither explicitly represented nor explicitly used in reasoning about the world. We will focus on a proposed solution in Sandewall (1972). A related approach is described in Hayes (1973). Sandewall proposes a modal operator, UNLESS, which takes a first order formula as argument. The intended interpretation of UNLESS(W) is “ W can not be proved” or in terms of consistency, “ $\neg W$ is consistent.” Sandewall proposes a single ‘frame inference rule’ which, in our notation for defaults, can be paraphrased as a default schema: For all relations R which take a state variable as an argument, and for all state transition functions f

$$\frac{R(\mathbf{x},s) : MR(\mathbf{x},f(\mathbf{x},s))}{R(\mathbf{x},f(\mathbf{x},s))}.$$

Intuitively this default schema formalizes the so-called ‘STRIPS assumption’ (Waldinger (1975)): Every action (state change) is assumed to leave every relation unaffected unless it is possible to deduce otherwise.

1.2. The non monotonic character of default reasoning

A fundamental feature of first order logic is that it is *monotonic*, i.e. if A and B are sets of first order formulae and $A \vdash w$ then $A \cup B \vdash w$. What was valid in the presence of information A remains valid when new information B is discovered. In contrast, any logic which presumes to formalize default reasoning must be non-monotonic. To see why consider a simple theory consisting of the single default : MA/B . Thus B may be believed. If subsequently $\neg A$ is discovered to hold (say through some observation) we then have the new theory : $MA/B, \neg A$ in which B

cannot be believed. One extremely interesting approach to the formalization of a non-monotonic logic of beliefs is described in McDermott and Doyle (1978).

This last example also demonstrates the need for some kind of mechanism for *revising beliefs* in the presence of new information. What this amounts to is the need to record, with each derived belief, the default assumptions made in deriving that belief. Should subsequent observations invalidate the default assumptions supporting some belief, then it must be removed from the data base. The Truth Maintenance System of Doyle (1978) is an heuristic implementations of just such a system for belief revision.

1.3. Defaults and incomplete knowledge

Consider the following natural pair of defaults which should be capable of peaceful co-existence:

“Assume a person’s hometown is that of his/her spouse.”

$$\frac{\text{SPOUSE}(x,y) \wedge \text{hometown}(y) = z : \text{M hometown}(x) = z}{\text{hometown}(x) = z}$$

“Assume a person’s hometown is where his/her employer is located.”

$$\frac{\text{EMPLOYER}(x,y) \wedge \text{location}(y) = z : \text{M hometown}(x) = z}{\text{hometown}(x) = z}$$

Suppose Mary’s spouse lives in Toronto while her employer is located in Vancouver. By the first default, Mary’s hometown is Toronto, and by the second it is Vancouver. To believe both is inconsistent since hometown is a function. If first we derive Toronto, we are then blocked from deriving Vancouver. On the other hand if we first derive Vancouver then the derivation of Toronto is blocked. We can derive one or the other, but not both simultaneously. From the point of view of conventional logical systems this is a perplexing, seemingly incoherent situation. From our intuitive perspective that default assumptions lead to beliefs, the example makes perfect sense; one can choose to believe that Mary’s hometown is Toronto or that it is Vancouver, but not both.² It would seem that defaults can sanction different sets of beliefs about a world.

Now a point of view that begins to make formal sense of this example is the following:

Imagine a first order formalization of what it is we know about any reasonably complex world. Since we cannot know everything about that world—there will be gaps in our knowledge—this first order theory will be incomplete. Nevertheless, there will arise situations in which it is necessary to act, to draw some inferences,

² There is another possibility, which is to believe the proposition $\text{hometown}(\text{Mary}) = \text{Toronto} \vee \text{hometown}(\text{Mary}) = \text{Vancouver}$. In terms of the theory which we develop in the remainder of this paper, this possibility amounts to refusing to choose between the Toronto extension and the Vancouver extension. See the following discussion.

despite the incompleteness of the knowledge base. That role of a default is to fill in some of the gaps in the knowledge base, i.e. to further complete the underlying incomplete first order theory, so as to permit the inferences necessary to act. Defaults therefore function somewhat like meta-rules; they are instructions about how to create an *extension* of this incomplete theory. Those formulae sanctioned by the defaults and which extend the theory can then be viewed as beliefs about the world. Now in general there are many different ways of extending an incomplete theory, which suggests that the default rules may be nondeterministic. Different applications of the defaults yield different extensions and hence different sets of beliefs about the world.

As an example of this point of view, consider the closed world assumption. This fully completes the underlying first order theory. Thus, the theory $p \vee q$ under the closed world defaults $: M \neg p / \neg p$ and $: M \neg q / \neg q$ has two different complete extensions, $\{p, \neg q\}$ and $\{\neg p, q\}$. In the first we choose to believe $\neg q$ by invoking the second default, whence we are forced to believe p . In the second we choose to believe $\neg p$ by invoking the first default.

It is this point of view—that defaults are rules for extending an underlying incomplete first order theory—which is adopted in this paper, and which motivates the subsequent development. The rest of this paper is devoted to articulating a logic for default reasoning. We see a two-fold task for such a logic:

(1) To provide a formal definition of the extensions to an underlying first order theory induced by a set of defaults.

(2) To provide a proof theory in the form of a procedure which, given a formula w , determines whether w can be believed, i.e., whether there is an extension induced by the defaults which contains w .

In this paper we propose a definition for the first task. The second will turn out to be intractable for default theories in general. However, by focusing upon a special class of defaults called normal defaults, we can provide a proof theory. Fortunately the class of normal defaults embraces a broad spectrum of naturally occurring defaults, for instance all of the examples of this section. In addition to providing a proof theory for normal defaults, we obtain the pessimistic result that in general the beliefs of a theory are not recursively enumerable. Finally, we determine conditions under which it is necessary to revise a set of derived beliefs when confronted with some new observations about a world.

2. Default Theories and their Extensions

In this section we provide a precise notion of a default theory, and then proceed to formally define the concept of an extension for such a theory. In view of the discussion of Section 1.3 these extensions are meant to formally specify the set of beliefs induced by a set of defaults, as a way of further completing some underlying incomplete collection of facts about a world.

2.1. Formal preliminaries

We shall be dealing with a first order language L_A consisting of the usual well formed formulae (wffs) over an alphabet A of variables, function and predicate letters, logical constants, and suitable punctuation signs. 0-ary function letters will sometimes be called constant letters. More specifically, L_A is the set of first order wffs which can be formed using an alphabet A consisting of countably many variables $x, y, z, x_1, y_1, z_1, \dots$, countably many function letters a, b, c, f, g, h, \dots , countably many predicate letters P, Q, R, \dots , the usual punctuation signs, and the standard logical constants \neg (not), \wedge (and), \vee (or), \supset (implies) and quantifiers (x) (for all x), $(\exists x)$ (there exists an x). As the alphabet A will usually remain fixed throughout this paper, we will generally write L instead of L_A for the first order language.

As usual, a wff is said to be *closed* iff it contains no free variables. For any set of closed wffs S , and any closed wff w , $S \vdash w$ means that w is first order provable from premises S . For any set of closed wffs $S \subseteq L$, define $\text{Th}_L(S) = \{w \mid w \in L, w \text{ is closed, and } S \vdash w\}$.

A *default* is any expression of the form

$$\frac{\alpha(\mathbf{x}) : M\beta_1(\mathbf{x}), \dots, M\beta_m(\mathbf{x})}{w(\mathbf{x})},$$

where $\alpha(\mathbf{x}), \beta_1(\mathbf{x}), \dots, \beta_m(\mathbf{x}), w(\mathbf{x})$ are wffs whose free variables are among those of $\mathbf{x} = x_1, \dots, x_n$. $\alpha(\mathbf{x})$ is called the *prerequisite* of the default, and $w(\mathbf{x})$ is its *consequent*. A default is *closed* iff none of $\alpha, \beta_1, \dots, \beta_m, w$ contains a free variable. None of the defaults of Section 1 is closed. An (artificial) example of a closed default is

$$\frac{(x)(P(x) \vee Q(x)) : M(\exists y)(z)R(y, z) \wedge P(y), M(\exists y)Q(y)}{(x)(\exists y)R(x, y) \wedge P(x)}.$$

A *default theory* is a pair (D, W) where D is a set of defaults and W a set of closed wffs. Notice that D and/or W need not be finite; they will be at most countably infinite however, in view of the countability of L . A default theory (D, W) is *closed* iff every default of D is closed.

2.2. The extensions of a closed default theory

We begin by defining the concept of an extension for closed default theories only. Later, in Section 7, we shall suitably generalize this notion for arbitrary theories.

The intuitive idea which must be captured is that of a set of defaults D inducing an extension of some underlying incomplete set of first order wffs W . In view of the discussion in Section 1.3 we cannot expect this extension to be unique. Any such extension will be interpreted as an acceptable set of beliefs that one may hold about the incompletely specified world W . There are three properties which can reasonably be expected of such an extension E :

- (1) It should contain W ; $W \subseteq E$.
 (2) It should be deductively closed; $\text{Th}_L(E) = E$.
 (3) Suppose $(\alpha : M\beta_1, \dots, M\beta_m/w)$ is a default. If $\alpha \in E$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$ (so that each of β_1, \dots, β_m is consistent with E) then $w \in E$. Thus, if α is believed, and if each of β_1, \dots, β_m can be consistently believed, then w is believed.

This motivates the following definition:

Definition 1. Let $\Delta = (D, W)$ be a closed default theory, so that every default of D has the form $(\alpha : M\beta_1, \dots, M\beta_m/w)$ where $\alpha, \beta_1, \dots, \beta_m, w$ are all closed wffs of L . For any set of closed wffs $S \subseteq L$ let $\Gamma(S)$ be the smallest set satisfying the following three properties:

- D1. $W \subseteq \Gamma(S)$
 D2. $\text{Th}_L(\Gamma(S)) = \Gamma(S)$
 D3. If $(\alpha : M\beta_1, \dots, M\beta_m/w) \in D$ and $\alpha \in \Gamma(S)$, and $\neg\beta_1, \dots, \neg\beta_m \notin S$ then $w \in \Gamma(S)$.

A set of closed wffs $E \subseteq L$ is an *extension* for Δ iff $\Gamma(E) = E$, i.e. iff E is a fixed point of the operator Γ .

The next theorem provides a more intuitive characterization of extensions.

Theorem 2.1. Let $E \subseteq L$ be a set of closed wffs, and let $\Delta = (D, W)$ be a closed default theory. Defin²

$$E_0 = W$$

and for $i \geq 0$

$$E_{i+1} = \text{Th}_L(E_i) \cup \left\{ w \mid \frac{\alpha : M\beta_1, \dots, M\beta_m}{w} \in D \right. \\ \left. \text{where } \alpha \in E_i \text{ and } \neg\beta_1, \dots, \neg\beta_m \notin E_i \right\}^1.$$

Then E is an extension for Δ iff

$$E = \bigcup_{i=0}^{\infty} E_i.$$

Proof. We begin by observing that $\bigcup_{i=0}^{\infty} E_i$ enjoys the following properties:

- D1'. $W \subseteq \bigcup_{i=0}^{\infty} E_i$
 D2'. $\text{Th}_L\left(\bigcup_{i=0}^{\infty} E_i\right) = \bigcup_{i=0}^{\infty} E_i$
 D3'. If $\frac{\alpha : M\beta_1, \dots, M\beta_m}{w} \in D$ and $\alpha \in \bigcup_{i=0}^{\infty} E_i$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$,
 then $w \in \bigcup_{i=0}^{\infty} E_i$.

² Note the occurrence of E in the definition of E_{i+1} .

Hence, by the minimality of $\Gamma(E)$, we have

$$\Gamma(E) \subseteq \bigcup_{i=0}^{\infty} E_i. \quad (2.1)$$

(\Rightarrow) We inductively prove $E_i \subseteq E$ for all $i \geq 0$, whence $\bigcup_{i=0}^{\infty} E_i \subseteq E$. Clearly, since $E = \Gamma(E)$, $E_0 \subseteq E$. Assume $E_i \subseteq E$ and consider $w \in E_{i+1}$. If $w \in \text{Th}_L(E_i)$, then since $E_i \subseteq E$ we have $w \in \text{Th}_L(E) = E$. Otherwise there is a default $(\alpha : M\beta_1, \dots, M\beta_m/w) \in D$, where $\alpha \in E_i$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$. Then since $E_i \subseteq E$, $\alpha \in E = \Gamma(E)$. Hence $w \in \Gamma(E)$ by D3 of Definition 1 and since $\Gamma(E) = E$, we have $w \in E$.

Thus $\bigcup_{i=0}^{\infty} E_i \subseteq E$. By (2.1) and the fact that $E = \Gamma(E)$ we have $E = \bigcup_{i=0}^{\infty} E_i$.

(\Leftarrow) We inductively prove $E_i \subseteq \Gamma(E)$ for all $i \geq 0$, whence $E = \bigcup_{i=0}^{\infty} E_i \subseteq \Gamma(E)$. By invoking (2.1) we will then have $E = \Gamma(E)$ whence E is an extension for Δ .

Clearly $E_0 \subseteq \Gamma(E)$, so assume $E_i \subseteq \Gamma(E)$ and consider $w \in E_{i+1}$. If $w \in \text{Th}_L(E_i)$, then since $E_i \subseteq \Gamma(E)$ we have $w \in \text{Th}_L(\Gamma(E)) = \Gamma(E)$. Otherwise there is a default $(\alpha : M\beta_1, \dots, M\beta_m/w) \in D$ where $\alpha \in E_i$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$. Then since $E_i \subseteq \Gamma(E)$, $\alpha \in \Gamma(E)$. Hence $w \in \Gamma(E)$ by D3 of Definition 1. Hence $E_{i+1} \subseteq \Gamma(E)$.

Example 2.1.

$$D = \left\{ \frac{:MA}{A}, \frac{:MB}{B}, \frac{:MC}{C} \right\} \quad W = \{B \supset \neg A \wedge \neg C\}$$

This has two extensions given by

$$E_1 = \text{Th}(W \cup \{A, C\}) \quad E_2 = \text{Th}(W \cup \{B\}).$$

Example 2.2.

$$D = \left\{ \frac{:MC}{\neg D}, \frac{:MD}{\neg E}, \frac{:ME}{\neg F} \right\} \quad W = \emptyset.$$

This is example (26) of McDermott and Doyle (1978) which is a variant of an example in Sandewall (1972). This theory has one extension given by

$$E = \text{Th}(\{\neg D, \neg F\})$$

Example 2.3.

$$D = \left\{ \frac{:MC}{\neg D}, \frac{:MD}{\neg C} \right\} \quad W = \emptyset$$

This is example (27) of McDermott and Doyle (1978). This theory has two extensions:

$$E_1 = \text{Th}(\{\neg C\}) \quad E_2 = \text{Th}(\{\neg D\}).$$

Example 2.4.

$$D = \left\{ \frac{:MA}{A}, \frac{B:MC}{C}, \frac{D \wedge A:ME}{E}, \frac{C \wedge E:M\neg A, M(D \vee A)}{F} \right\}$$

$$W = \{B, C \supset D \vee A, A \wedge C \supset \neg E\}$$

This has three extensions:

$$E_1 = \text{Th}(W \cup \{A, C\}) \quad E_2 = \text{Th}(W \cup \{A, E\}) \quad E_3 = \text{Th}(W \cup \{C, E, F\}).$$

Example 2.5.

$$D = \left\{ \frac{A:M(\text{Ex})P(x)}{(\text{Ex})P(x)}, \frac{:MA}{A}, \frac{:M\neg A}{\neg A} \right\} \quad W = \emptyset$$

This has two extensions:

$$E_1 = \text{Th}(\{\neg A\}) \quad E_2 = \text{Th}(\{A, (\text{Ex})P(x)\}).$$

What is interesting about this example is that in the second extension, the existence of some individual is believed, while this belief is not held in the first extension. *Different extensions for the same default theory can support different ontologies.* Defaults leading to an ontological commitment of some kind are quite common. For example, "Most American adults have a car".

$$\frac{\text{AMERICAN}(x) \wedge \text{ADULT}(x) : M(\text{Ey})\text{CAR}(y) \wedge \text{OWNS}(x,y)}{(\text{Ey})\text{CAR}(y) \wedge \text{OWNS}(x,y)}$$

Unfortunately, there are default theories with no extensions.

Example 2.6.

$$D = \left\{ \frac{:MA}{\neg A} \right\} \quad W = \emptyset$$

This theory has no extension.

It is easy to see that in general default theories are *non-monotonic*. By this we mean that if $\Delta = (D, W)$ is a default theory with an extension E , D' is a set of defaults and W' is a set of first order wffs then $\Delta' = (D \cup D', W \cup W')$ may have no extension E' such that $E \subseteq E'$.

We shall say that a closed default theory has an *inconsistent extension* iff one of its extensions is the set of all closed wffs of L . As immediate corollaries of Theorem 2.1 we have:

Corollary 2.2. *A closed default theory (D, W) has an inconsistent extension iff W is inconsistent.*

Corollary 2.3. *If a closed default theory has an inconsistent extension then this is its only extension.*

A closed default theory is *consistent* iff it has a consistent extension. By Corollary 2.2 if (D, W) is consistent then W is consistent.

Theorem 2.4 (Minimality of Extensions). *If E and F are extensions for a closed default theory (D, W) and if $E \subseteq F$, then $E = F$.*

Proof. By Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$, $F = \bigcup_{i=0}^{\infty} F_i$. We inductively prove $F_i \subseteq E_i$ for all $i \geq 0$, whence $F \subseteq E$ whence $E = F$.

Trivially $F_0 \subseteq E_0$. Assume $F_i \subseteq E_i$ and consider $w \in F_{i+1}$. If $w \in \text{Th}_L(F_i)$ then, since $F_i \subseteq E_i$, $w \in \text{Th}_L(E_i) \subseteq E_{i+1}$. Otherwise there is a default $(\alpha : M\beta_1, \dots, M\beta_m)/w) \in D$ where $\alpha \in F_i$ and $\neg\beta_1, \dots, \neg\beta_m \notin F$. Since $F_i \subseteq E_i$ and $E \subseteq F$ we have $\alpha \in E_i$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$. Hence $w \in E_{i+1}$.

Definition 2. Suppose $\Delta = (D, W)$ is a closed default theory and E is an extension for Δ . The set of *generating defaults* for E with respect to Δ is defined to be

$$\text{GD}(E, \Delta) = \left\{ \frac{\alpha : M\beta_1, \dots, M\beta_m}{w} \in D \mid \alpha \in E \text{ and } \neg\beta_1, \dots, \neg\beta_m \notin E \right\}.$$

If D is any set of defaults (not necessarily closed) then

$$\text{CONSEQUENTS}(D) = \left\{ w(\mathbf{x}) \mid \frac{\alpha(\mathbf{x}) : M\beta_1(\mathbf{x}), \dots, M\beta_m(\mathbf{x})}{w(\mathbf{x})} \in D \right\},$$

i.e., it is the set of consequents of the defaults of D .

The next theorem justifies the terminology of 'generating default'.

Theorem 2.5. *Suppose E is an extension for a closed default theory $\Delta = (D, W)$. Then $E = \text{Th}_L(W \cup \text{CONSEQUENTS}(\text{GD}(E, \Delta)))$*

Proof. Using Theorem 2.1 we inductively prove that $E_i \subseteq \text{RHS}$ for all $i \geq 0$, whence $E = \bigcup_{i=0}^{\infty} E_i \subseteq \text{RHS}$. Trivially $E_0 \subseteq \text{RHS}$. Assume $E_i \subseteq \text{RHS}$ and let $\theta \in E_{i+1}$. If $\theta \in \text{Th}_L(E_i)$ then since $\text{Th}_L(E_i) \subseteq \text{RHS}$, $\theta \in \text{RHS}$. Otherwise θ is w where $(\alpha : M\beta_1, \dots, M\beta_m/w) \in D$, $\alpha \in E_i$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$. Since $\alpha \in E_i$, $\alpha \in E$. Hence $(\alpha : M\beta_1, \dots, M\beta_m/w) \in \text{GD}(E, \Delta)$ so that

$\theta \in \text{CONSEQUENTS}(\text{GD}(E, \Delta)) \subseteq \text{RHS}$.

Next we prove $\text{RHS} \subseteq E$. It is clearly sufficient to prove

$W \cup \text{CONSEQUENTS}(\text{GD}(E, \Delta)) \subseteq E$ since $\text{Th}_L(E) = E$.

Obviously $W \subseteq E$ so it is sufficient to prove $\text{CONSEQUENTS}(\text{GD}(E, \Delta)) \subseteq E$. If $w \in \text{CONSEQUENTS}(\text{GD}(E, \Delta))$ then there is a default $(\alpha : M\beta_1, \dots, M\beta_m/w) \in D$ such that $\alpha \in E$ and $\neg\beta_1, \dots, \neg\beta_m \notin E$. Since, by Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$ then $\alpha \in E_i$ for some i which means $w \in E_{i+1} \subseteq E$.

We shall later require the following result:

Theorem 2.6. *Suppose that E is an extension for a closed default theory (D, W) , and that $B \subseteq E$. Then E is also an extension for $(D, W \cup B)$.*

Proof. By Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$. Let $F_0 = W \cup B$ and for $i \geq 0$

$$F_{i+1} = \text{Th}_L(F_i) \cup \left\{ w \mid \frac{\alpha: M\beta_1, \dots, M\beta_m}{w} \in D, \right. \\ \left. \text{where } \alpha \in F_i \text{ and } \neg\beta_1, \dots, \neg\beta_m \notin E \right\}.$$

We prove that $\bigcup_{i=0}^{\infty} F_i = E$ whence E will be an extension for $(D, W \cup B)$. An easy inductive proof establishes that $E_i \subseteq F_i$ for all $i \geq 0$ so that $E \subseteq \bigcup_{i=0}^{\infty} F_i$. To prove $\bigcup_{i=0}^{\infty} F_i \subseteq E$ we inductively prove that for each $i \geq 0$ there is a $k \geq 0$ such that $F_i \subseteq E_k$. Now $B \subseteq E = \bigcup_{i=0}^{\infty} E_i$ so $B \subseteq E_k$ for some k . Since $W \subseteq E_j$ for all $j \geq 0$, $F_0 = W \cup B \subseteq E_k$. The inductive step is straightforward.

There are a number of close parallels between default logic and the non-monotonic logic of McDermott and Doyle (1978). The necessary translation from our logic to theirs is to map a default $\alpha(x): M\beta_1(x), \dots, M\beta_m(x)/w(x)$ into the modalized formula

$$(x). \alpha(x) \wedge M\beta_1(x) \wedge \dots \wedge M\beta_m(x) \supset w(x). \quad (2.2)$$

Default theories thus map into a special class of non-monotonic theories, namely those non-monotonic theories whose modalized formulae all have the form (2.2). The approach of McDermott and Doyle is therefore far more general than ours, since they admit arbitrary modalized formulae into their theory. However, when relativized to formulae of the form (2.2) non-monotonic logic has many similarities with default logic. The extensions of a default theory obviously correspond to the fixed points of a non-monotonic theory. Both admit theories with no extensions and fixed points respectively. Corollaries 2.2 and 2.3, and Theorem 2.4 have their analogues in non-monotonic logic.

A major distinction between the two approaches arises from the different ways that extensions and non-monotonic fixed points are defined. While both are defined to be fixed points of certain operators, these operators differ significantly. As a result there are default theories whose extensions differ from the fixed points of the corresponding non-monotonic theories, even when the modalized elements of the latter are ignored. As an example, consider the default theory $\Delta = (\{A: MB/B, C: MD/D\}, \{A \vee C\})$. This has the unique extension $\text{Th}(\{A \vee C\})$. The corresponding non-monotonic theory is $\{A \wedge MB \supset B, C \wedge MD \supset D, A \vee C\}$ and this has a unique fixed point which contains $B \vee D$. In some instances there are default theories which have an extension while the corresponding non-monotonic theory has no fixed point. An example is given following the proof of Theorem 3.1 below.

There are also some significant differences in emphasis between our two approaches. We avoid modal logic or any modal concepts entirely and work only within a first order framework. For us defaults function as meta-rules whose role it is to further complete an underlying incomplete first order theory whereas for

McDermott and Doyle defaults are formulae of a modal logic. In addition they introduce the notion of the theorems of a non-monotonic theory as the intersection of all of the fixed points of that theory and they proceed to give a proof theory for the sentential case. While we could analogously define the theorems of a default theory as the intersection of its extensions, we choose not to pursue this point of view. Instead our position is that the purpose of default reasoning is to determine *one* consistent set of beliefs about a world, i.e. *one* extension, and to reason within this extension until such time as the evidence at hand forces a revision of those beliefs, in which case a switch to a new extension may be called for. This seems also to be the point of view of the Truth Maintenance System of Doyle (1978). Accordingly, the proof theory developed in the next few sections, is not for theoremhood but for 'believability'. By this we mean a proof theory which, given a wff β , determines whether or not there is an extension containing β , rather than determining whether β is in all extensions.

In general the relationship between default and non-monotonic logics appears to be complex. A few results relating the two will be contained in a forthcoming paper (Reiter (1980)).

In an attempt to formalize the ideas in Sandewall (1972), Kramosil postulated a logic which admits deduction rules with 'negative premises' (Kramosil (1975)). In their intended interpretation these deduction rules correspond in a straightforward way to our closed defaults. Unfortunately, Kramosil's notion of theoremhood for his logic fails to capture a suitable intuitive interpretation of a default rule with the result that he is led to the conclusion that either the default rules are irrelevant, or the resulting theory is meaningless.

3. Normal Default Theories

The fact that some default theories have no extension is not very encouraging for a general theory of default reasoning. In view of this, it is natural to seek out restricted default theories for which extensions can be proved to exist. Fortunately, a very large and natural class of such default theories immediately presents itself. To emphasize the common pattern which defines this class we review the examples of Section 1.

The Closed World Default:

$$\frac{M \neg R(x_1, \dots, x_n)}{\neg R(x_1, \dots, x_n)}$$

for each n -ary relation R of the data base.

Exceptions:

$$\frac{\text{BIRD}(x) : \text{MFLY}(x)}{\text{FLY}(x)}$$

The Frame Default Schema:

$$\frac{R(\mathbf{x},s) : MR(\mathbf{x},f(\mathbf{x},s))}{R(\mathbf{x},f(\mathbf{x},s))}$$

for each relation R and state transition function f .

KRL's Palo Alto Default:

$$\frac{\text{PERSON}(x) : \text{Mhometown}(x) = \text{Palo Alto}}{\text{hometown}(x) = \text{Palo Alto}}$$

The Hometown Example:

$$\frac{\text{SPOUSE}(x,y) \wedge \text{hometown}(y) = z : \text{Mhometown}(x) = z}{\text{hometown}(x) = z}$$

$$\frac{\text{EMPLOYER}(x,y) \wedge \text{location}(y,z) : \text{Mhometown}(x) = z}{\text{hometown}(x) = z}$$

All of these examples share a common pattern; they all have the form $\alpha(\mathbf{x}) : Mw(\mathbf{x})/w(\mathbf{x})$. In fact I know of no naturally occurring default which cannot be represented in this form. Accordingly, we propose to focus attention on such defaults. A *normal* default is one which has the above form for wffs $\alpha(\mathbf{x})$, $w(\mathbf{x}) \in L$. Notice that, although in each of the examples the wff $w(\mathbf{x})$ is a literal, we are adopting a more general point of view in which $w(\mathbf{x})$ is any wff, possibly involving quantifiers, whose free variables are among \mathbf{x} . Similarly, $\alpha(\mathbf{x})$ is arbitrary. A default theory (D,W) is *normal* iff every default of D is normal.

The rest of this section deals with properties of closed normal default theories, i.e. theories all of whose defaults have the form $\alpha : Mw/w$ for closed wffs α and w . Although this seems not to be a very interesting class, (None of the examples of Section 1 is closed.) the results which are derived for it will all generalize in Section 7.

3.1. Closed normal default theories

In this section we derive a variety of results about closed normal default theories which will provide some insight into their structure. The most important such result is the following:

Theorem 3.1. *Every closed normal default theory has an extension.*

Proof. Let $\Delta = (D,W)$ be a closed normal default theory. If W is inconsistent then by Corollary 2.2, Δ has an (inconsistent) extension. Hence assume W is consistent. We construct an extension for Δ as follows: Let $E_0 = W$. For $i \geq 0$ let T_i be a maximal set of closed wffs such that

- (1) $E_i \cup T_i$ is consistent, and
- (2) If $u \in T_i$ then for some $(\alpha : Mw/w) \in D$, u is w , where $\alpha \in E_i$.

Define $E_{i+1} = \text{Th}_L(E_i) \cup T_i$ and $E = \bigcup_{i=0}^{\infty} E_i$. We prove E is an extension for Δ by proving that

$$T_i = \left\{ w \left| \frac{\alpha: Mw}{w} \in D, \text{ where } \alpha \in E_i \text{ and } \neg w \notin E \right. \right\}$$

and invoking Theorem 2.1.

Clearly $T_i \subseteq \text{RHS}$. Assume $T_i \neq \text{RHS}$ so there is a wff $u \in \text{RHS} - T_i$. By the maximality of T_i we must have that $E_i \cup T_i \cup \{u\}$ is inconsistent, i.e. $\text{Th}_L(E_i) \cup T_i \cup \{u\}$ is inconsistent, i.e. $E_{i+1} \cup \{u\}$ is inconsistent, and, since $E_{i+1} \subseteq E$, $E \cup \{u\}$ is inconsistent. Since clearly $\text{Th}_L(E) = E$ we must have $\neg u \in E$ which contradicts the fact that $u \in \text{RHS}$.

For the benefit of those familiar with the non-monotonic logic of McDermott and Doyle (1978) it is worth pointing out that Theorem 3.1 does not hold for their logic in the sense that there are 'normal non-monotonic theories' with no fixed point. The simplest counterexample I know of is the following:

$$\left. \begin{array}{ccc} \frac{\alpha_1: Mw_1}{w_1} & \frac{\alpha_2: Mw_2}{w_2} & \frac{\alpha_3: Mw_3}{w_3} \end{array} \right\} D$$

$$\left. \begin{array}{ccc} w_1 \supset \alpha_2 & w_2 \supset \alpha_3 & w_3 \supset \alpha_1 \\ w_1 \supset \neg w_2 & w_2 \supset \neg w_3 & w_3 \supset \neg w_1 \end{array} \right\} W$$

This closed normal default theory has the unique extension $\text{Th}(W)$. The corresponding non-monotonic logic would represent the defaults by

$$\alpha_1 \wedge Mw_1 \supset w_1 \quad \alpha_2 \wedge Mw_2 \supset w_2 \quad \alpha_3 \wedge Mw_3 \supset w_3$$

and these together with W define a non-monotonic theory with no fixed point.

Theorem 3.2. (Semi-monotonicity). *Suppose D and D' are sets of closed normal defaults with $D' \subseteq D$. Let E' be an extension for the closed normal default theory $\Delta' = (D', W)$ and let $\Delta = (D, W)$. Then Δ has an extension E such that*

- (1) $E' \subseteq E$ and
- (2) $\text{GD}(E', \Delta') \subseteq \text{GD}(E, \Delta)$.

Proof. If W is inconsistent then by Corollaries 2.2 and 2.3 E' is inconsistent and we can take $E = E'$. In this case the theorem trivially holds. Hence, assume W is consistent, whence E' is consistent.

By Corollary 2.2, (D, E') has a consistent extension E . Trivially, $E' \subseteq E$ so if we prove that E is an extension for Δ then the first half of the theorem will be established.

Towards that end, define $F_0 = W$ and for $i \geq 0$

$$F_{i+1} = \text{Th}_L(F_i) \cup \left\{ w \left| \frac{\alpha: Mw}{w} \in D, \text{ where } \alpha \in F_i \text{ and } \neg w \notin E \right. \right\}$$

We shall prove that $E = \bigcup_{i=0}^{\infty} F_i$ in which case by Theorem 2.1 E will be an extension for Δ . First, however, we prove that $E' \subseteq \bigcup_{i=0}^{\infty} F_i$. Since E' is an extension for (D', W) , then by Theorem 2.1, $E' = \bigcup_{i=0}^{\infty} E'_i$ where $E'_0 = W$ and for $i \geq 0$

$$E'_{i+1} = \text{Th}_L(E'_i) \cup \left\{ w \mid \frac{\alpha: Mw}{w} \in D', \text{ where } \alpha \in E'_i \text{ and } \neg w \notin E' \right\}.$$

We inductively prove that $E'_i \subseteq F_i$ for all $i \geq 0$. Trivially, $E'_0 \subseteq F_0$. Assume $E'_i \subseteq F_i$ and consider $w \in E'_{i+1}$. If $w \in \text{Th}_L(E'_i)$ then, since $E'_i \subseteq F_i$, $w \in \text{Th}_L(F_i) \subseteq F_{i+1}$. Otherwise, for some $(\alpha: Mw/w) \in D'$, $\alpha \in E'_i$ and $\neg w \notin E'$. Since $D' \subseteq D$ and $E'_i \subseteq F_i$ then $\alpha \in F_i$ and $(\alpha: Mw/w) \in D$. We prove $\neg w \in E$ whence $w \in F_{i+1}$. For if $\neg w \in E$, then since $w \in E'_{i+1} \subseteq E' \subseteq E$, E would be inconsistent, a contradiction.

Hence $E' \subseteq \bigcup_{i=0}^{\infty} F_i$. We now prove $E = \bigcup_{i=0}^{\infty} F_i$. Since E is an extension for (D, E') then by Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$, where $E_0 = E'$ and for $i \geq 0$

$$E_{i+1} = \text{Th}_L(E_i) \cup \left\{ w \mid \frac{\alpha: Mw}{w} \in D, \text{ where } \alpha \in E_i \text{ and } \neg w \notin E \right\}.$$

We first inductively prove that for all $i \geq 0$ $E_i \subseteq \bigcup_{i=0}^{\infty} F_i$, whence $E = \bigcup_{i=0}^{\infty} E_i \subseteq \bigcup_{i=0}^{\infty} F_i$. We have already proved that $E_0 = E' \subseteq \bigcup_{i=0}^{\infty} F_i$. Assume $E_j \subseteq \bigcup_{i=0}^{\infty} F_i$ and consider $w \in E_{j+1}$. If $w \in \text{Th}_L(E_j)$ then since $\bigcup_{i=0}^{\infty} F_i$ is closed under Th_L , $w \in \bigcup_{i=0}^{\infty} F_i$. Otherwise, for some $(\alpha: Mw/w) \in D$, $\alpha \in E_j$ and $\neg w \notin E$. Hence $\alpha \in \bigcup_{i=0}^{\infty} F_i$ so that $\alpha \in F_i$ for some i . Since $\neg w \notin E$, $w \in F_{i+1} \subseteq \bigcup_{i=0}^{\infty} F_i$.

Hence $E \subseteq \bigcup_{i=0}^{\infty} F_i$. The opposite inclusion, $\bigcup_{i=0}^{\infty} F_i \subseteq E$ has a simple inductive proof based upon $F_i \subseteq E_i$ for all $i \geq 0$.

There remains to prove $\text{GD}(E', \Delta') \subseteq \text{GD}(E, \Delta)$. To that end, suppose $(\alpha: Mw/w) \in \text{GD}(E', \Delta')$. Then $(\alpha: Mw/w) \in D'$, $\alpha \in E'$ and $\neg w \notin E'$. Since $D' \subseteq D$ and $E' \subseteq E$, $(\alpha: Mw/w) \in D$ and $\alpha \in E$. It remains only to prove that $\neg w \notin E$. Now by Theorem 2.5,

$$E' = \text{Th}_L(W \cup \text{CONSEQUENTS}(\text{GD}(E', \Delta'))).$$

Hence $w \in E'$. Since $E' \subseteq E$ and since E is consistent, $\neg w \notin E$.

Theorem 3.2 is a surprising and fortuitous result, one which of course is not true for arbitrary default theories. As we shall see in Section 4, it makes possible a proof theory which is local with respect to the defaults entering into the proof. To see what is meant by this, consider the usual first order proof theory. If A is some set of first order axioms, and if we determine a proof of some wff T using only a subset A' of A then we know that T is a theorem of the first order theory A . This of course is nothing else but a restatement of the monotonicity property of first order logic. It is precisely this property which admits a local proof theory for first order logic, local in the sense that one need not necessarily compute on all of the axioms in A to determine a proof of T .

Now default theories (D, W) in general are non-monotonic in both D and W , i.e. by augmenting either D or W we can so change the extensions of (D, W) that

none of these is a subset of any extension of the augmented theory. This means that any proof procedure for general default theories must necessarily compute on all of the defaults of D and all of the wffs of W —no locality is possible. Indeed, the sentential non-monotonic proof theory of McDermott and Doyle (1978) has precisely this global characteristic; every axiom must be taken into account before an attempted proof can succeed.

What Theorem 3.2 tells us is that for closed normal default theories there is a proof procedure which is local with respect to the defaults so that proofs can be constructed which ignore some of the defaults. Indeed, the proof procedure of Section 4 and its refinement in Section 5 have just this property. In view of the large number of defaults which are likely to be in force in the representation of any interesting domain of knowledge this is a very real computational advantage. Of course this kind of locality cannot be expected with respect to W for otherwise closed normal default theories would be monotonic. And in fact the proof theory of Section 4 requires a satisfiability test involving all of W , which is precisely the kind of global property on W one would expect.

The following theorem tells us that one can attempt to simultaneously hold beliefs in two distinct extensions only at the risk of inconsistency.

Theorem 3.3 (Orthogonality of Extensions). *If a closed normal default theory (D, W) has distinct extensions E and F then $E \cup F$ is inconsistent.*

Proof. By Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$ and $F = \bigcup_{i=0}^{\infty} F_i$ where $E_0 = W$ and for $i \geq 0$

$$E_{i+1} = \text{Th}_L(E_i) \cup \left\{ w \mid \frac{\alpha: M_w}{w} \in D, \text{ where } \alpha \in E_i \text{ and } \neg w \notin E \right\}$$

Similarly for the F_i .

Since E and F are distinct, there is a least integer i such that $E_{i+1} \neq F_{i+1}$ in which case $E_i = F_i$. Hence for some $(\alpha: M_w/w) \in D$ it is the case that $\alpha \in E_i = F_i$, $\neg w \notin E$ and $w \in E_{i+1}$ but $w \notin F_{i+1}$. But if $\alpha \in F_i$ and $w \notin F_{i+1}$ then $\neg w \in F$. Hence $w \in E$ and $\neg w \in F$ so $E \cup F$ is inconsistent.

Corollary 3.4. *Suppose $\Delta = (D, W)$ is a closed normal default theory such that $W \cup \text{CONSEQUENTS}(D)$ is consistent. Then Δ has a unique extension.*

Proof. Suppose on the contrary that Δ has two distinct extensions E_1 and E_2 . By Theorem 2.5,

$$E_i = \text{Th}_L(W \cup \text{CONSEQUENTS}(\text{GD}(E_i, \Delta))) \quad \text{for } i = 1, 2.$$

By Theorem 3.3, $E_1 \cup E_2$ is inconsistent. Now

$$E_i \subseteq \text{Th}_L(W \cup \text{CONSEQUENTS}(D)) \quad \text{for } i = 1, 2$$

since $\text{GD}(E_i, \Delta) \subseteq D$. Hence

$$E_1 \cup E_2 \subseteq \text{Th}_L(W \cup \text{CONSEQUENTS}(D))$$

But $W \cup \text{CONSEQUENTS}(D)$ is consistent, by hypothesis, and this contradicts the inconsistency of $E_1 \cup E_2$.

Theorem 3.5. *Suppose $\Delta = (D, W)$ is a closed normal default theory, and that $D' \subseteq D$. Suppose further that E'_1 and E'_2 are distinct extensions of (D', W) . Then Δ has distinct extensions E_1 and E_2 such that $E'_1 \subseteq E_1$ and $E'_2 \subseteq E_2$.*

Remark. This theorem has the interesting interpretation that the addition of new closed normal defaults to a closed normal default theory (D', W) can never lead to a default theory (D, W) with fewer extensions than the original. *The number of extensions of a closed normal default theory is monotone non-decreasing under the addition of closed normal defaults.*

Proof. By Theorem 3.2 there are extensions E_1 and E_2 for Δ such that $E'_1 \subseteq E_1$ and $E'_2 \subseteq E_2$. Suppose that $E_1 = E_2$. Then $E'_1 \cup E'_2 \subseteq E_1$. But by Theorem 3.3, $E'_1 \cup E'_2$ is inconsistent. Hence E_1 is inconsistent so by Corollary 2.2 W is inconsistent. But this contradicts the fact that (D', W) has distinct extensions E_1 and E_2 , by Corollary 2.3.

4. Proof Theory: Closed Normal Default Theories

In this section we provide a proof theory for closed normal default theories. By a proof theory here we mean a method for answering the question “Given β , can β be believed?”. We interpret this as the following formal task:

Given a closed normal default theory Δ and a closed wff $\beta \in L$, determine whether Δ has an extension E such that $\beta \in E$.

Having devised such a proof theory we shall discover that no procedure exists which, for every Δ and β such that β is a legitimate belief with respect to Δ , will determine a proof of β , i.e. in general beliefs are not recursively enumerable.

4.1. Default proofs

We shall require the following notation: If D is a finite set of closed normal defaults, define $\text{PREREQUISITES}(D) = \bigwedge \alpha$ where the conjunction is taken over all wffs α such that $(\alpha : M_w/w) \in D$. Thus $\text{PREREQUISITES}(D)$ is the conjunction of all of the prerequisites of the defaults of D .

Definition 3. Let $\Delta = (D, W)$ be a closed normal default theory, and $\beta \in L$ a closed wff. A finite sequence D_0, \dots, D_k of finite subsets of D is a *default proof of β* with respect to Δ iff

- P1. $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$
- P2. For $1 \leq i \leq k$
 $W \cup \text{CONSEQUENTS}(D_i) \vdash \text{PREREQUISITES}(D_{i-1})$.
- P3. $D_k = \emptyset$
- P4. $W \cup \bigcup_{i=0}^k \text{CONSEQUENTS}(D_i)$ is satisfiable.

Example 4.1. $\Delta = (D, W)$ has defaults

$$\begin{aligned} \delta_1 &= \frac{E \vee F : \mathbf{M}(A \wedge F)}{A \wedge F} & \delta_2 &= \frac{A : \mathbf{M}B}{B} \\ \delta_3 &= \frac{A \wedge E : \mathbf{M}C}{C} & \delta_4 &= \frac{: \mathbf{M} \neg E}{\neg E} \\ W &= \{C \supset D, A \wedge B \supset E, E \vee D, D \supset F\} \end{aligned}$$

Then $\{\delta_3\}$, $\{\delta_1, \delta_2\}$, $\{\delta_1\}$, $\{\}$ is a default proof of D with respect to Δ . So is $\{\delta_4\}$, $\{\}$. Notice that $\{\delta_3, \delta_4\}$, $\{\delta_1, \delta_2\}$, $\{\}$ is not a default proof of $D \wedge \neg E$, for, while it satisfies conditions P1–P3, it fails to satisfy P4.

The notion of a default proof does not actually provide a proof procedure in the conventional sense of that term. For it gives no method for determining the D_i , nor does it specify a first order proof procedure for verifying P1 or P2. Nor for that matter does it specify how to verify condition P4. In fact most readers will have already balked at P4 given that the first order satisfiable wffs are not recursively enumerable. The best way to view this definition is as a bare bones set of conditions on a sequence of default sets D_0, \dots, D_k . As we shall see, whenever these conditions are satisfied the wff β can be believed. Conversely, as we shall also see, whenever β can be believed then some sequence of defaults will satisfy these conditions. Once this completeness result has been established, we will proceed to fill in some of the obvious computational gaps in this computationally spare notion of a default proof.

Accordingly, we now proceed to prove the completeness of default proofs for closed normal default theories.

Lemma 4.1. *Suppose (D, W) is a closed normal default theory with $(\beta : \mathbf{M}v/v) \in D$. If $W \vdash \beta$ and $W \cup \{v\}$ is consistent, then any extension for the default theory $(D, W \cup \{v\})$ is also an extension for (D, W) .*

Proof. Suppose $(D, W \cup \{v\})$ has an extension E . By Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$, where $E_0 = W \cup \{v\}$ and for $i \geq 0$

$$E_{i+1} = \text{Th}_L(E_i) \cup \left\{ w \mid \frac{\alpha : \mathbf{M}w}{w} \in D, \text{ where } \alpha \in E_i \text{ and } \neg w \notin E \right\}$$

Define $F_0 = W$ and for $i \geq 0$

$$F_{i+1} = \text{Th}_L(F_i) \cup \left\{ w \mid \frac{\alpha : \mathbf{M}w}{w} \in D, \text{ where } \alpha \in F_i \text{ and } \neg w \notin E \right\}.$$

We prove $E = \bigcup_{i=0}^{\infty} F_i$ whence by Theorem 2.1 E is an extension for (D, W) .

Clearly, $F_i \subseteq E_i$ for all $i \geq 0$, so that $\bigcup_{i=0}^{\infty} F_i \subseteq \bigcup_{i=0}^{\infty} E_i = E$. We inductively prove that for all i , $E_i \subseteq F_{i+2}$ whence $\bigcup_{i=0}^{\infty} E_i \subseteq \bigcup_{i=0}^{\infty} F_i$.

To prove $E_0 \subseteq F_2$ it is obviously sufficient to prove $v \in F_2$. Since $W \vdash \beta$, $\beta \in \text{Th}_L(F_0) \subseteq F_1$. Since $W \cup \{v\}$ is consistent, and since E is an extension for $(D, W \cup \{v\})$, then by Corollary 2.2, E is consistent. Since $v \in E$, $\neg v \notin E$. Hence $\beta \in F_1$, $\neg v \notin E$ so $v \in F_2$.

For the inductive step, assume $E_i \subseteq F_{i+2}$ and consider $w \in E_{i+1}$. If $w \in \text{Th}_L(E_i)$ then $w \in \text{Th}_L(F_{i+2}) \subseteq F_{i+3}$. Otherwise for some $(\alpha : Mw/w) \in D$, $\alpha \in E_i$ and $\neg w \notin E$ so that $\alpha \in F_{i+2}$ and $\neg w \notin E$ whence $w \in F_{i+3}$.

Corollary 4.2. *Suppose (D, W) is a closed normal default theory, and that $D' \subseteq D$ such that $W \cup \text{CONSEQUENTS}(D')$ is consistent and such that $W \vdash \text{PREREQUISITES}(D')$. Then any extension for the default theory $(D, W \cup \text{CONSEQUENTS}(D'))$ is also an extension for (D, W) .*

The next theorem provides one half of the completeness result we seek for default proofs.

Theorem 4.3. *Let $\Delta = (D, W)$ be a closed normal default theory, and let $\beta \in L$ be a closed wff. If β has a default proof D_0, \dots, D_k with respect to Δ , then Δ has an extension E such that $\beta \in E$.*

Proof. Consider the closed normal default theory $\Delta' = (\bigcup_{i=0}^k D_i, W)$. We prove that Δ' has an extension E' such that $\beta \in E'$. Since $\bigcup_{i=0}^k D_i \subseteq D$, then by semi-monotonicity (Theorem 3.2) we will be assured that Δ has an extension E such that $E' \subseteq E$, from which the theorem will then follow.

Now by property P2 of default proofs $W \cup \text{CONSEQUENTS}(D_k) \vdash \text{PREREQUISITES}(D_{k-1})$ and since $D_k = \emptyset$, $W \vdash \text{PREREQUISITES}(D_{k-1})$. Also, from property P4 of default proofs it follows that $W \cup \text{CONSEQUENTS}(D_{k-1})$ is consistent. Hence by Corollary 4.2, any extension for $\Delta'_1 = (\bigcup_{i=0}^k D_i, W \cup \text{CONSEQUENTS}(D_{k-1}))$ is also an extension for Δ' , and since Δ'_1 is a closed normal default theory it has such an extension, by Theorem 3.1.

Again by property P2 of default proofs $W \cup \text{CONSEQUENTS}(D_{k-1}) \vdash \text{PREREQUISITES}(D_{k-2})$ and by property P4 $W \cup \text{CONSEQUENTS}(D_{k-1}) \cup \text{CONSEQUENTS}(D_{k-2})$ is consistent so again by Corollary 4.2:

$$\Delta'_2 = \left(\bigcup_{i=0}^k D_i, W \cup \text{CONSEQUENTS}(D_{k-1}) \cup \text{CONSEQUENTS}(D_{k-2}) \right)$$

has an extension which is also an extension for Δ'_1 and hence which is also an extension for Δ' .

We can continue in this way, finally arriving at a default theory

$$\Delta'_k = \left(\bigcup_{i=0}^k D_i, W \cup \bigcup_{i=1}^k \text{CONSEQUENTS}(D_{k-i}) \right),$$

which has an extension E' which is also an extension for Δ' . Since clearly $W \cup \text{CONSEQUENTS}(D_0) \subseteq E'$ and since $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$, then $\beta \in E'$.

Suppose that β has a default proof $P_\beta = D_0, \dots, D_k$. Then the *default support* of this proof is defined to be $DS(P_\beta) = \bigcup_{i=0}^k D_i$. The default support of a proof is simply the set of all default rules which are invoked in that proof. Notice that condition P4 in the definition of a default proof P_β is the same as the requirement that $W \cup \text{CONSEQUENTS}(DS(P_\beta))$ be satisfiable.

Lemma 4.4. *Suppose that $\Delta = (D, W)$ is a closed normal default theory, and that $\beta', \beta'' \in L$ are closed wffs. Then $\beta' \wedge \beta''$ has a default proof with respect to Δ , iff β' and β'' have default proofs $P_{\beta'}$ and $P_{\beta''}$ respectively with respect to Δ , such that $W \cup \text{CONSEQUENTS}(DS(P_{\beta'}) \cup DS(P_{\beta''}))$ is satisfiable.*

Proof (\Rightarrow). Suppose P is a default proof of $\beta' \wedge \beta''$ with respect to Δ . Then trivially P is also a default proof of β' and of β'' and the result immediately follows.

(\Leftarrow). Assume $P_{\beta'} = D'_0, \dots, D'_m$ and $P_{\beta''} = D''_0, \dots, D''_n$ and suppose $n \geq m$. Then clearly

$$P = D'_0 \cup D''_0, \dots, D'_m \cup D''_m, D''_{m+1}, \dots, D''_n$$

is a default proof of $\beta' \wedge \beta''$ with respect to Δ .

Corollary 4.5. *Suppose that $\Delta = (D, W)$ is a closed normal default theory, and that $\theta_1, \dots, \theta_r$ have default proofs $P_{\theta_1}, \dots, P_{\theta_r}$ respectively with respect to Δ . Suppose further that $W \cup \text{CONSEQUENTS}(\bigcup_{i=1}^r DS(P_{\theta_i}))$ is satisfiable. Then $\theta_1 \wedge \dots \wedge \theta_r$ has a default proof P with respect to Δ , such that $DS(P) = \bigcup_{i=1}^r DS(P_{\theta_i})$.*

Proof. Notice that, in the proof of the \Leftarrow half of Lemma 4.4, the default proof P of $\beta' \wedge \beta''$ has the property that $DS(P) = DS(P_{\beta'}) \cup DS(P_{\beta''})$. The corollary easily follows from this.

Lemma 4.6. *Suppose α has a default proof P_α with respect to a closed normal default theory $\Delta = (D, W)$. Suppose further that the wff $\alpha \supset \beta$ is valid. Then P_α is also a default proof of β with respect to Δ .*

Proof. Suppose $P_\alpha = D_0, \dots, D_k$. Then $W \cup \text{CONSEQUENTS}(D_0) \vdash \alpha$. Since $\alpha \supset \beta$ is valid, $\vdash \alpha \supset \beta$ so trivially $W \cup \text{CONSEQUENTS}(D_0) \vdash \alpha \supset \beta$. Hence by modus ponens, $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$ so that P_α is also a default proof of β .

The next theorem provides the remaining half of the completeness result for default proofs.

Theorem 4.7. *Suppose that E is an extension for a consistent closed normal default theory $\Delta = (D, W)$, and that $\beta \in E$. Then β has a default proof with respect to Δ .*

Proof. By Theorem 2.1, $E = \bigcup_{i=0}^{\infty} E_i$, where $E_0 = W$ and for $i \geq 0$.

$$E_{i+1} = \text{Th}_L(E_i) \cup \left\{ w \mid \frac{\alpha : M_w}{w} \in D, \text{ where } \alpha \in E_i \text{ and } \neg w \notin E \right\}.$$

Since $\beta \in E$, $\beta \in E_i$ for some i . We shall inductively prove the following result, for all $i \geq 0$:

If $\theta \in E_i$, then there is a default proof P_θ of θ with respect to Δ such that $DS(P_\theta) \subseteq GD(E, \Delta)$.

The case $i = 0$ is trivial; if $\theta \in E_0$ then since W is consistent (Corollary 2.2), $P_\theta = \emptyset$ is a definite proof of θ with $DS(P_\theta) = \emptyset$.

Now assume the induction hypothesis for all wffs of E_i and consider $\theta \in E_{i+1}$.

Case 1. $\theta \in Th_L(E_i)$

Then there exist finitely many wffs $\theta_1, \dots, \theta_r \in E_i$ such that $\{\theta_1, \dots, \theta_r\} \vdash \theta$. By the induction hypothesis, each θ_j has a default proof P_{θ_j} such that $DS(P_{\theta_j}) \subseteq GD(E, \Delta)$. Hence $\bigcup_{j=1}^r DS(P_{\theta_j}) \subseteq GD(E, \Delta)$. By Theorem 2.5

$E = Th_L(W \cup CONSEQUENTS(GD(E, \Delta)))$

and since E is consistent so is $W \cup CONSEQUENTS(GD(E, \Delta))$. Hence, so is

$$W \cup CONSEQUENTS\left(\bigcup_{j=1}^r DS(P_{\theta_j})\right)$$

so by Corollary 4.5, $\theta_1 \wedge \dots \wedge \theta_r$ has a default proof P such that

$$DS(P) = \bigcup_{j=1}^r DS(P_{\theta_j}) \subseteq GD(E, \Delta). \quad (4.1)$$

Since $\{\theta_1, \dots, \theta_r\} \vdash \theta$, $\theta_1 \wedge \dots \wedge \theta_r \supset \theta$ is valid. Hence by Lemma 4.6 P is also a default proof of θ , and by (4.1) $DS(P) \subseteq GD(E, \Delta)$.

Case 2. θ is w where $(\alpha : Mw/w) \in D$, $\alpha \in E_i$ and $\neg w \notin E$.

By the induction hypothesis, α has a default proof $P_\alpha = D_0, \dots, D_k$ and $DS(P_\alpha) \subseteq GD(E, \Delta)$. Consider the sequence $P_w = \{\alpha : Mw/w\}, D_0, \dots, D_k$. This clearly satisfies conditions P1–P3 of the definition of a default proof of w . Now $(\alpha : Mw/w) \in GD(E, \Delta)$, and since $DS(P_\alpha) \subseteq GD(E, \Delta)$, then $DS(P_w) \subseteq GD(E, \Delta)$. Moreover, by the same argument as in Case 1, we have that $W \cup CONSEQUENTS(DS(P_w))$ is consistent. Hence P_w satisfies property P4 so that P_w is a default proof of w .

By combining Theorems 4.3 and 4.7 we obtain the following completeness result:

Theorem 4.8. *Let $\beta \in L$ be a closed wff. A consistent closed normal default theory Δ has an extension E such that $\beta \in E$ iff β has a default proof with respect to Δ .*

While this completeness result is gratifying we still have no computationally reasonable way of deriving default proofs. We shall defer this question to Section 5. In the interim we show that in a very real sense, default proofs cannot always be derived, even when they exist.

4.2. The extension membership problem

It is known that there is no decision procedure for the validity problem of first order logic (Church (1936)). What this amounts to is that there is no algorithm which, when presented with a closed first order wff, will determine whether or not that wff is valid. The validity problem is, however, semidecidable, i.e. there is a procedure (namely first order proof theory) which will confirm the validity of a closed wff (by finding a proof for it) if that wff is valid, but which fails to terminate on some non-valid wffs. Another way of expressing this state of affairs is by saying that the closed valid wffs are recursively enumerable, but that there is no recursive enumeration of the closed non-valid wffs. From this it follows that the set of closed satisfiable wffs is not recursively enumerable.

Now the definition of a default proof appeals, in condition P4, to a test for the satisfiability of some set of first order wffs. In view of the fact that the closed satisfiable first order wffs are not recursively enumerable, this suggests that there may not be a recursive enumeration of the set of beliefs of an arbitrary closed normal default theory, i.e. that the extension membership problem is not even semi-decidable. By the *extension membership problem* we mean the following:

Given a closed default theory Δ and a closed wff $\beta \in L$, is there an extension for Δ which contains β ?

We now confirm this intuition that the extension membership problem for closed normal default theories is not semi-decidable, or what amounts to the same thing, that the union of the extensions of an arbitrary closed normal default theory is not recursively enumerable.

Given a default theory Δ , let $B(\Delta) = \cup E$ where the union is taken over all extensions E of Δ . Let w_1, w_2, \dots , be a recursive enumeration of the closed wffs of L . For $i = 1, 2, \dots$, let Δ_i be the closed normal default theory $(\{ : M w_i / w_i \}, \emptyset)$ so that Δ_i has the single default rule $: M w_i / w_i$, and no first order axioms. Then clearly Δ_i has a unique extension E_i . E_i and therefore $B(\Delta_i)$ is empty if w_i is unsatisfiable, and is $\text{Th}_L(\{w_i\})$ if w_i is satisfiable. Hence $\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} B(\Delta_i)$ is the set of closed satisfiable wffs of L .

We shall argue that there is no recursive enumeration of $B(\Delta)$ for arbitrary closed normal default theories Δ . For if there were, then there is a recursive function $f(\Delta, i)$, defined on closed normal default theories Δ and nonnegative integers i , such that $\text{range}(f(\Delta, \cdot)) = B(\Delta)$. In particular, $\text{range}(f(\Delta_j, \cdot)) = B(\Delta_j)$. Then it is known (see e.g. Rogers 1967) that there is a recursive function $g(i)$ defined on the nonnegative integers i such that $\text{range}(g) = \bigcup_{j=1}^{\infty} B(\Delta_j)$, i.e. $\bigcup_{j=1}^{\infty} B(\Delta_j)$, the set of closed satisfiable wffs of L , is recursively enumerable, a contradiction.

Theorem 4.9. *The extension membership problem for closed normal default theories is not semi-decidable.*

This result assures us that the appeal to satisfiability in condition P4 of the definition of a default proof is not merely an accidental feature of our particular brand of proof theory. Rather, any proof theory whatever for closed normal default theories must somehow appeal to some inherently non-semi-decidable process. This extremely pessimistic result forces the conclusion that any computational treatment of defaults must necessarily have an heuristic component and will, on occasion, lead to mistaken beliefs. Given the faulty nature of human common sense reasoning, this is perhaps the best one could hope for in any event.

Of course there are decidable subcases of the extension membership problem. It is not difficult to see from the definition of a default proof that provided

$$W \cup \left\{ \alpha \mid \frac{\alpha: M_w}{w} \in D \right\} \cup \left\{ w \mid \frac{\alpha: M_w}{w} \in D \right\}$$

belongs to a decision class for first order provability then the extension membership problem will be decidable. In particular then, this problem is decidable for the sentential case, i.e. when each α and w and each formula of W is a sentential wff. It is also decidable in the monadic case, i.e. when each of these is a formula of the monadic predicate calculus.

It is also of some interest to contrast Theorem 4.9 with the corresponding result for first order theories. It is well known that the set of theorems of any first order theory is recursively enumerable. If the elements of the extensions of a default theory are viewed as the 'theorems' of that theory, then Theorem 4.9 tells us that the 'theorems' of an arbitrary default theory are not recursively enumerable. This then provides another major distinction between default logic and first order logic, in addition to that of the non-monotonicity of default logic.

5. Top Down Default Proofs: Interface with a Resolution Theorem Prover

The definition of a default proof naturally admits both top down and bottom up search procedures. The top down approach can be loosely described as follows: given β , determine a subset D_0 of the defaults such that $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$. For $i \geq 1$, if D_{i-1} has been determined, then determine a subset D_i of the defaults such that

$$W \cup \text{CONSEQUENTS}(D_i) \vdash \text{PREREQUISITES}(D_{i-1}).$$

If, for some k , $W \vdash \text{PREREQUISITES}(D_{k-1})$ and if $W \cup \bigcup_{i=0}^{k-1} \text{CONSEQUENTS}(D_i)$ is satisfiable, then a proof of β has been found.

The bottom up approach is as follows: given β , determine whether $W \vdash \beta$. If so, and if W is satisfiable return success. Otherwise, determine a subset D_0 of the defaults such that $W \vdash \text{PREREQUISITES}(D_0)$. If $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$ and if $W \cup \text{CONSEQUENTS}(D_0)$ is satisfiable, return success. Otherwise, for $i \geq 1$, assuming that D_{i-1} has been determined, determine a subset D_i of the

defaults such that $W \cup \bigcup_{j=0}^{i-1} \text{CONSEQUENTS}(D_j) \vdash \text{PREREQUISITES}(D_i)$. If $W \cup \bigcup_{j=0}^i \text{CONSEQUENTS}(D_j) \vdash \beta$ and if $W \cup \bigcup_{j=0}^i \text{CONSEQUENTS}(D_j)$ is satisfiable, return success.⁴

The bottom up search procedure does not appear to be computationally very promising. We shall focus instead on the top down approach. At first blush this does not look too promising either. How, for example, are we to determine a D_0 such that $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$? Are we to simply randomly select subsets D_0 of D and try each in turn? The solution lies with the appropriate choice of a first order proof procedure. For if a top down theorem prover is used, then the goal wff β can help to select out a suitable subset D_0 of D .

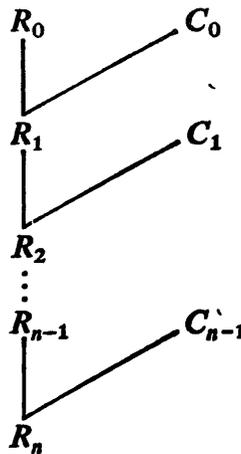


FIG. 1. A linear resolution proof.

Linear resolution (Loveland (1970), Luckham (1970)) provides just such a top down theorem prover. A *linear resolution proof of β* from some set of clauses S has the form of Fig. 1, where

- (1) The *top clause*, R_0 , is a clause of $\neg\beta$.
- (2) For $1 \leq i \leq n$, R_i is a resolvent of R_{i-1} and C_{i-1} .
- (3) For $0 \leq i \leq n-1$, $C_i \in S$ or C_i is a clause of $\neg\beta$ or C_i is R_j for some $j < i$.
- (4) $R_n = \square$, the empty clause.

Linear resolution is known to be complete. It is also a suitable generalization of conventional incomplete back-chaining (subgoaling) proof procedures which reason backward from a goal β until a suitable set of premises is reached.

Now the idea is to use linear resolution to allow the goal wff β to help select out a suitable subset D_0 of D . To do this requires an appropriate representation of a closed normal default theory (D, W) . Specifically, assume with no loss of generality that W is a set of clauses. In addition we shall require a special clausal form of the consequents of each default of D . If $\delta = (\alpha : Mw/w) \in D$, suppose C_1, \dots, C_r

⁴ This bottom up approach is not immediately apparent from the definition of a default proof. It can be shown to follow from that definition. We omit the details.

are all of the clauses of w . Then for $1 \leq i \leq r$ the ordered pair $(C_i, \{\delta\})$ is called a *consequent clause* of the default δ . In effect, a consequent clause of δ will function as an ordinary clause of w indexed by the name of the default giving rise to that clause.

Example 5.1. Consider a closed normal default theory $\Delta = (D, W)$ with defaults

$$\begin{aligned} \delta_1 &= \frac{:MA}{A} & \delta_2 &= \frac{(Ex)(Ey)B(x, y):MC}{C} \\ \delta_3 &= \frac{D \vee A:ME}{E} & \delta_4 &= \frac{C \wedge E:M(Ex)(y)F(x, y) \wedge G(x)}{(Ex)(y)F(x, y) \wedge G(x)}. \end{aligned}$$

Suppose the clausal form of W is

$$C_1 = B(x, f(x)) \quad C_2 = \neg C \vee D \vee A \quad C_3 = \neg C \vee \neg E \vee \neg A.$$

The consequent clauses are:

$$(A, \{\delta_1\}) \quad (C, \{\delta_2\}) \quad (E, \{\delta_3\}) \\ (F(a, y), \{\delta_4\}), (G(a), \{\delta_4\});$$

a is a Skolem constant introduced in forming the clausal form of the consequent of δ_4 .

If $\Delta = (D, W)$ is a closed normal default theory, where W is a set of (ordinary) clauses, define

$$\begin{aligned} \text{CLAUSES}(\Delta) = \\ \{(C, \{\delta\}) \mid \delta \in D \text{ and } (C, \{\delta\}) \text{ is a consequent clause of } \delta\} \\ \cup \{(C, \{\}) \mid C \in W\}. \end{aligned}$$

In general, we shall refer to an ordered pair (C, D) where C is a clause and D a set of defaults as an *indexed clause*; C will be said to be *indexed by* D . Thus the elements of $\text{CLAUSES}(\Delta)$ are all indexed clauses, where each clause is indexed by the default which gave rise to it. In what follows, we shall usually not distinguish between $(C, \{\})$ i.e. a clause C indexed by the empty set of defaults, and C itself.

Now we propose to invoke linear resolution as the first order theorem proving component of a default theorem prover. The representation of Δ to be used by the linear resolution theorem prover will be $\text{CLAUSES}(\Delta)$, i.e. a set of indexed clauses. Hence the notion of a resolvent of two indexed clauses must be defined. If (C_1, D_1) and (C_2, D_2) are indexed clauses, and if R is a usual resolvent of C_1 and C_2 , then $(R, D_1 \cup D_2)$ is a *resolvent* of the indexed clauses (C_1, D_1) and (C_2, D_2) . Then a linear resolution proof of a wff β from some set S of indexed clauses is the same as that using ordinary (non-indexed clauses) except that all of the clauses in the proof are indexed, i.e. it has the form of Fig. 1, where

(1) The top clause, R_0 , is a clause of $\neg\beta$.

(2) For $1 \leq i \leq n$, R_{i-1} and C_{i-1} are indexed clauses and R_i is a resolvent of these indexed clauses.

- (3) For $0 \leq i \leq n-1$, $C_i \in S$ or C_i is a clause of $\neg\beta$ or C_i is R_j for some $j < i$.
 (4) $R_n = (\square, D)$ for some set of defaults D .

We shall say that such a linear resolution proof of β returns D .

Now recall the original problem: Given β and a closed normal default theory $\Delta = (D, W)$ how are we to determine a D_0 such that $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$? Suppose we can determine a linear resolution proof of β from the set of indexed clauses $\text{CLAUSES}(\Delta)$, and that this proof returns D_0 . Then it is immediately apparent that $W \cup \text{CONSEQUENTS}(D_0) \vdash \beta$.

Now having determined D_0 , there will in general be the problem of determining D_1 such that $W \cup \text{CONSEQUENTS}(D_1) \vdash \text{PREREQUISITES}(D_0)$. Again, we can proceed in a top down fashion by determining a linear resolution proof of $\text{PREREQUISITES}(D_0)$ from $\text{CLAUSES}(\Delta)$ and this proof will return the required D_1 . And so on.

These considerations motivate the following definition:

Definition 4. A *top down default proof* of β with respect to a closed normal default theory $\Delta = (D, W)$ is a sequence of linear resolution proofs L_0, \dots, L_k such that

- (1) L_0 is a linear resolution proof of β from $\text{CLAUSES}(\Delta)$.
- (2) For $0 \leq i \leq k$, L_i returns D_i .
- (3) For $1 \leq i \leq k$, L_i is a linear resolution proof of $\text{PREREQUISITES}(D_{i-1})$ from $\text{CLAUSES}(\Delta)$.
- (4) $D_k = \emptyset$.
- (5) $W \cup \bigcup_{i=0}^k \text{CONSEQUENTS}(D_i)$ is satisfiable.

Example 5.2. This is the closed normal default theory of Example 4.1, which has two extensions given by

$$E_1 = \text{Th}(W \cup \{A \wedge F, B, C\})$$

$$E_2 = \text{Th}(W \cup \{A \wedge F, \neg E\}),$$

both of which contain D . Figure 2(a) gives a top down default proof of D as an element of E_2 , and Fig. 2(b) is a proof of D as an element of E_1 .

Example 5.3. This is the default theory of Example 5.1. It has three extensions:

$$E_1 = \text{Th}(W \cup \{A, C\}),$$

$$E_2 = \text{Th}(W \cup \{A, E\}),$$

$$E_3 = \text{Th}(W \cup \{C, E, (\text{Ex})(y)F(x, y) \wedge G(x)\}).$$

Clearly, $(\text{Ex})G(x) \in E_3$. Figure 3(a) represents a top down default proof of $(\text{Ex})G(x)$. This does indeed represent a default proof since δ_2, δ_3 and δ_4 are all of the defaults which enter into the proof and $W \cup \text{CONSEQUENTS}(\{\delta_2, \delta_3, \delta_4\}) = W \cup \{C, E, (\text{Ex})(y)F(x, y) \wedge G(x)\}$ is satisfiable.

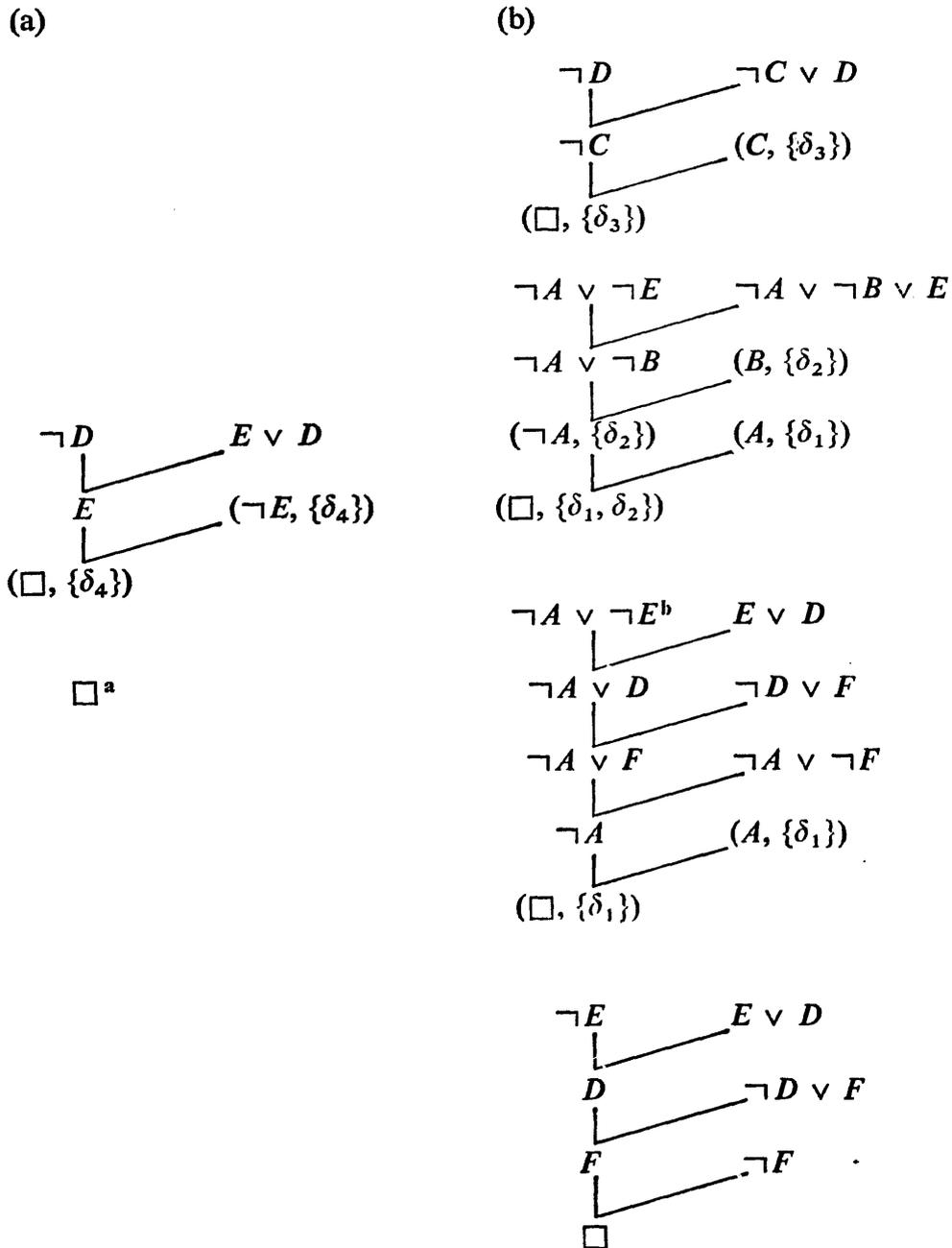


FIG. 2.

a Since the prerequisite of δ_4 is T , the identically true clause, its negation is \square , which completes the proof.

b This is one of the clauses of the negation of PREREQUISITES $(\{\delta_1, \delta_2\}) = A \wedge (E \vee F)$.

c This is the other clause of the negation of PREREQUISITES $(\{\delta_1, \delta_2\})$.

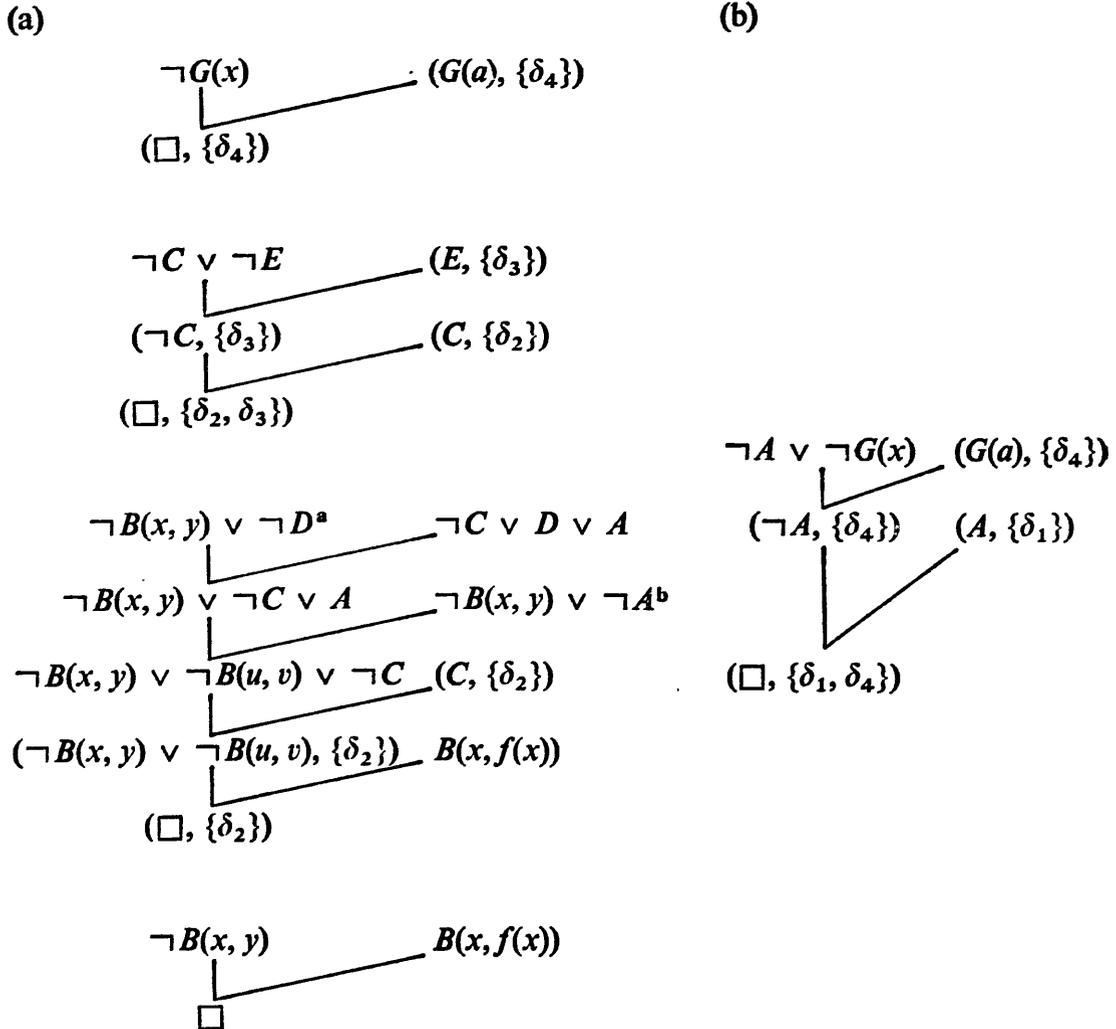


FIG. 3.

^a This is one of the clauses of the negation of $\text{PREREQUISITES}(\{\delta_2, \delta_3\}) = (\exists x)(\exists y)B(x, y) \wedge (D \vee A)$.

^b This is the other clause of the negation of $\text{PREREQUISITES}(\{\delta_2, \delta_3\})$.

Notice that no extension contains $A \wedge (\exists x)G(x)$. To see how an attempted top down default proof of this wff gets blocked, consider Fig. 3(b), which is the initial linear refutation of $\neg A \vee \neg G(x)$. Since $\text{PREREQUISITES}(\{\delta_1, \delta_4\}) = C \wedge E$ we would next construct a linear refutation of $\neg C \vee \neg E$. We have already done this in Fig. 3(a). The refutation there derives $(\square, \{\delta_2, \delta_3\})$ as its terminal node. There is no need to proceed further since $W \cup \text{CONSEQUENTS}(\{\delta_1, \delta_4, \delta_2, \delta_3\})$ is unsatisfiable, so the attempted top down default proof of $A \wedge (\exists x)G(x)$ fails.

It is not difficult to show, using the completeness of linear resolution for first order logic, that β has a default proof iff there is a top down default proof of β .

This result, when coupled with the completeness of default proofs (Theorem 4.8) yields the following:

Theorem 5.1 (Completeness of Top Down Default Proofs). *Let $\Delta = (D, W)$ be a consistent closed normal default theory and let $\beta \in L$ be a closed wff. Then Δ has an extension containing β iff there is a top down default proof of β with respect to Δ .*

This result is of some computational importance. Top down default proofs provide a coherent method for determining the necessary defaults to invoke in establishing a wff β . Moreover, at each stage the current goal (i.e. β or $\text{PRE-REQUISITES}(D_{i-1})$) provides guidance in a top down fashion in the selection of the defaults necessary for establishing that goal. Another important feature of top down default proofs is that they can be implemented in such a way as to favour 'default free' proofs of β , i.e. the linear proof L_0 can seek first to establish β using W alone, only introducing consequent clauses when this fails. Such wffs β , free of default assumptions, have a much firmer status than those which rest upon one or more default assumptions. For example they have a monotonic character in the sense that they need never be retracted under the addition of new first order facts about the world. And they need never take part in the process of belief revision which is required whenever inconsistencies arise. (See Section 6 for a discussion of this issue.) Finally, top down default proofs may invoke any of a variety of complete linear resolution strategies, among which some quite restrictive methods are known e.g. C-ordered linear resolution (Reiter (1971)) and SL resolution (Kowalski and Kuehner (1971)).

It is of some interest to reflect upon the nature of top down default proofs in view of the discussion following the proof of Theorem 3.2. For one can see quite clearly how it is that the top down linear resolution proofs L_i select out a subset of all of the defaults. Any default not selected out in this way is irrelevant to the proof and need not be further considered. This is precisely the property of locality with respect to the defaults referred to in that discussion. We also pointed out there that no such locality principle can be expected with respect to W . This expectation is realized in the final defining property of top down default proofs, namely that $W \cup \bigcup_{i=0}^k \text{CONSEQUENTS}(D_i)$ be satisfiable, for satisfiability is clearly a global property.

6. Closed Normal Default Theories and the Revision of Beliefs

The point of view of this section is that the role of a default reasoning program is to successively derive new beliefs in response to some task at hand. Now an extension specifies one coherent view of an incompletely specified world; moreover many such coherent views are possible, one for each extension of the default theory. We are free to choose any one of these extensions as our current view of the world; the particular choice made will usually be conditioned by some overall objective, e.g. some task to be performed, some decision to be made, etc.

From this perspective, the default reasoner cannot entertain just any set of derived beliefs; they must all belong to some common extension. This requirement gives rise to a variety of problems:

(1) How can one determine when a set of derived beliefs is a subset of some extension?

(2) Suppose given a set B of derived beliefs all of which belong to some common extension of (D, W) . Now imagine updating W with some new facts (observations, perhaps) about the world. How can one determine whether B is a subset of some extension of this new default theory? That is can we continue to entertain our previously held beliefs in the presence of new information about the world?

(3) This is the same as (2), except that D is updated rather than W . Can we continue to entertain our previously held beliefs in the presence of new defaults?

It is the purpose of this section to provide answers to these questions.

6.1. Derived beliefs: maintaining a common extension

We envisage a default reasoning program as functioning in the following way:

Initially, some closed normal default theory $\Delta_0 = (D, W)$ is specified. The default reasoner then determines some default proof P_{β_0} with respect to Δ_0 . Now that β_0 has been derived, it should be made available as a premise for future derivations, i.e. the 'current default theory' becomes $\Delta_1 = (D, W \cup \{\beta_0\})$. If now a default proof P_{β_1} is determined with respect to Δ_1 then both derived beliefs β_0 and β_1 should be available as premises for future derivations so that the 'current default theory' becomes $\Delta_2 = (D, W \cup \{\beta_0, \beta_1\})$. And so on. In effect, we want to view derived beliefs as lemmas to be subsequently used in the derivation of new beliefs. What is required is that the default reasoner, by proceeding in this way, will be generating 'an approximation' to some extension for the original default theory Δ_0 , i.e. the set of derived beliefs $\{\beta_0, \beta_1, \dots\}$ must be a subset of some extension for Δ_0 . The following example demonstrates that this requirement is not always fulfilled:

Example 6.1.

$$\Delta_0 = \left(\left\{ \frac{:MA}{A}, \frac{B:M\neg A}{\neg A} \right\}, \{A \supset B\} \right).$$

Then $\beta_0 = B$ is a derived belief of Δ_0 . Also $\beta_1 = \neg A$ is a derived belief of

$$\Delta_1 = \left(\left\{ \frac{:MA}{A}, \frac{B:M\neg A}{\neg A} \right\}, \{A \supset B, B\} \right),$$

yet β_1 is not in any extension for the original default theory Δ_0 .

Some care must therefore be exercised in using derived beliefs as lemmas. The following theorem provides a sufficient condition for the proper use of derived beliefs as lemmas in the derivation of new beliefs.

Theorem 6.1. *Let $\Delta_0 = (D, W)$ be a closed normal default theory. In general, suppose Δ_i has been determined and that P_{β_i} is a default proof of β_i with respect to Δ_i . Let $\Delta_{i+1} = (D, W \cup \{\beta_0, \dots, \beta_i\})$. For any $n \geq 0$, if*

$$W \cup \bigcup_{i=0}^n \text{CONSEQUENTS}(\text{DS}(P_{\beta_i})) \text{ is consistent} \quad (6.1)$$

then Δ_0 has an extension E_0 such that $\{\beta_0, \dots, \beta_n\} \subseteq E_0$.

Proof. Let

$$\Delta'_0 = \left(\bigcup_{j=0}^n \text{DS}(P_{\beta_j}), W \right)$$

and

$$\Delta'_i = \left(\bigcup_{j=0}^n \text{DS}(P_{\beta_j}), W \cup \{\beta_0, \dots, \beta_{i-1}\} \right) \quad \text{for } 1 \leq i \leq n.$$

We first inductively prove that for $0 \leq i \leq n$

$$W \cup \{\beta_0, \dots, \beta_{i-1}\} \cup \bigcup_{j=0}^n \text{CONSEQUENTS}(\text{DS}(P_{\beta_j})) \text{ is consistent.}$$

The case $i = 0$ is simply the hypothesis (6.1). Assume the result for i and consider

$$W \cup \{\beta_0, \dots, \beta_i\} \cup \bigcup_{j=0}^n \text{CONSEQUENTS}(\text{DS}(P_{\beta_j})) \quad (6.2)$$

Now P_{β_i} is a default proof of β_i with respect to Δ_i . Hence

$$W \cup \{\beta_0, \dots, \beta_{i-1}\} \cup \text{CONSEQUENTS}(\text{DS}(P_{\beta_i})) \vdash \beta_i.$$

Thus, the consistency of (6.2) reduces to that of the induction hypothesis.

This result coupled with Corollary 3.4 assures us that for $0 \leq i \leq n$ Δ'_i has a unique extension E'_i . Since P_{β_i} is a default proof of β_i with respect to Δ_i it is also a default proof of β_i with respect to Δ'_i and since E'_i is the unique extension for Δ'_i , $\beta_i \in E'_i$. Hence $\beta_0 \in E'_0$, so by Theorem 2.6 E'_0 is an extension for Δ'_1 . Since Δ'_1 has a unique extension E'_1 then $E'_0 = E'_1$. Since $\beta'_1 \in E'_1$ we can apply Theorem 2.6 again to conclude that E'_1 is an extension for Δ'_2 and again by uniqueness $E'_1 = E'_2$. We can continue in this way to obtain $E'_0 = E'_1 = \dots = E'_n$ and since, for $0 \leq i \leq n$, $\beta_i \in E'_i$ we have $\{\beta_0, \dots, \beta_n\} \subseteq E'_0$.

Finally, since $\bigcup_{j=0}^n \text{DS}(P_{\beta_j}) \subseteq D$, then by Theorem 3.2, $\Delta_0 = (D, W)$ has an extension E_0 such that $E'_0 \subseteq E_0$.

This last theorem provides a mechanism by which a default reasoning program can be assured that all of its current beliefs have a common extension: Let B be the current set of derived beliefs. For each $\beta \in B$ let P_β be the default proof of β . P_β may invoke other β 's of B as lemmas. Then the members of B all share a common extension if $W \cup \bigcup_{\beta \in B} \text{DS}(P_\beta)$ is consistent.

The question arises "How does the default reasoner proceed in the event its derived beliefs B do not satisfy this consistency property?" This raises some quite

complex issues which we do not address in this paper. However, the basic flavour of the required approach is clear. To restore the consistency property, the system must initiate a process of *belief revision* as follows:

- (1) Attempt to rederive some minimal subset of the derived beliefs B in order that the resulting new set of default proofs has the consistency property.
- (2) Failing this, reject some minimal subset of B .

This process of belief revision is one of the tasks performed by the Truth Maintenance System of Doyle (1978).

6.2. Derived beliefs and the assimilation of new information

Section 1.2 noted the need for belief revision in the event that a belief depends upon some default assumptions which are subsequently violated by new first order facts about a world. Our purpose now is to make this idea precise, and to formulate a condition under which belief revision will not be necessary.

Theorem 6.2. *Let $\Delta_0 = (D, W)$ be a closed normal default theory, and let $\{\beta_0, \dots, \beta_n\}$ be a set of derived beliefs determined as in the statement of Theorem 6.1. Suppose further that $F \subseteq L$ is a set of closed wffs. If*

$$W \cup F \cup \bigcup_{i=0}^n \text{CONSEQUENTS}(\text{DS}(P_{\beta_i})) \text{ is consistent} \quad (6.3)$$

then $(D, W \cup F)$ has an extension E such that $\{\beta_0, \dots, \beta_n\} \subseteq E$.

Proof. Let

$$\Delta'_0 = (D, W \cup F)$$

and

$$\Delta'_i = (D, W \cup F \cup \{\beta_0, \dots, \beta_{i-1}\}) \quad \text{for } 1 \leq i \leq n.$$

Now P_{β_i} is a default proof of β_i with respect to Δ_i . By the consistency property (6.3), $W \cup F \cup \text{CONSEQUENTS}(\text{DS}(P_{\beta_i}))$ is consistent so that P_{β_i} is also a default proof of β_i with respect to Δ'_i . Hence, by Theorem 6.1, $\Delta'_0 = (D, W \cup F)$ has an extension E such that $\{\beta_0, \dots, \beta_n\} \subseteq E$.

Theorem 6.2 provides a mechanism with which a default reasoning program can be assured that all of its current beliefs $\{\beta_0, \dots, \beta_n\}$ with respect to a theory (D, W) have a common extension in the theory $(D, W \cup F)$ obtained by updating W with some new facts F about the world. Again, the problem of belief revision arises whenever the consistency property (6.3) fails to hold.

6.3. Derived beliefs and the assimilation of new defaults

We consider here the problem of updating a closed normal default theory with new defaults, and how this may influence our confidence in some set of previously derived beliefs.

Theorem 6.3. *Let $\Delta_0 = (D, W)$ be a closed normal default theory, and let $\{\beta_0, \dots, \beta_n\}$ be a set of derived beliefs determined as in the statement of Theorem 6.1. Suppose further that D' is a set of closed normal defaults. If (D, W) has an extension E such that $\{\beta_0, \dots, \beta_n\} \subseteq E$ then $(D \cup D', W)$ has an extension E' such that $\{\beta_0, \dots, \beta_n\} \subseteq E'$.*

Proof. The proof is a simple consequence of semi-monotonicity (Theorem 3.2).

Theorem 6.3 provides the only truly comforting result *vis à vis* belief revision. It guarantees that provided the derived beliefs β_0, \dots, β_n have a common extension in (D, W) then they will continue to have a common extension in $(D \cup D', W)$. Updating a closed normal default theory with new defaults cannot affect old beliefs.

7. Arbitrary Default Theories

The technical results of the previous sections have been derived for closed default theories, i.e. theories all of whose defaults have the form $\alpha : M\beta_1, \dots, M\beta_m / w$, where $\alpha, \beta_1, \dots, \beta_m, w$ are all closed first order wffs. In view of the examples of Section 1, none of which is a closed default, the genuinely interesting cases involve open defaults. An *open* default has the form $\alpha(\mathbf{x}) : M\beta_1(\mathbf{x}), \dots, M\beta_m(\mathbf{x}) / w(\mathbf{x})$ where at least one of $\alpha(\mathbf{x}), \beta_1(\mathbf{x}), \dots, \beta_m(\mathbf{x}), w(\mathbf{x})$ contains free variables in \mathbf{x} . The purpose of this section is to generalize our earlier results on closed default theories to the case of open defaults. We do this by first defining the notion of an extension for default theories with open defaults. Then we focus upon normal default theories and generalize the proof theory of Section 5 for this case. Finally, the results of Section 6 on belief revision are suitably generalized.

7.1. The extensions for arbitrary default theories

The first task at hand is to suitably generalize the concept of an extension from closed default theories to arbitrary theories. Intuitively, we want to interpret the open default $\alpha(\mathbf{x}) : M\beta_1(\mathbf{x}), \dots, M\beta_m(\mathbf{x}) / w(\mathbf{x})$ as saying something like: "For all individuals x_1, \dots, x_n , if $\alpha(\mathbf{x})$ is believed and each of $\beta_1(\mathbf{x}), \dots, \beta_m(\mathbf{x})$ can be consistently believed, then one is permitted to believe $w(\mathbf{x})$." The initial problem then is to specify exactly what are the individuals of a default theory.

To begin with a default theory is defined over the language L_A whose alphabet is A . If $F \subseteq A$ is the set of all function letters of A (including the 0-ary function letters, or constant letters) then any term constructable using the function letters of F should count as an individual of the default theory, so that the free variables of an open default should at least be seen as ranging over these terms. This seems fine with respect to these explicitly nameable individuals. However, there will generally be implicitly defined individuals which are introduced via existential quantifiers. For example, the theory

$$\frac{:MP(x)}{P(x)} \quad (Ex)Q(x) \quad \neg P(a)$$

has an explicitly named individual a , together with an implicitly defined individual which satisfies Q . Intuitively one would expect of this implicitly defined individual that it satisfy P in the extension for this theory. In order to accommodate this intuition, we introduce a new constant letter α distinct from any element of F to denote this implicitly defined individual and replace $(\text{Ex})Q(x)$ by $Q(\alpha)$. This immediately yields $P(\alpha)$ in the extension whence $(\text{Ex})Q(x) \wedge P(x)$ is in the extension.⁵

What the previous discussion amounts to is the suggestion that each wff of W be replaced by its Skolemized form. If Σ is the set of Skolem functions so introduced, then we would like to view the individuals constructable from the elements of $F \cup \Sigma$ as the individuals of the default theory, and the free variables of an open default as ranging over these individuals.

While this feels closer to the truth, there still remains the problem of implicitly defined individuals which may be introduced by default. For example, in the theory

$$\frac{:\text{M}(\text{Ex})P(x)}{(\text{Ex})P(x)} \quad \frac{:\text{M}Q(x)}{Q(x)}$$

there is an implicitly defined individual which is introduced by the first default, and which satisfies P . One would expect, by the second default, that this individual will also satisfy Q , i.e. one would expect $(\text{Ex})P(x) \wedge Q(x)$ to be in the extension. As before, we introduce a new constant letter α to denote this individual, and represent the first default by its Skolemized form $:\text{M}P(\alpha)/P(\alpha)$. This immediately yields $P(\alpha)$ in the extension and then, by the second default, $Q(\alpha)$ is in the extension, and hence so is $(\text{Ex})P(x) \wedge Q(x)$.

In general then, if Σ is the set of Skolem functions introduced by Skolemizing W as well as all of the defaults, then we shall view the individuals constructable from the elements of $F \cup \Sigma$ as the individuals of the default theory, and the free variables of an open default as ranging over these individuals. We proceed to make these ideas precise.

Let w be a closed wff of L_A . With no loss in generality assume w is in prenex form. Then the *Skolemized form* of w is obtained as follows (Robinson (1965)): replace each existentially quantified variable y of w by $\sigma(x_1, \dots, x_n)$ where $(x_1), \dots, (x_n)$ are all of the universal quantifiers preceding $(\text{E}y)$ in the prefix of w . σ must be a new function letter distinct from any in the alphabet A and distinct from any other such function letters previously introduced. Do this for all existentially quantified variables of w , and then delete all of w 's quantifiers. The result is a quantifier free formula. The new function letters σ so introduced are called *Skolem functions*.

⁵ It would appear from the definition of a fixed point in McDermott and Doyle (1978) that $(\text{Ex})Q(x) \wedge P(x)$ is not in the fixed point of this theory, which suggests some fundamental differences between default logic and their non-monotonic logic.

We also require the notion of the *Skolemized form* of a default

$$\frac{\alpha(\mathbf{x}) : M\beta_1(\mathbf{x}), \dots, M\beta_m(\mathbf{x})}{w(\mathbf{x})}$$

This is obtained by replacing the wff $w(\mathbf{x})$ by the Skolemized form of $(\mathbf{x})w(\mathbf{x})$. For example the Skolemized form of

$$\frac{(E y)(z)P(x, y, z) : M(E y)Q(x, y)}{(E y)(z)(E w)R(x, y, z, w)}$$

is

$$\frac{(E y)(z)P(x, y, z) : M(E y)Q(x, y)}{R(x, f(x), z, g(x, z))},$$

where f and g are Skolem functions. Notice that the ‘top half’ of a default remains unchanged in converting it to its Skolemized form.

A default theory (D, W) is in *Skolemized form* iff all of the defaults of D , and all of the wffs of W are in Skolemized form.

Let G be a set of function letters. Then $H(G)$ is the smallest set of terms such that $a \in H(G)$ for all constant letters $a \in G$, and if $t_1, \dots, t_n \in H(G)$ and $g \in G$ is an n -ary function letter then $g(t_1, \dots, t_n) \in H(G)$. $H(G)$ is simply the set of all terms constructable using the function letters of G .

Suppose $\Delta = (D, W)$ is a Skolemized default theory (not necessarily normal). Suppose further that Σ is the set of Skolem functions of Δ , and that F is the set of function letters of the alphabet A . Define

$$\begin{aligned} \text{CLOSED-DEFAULTS}(\Delta) \\ = \{ \delta(\mathbf{g}) \mid \delta(\mathbf{x}) \in D \text{ and } \mathbf{g} \text{ is a tuple of ground terms of } H(F \cup \Sigma) \} \end{aligned}$$

Thus $\text{CLOSED-DEFAULTS}(\Delta)$ is the set of all ground instances over $H(F \cup \Sigma)$ of the defaults of D . Notice that each such ground instance of a default is a closed default. The extensions for a general (not necessarily closed) default theory can now be defined as follows:

Definition 5. E is an *extension* for $\Delta = (D, W)$ iff E is an extension for the closed default theory $\text{CLOSED}(\Delta) = (\text{CLOSED-DEFAULTS}(\Delta), W)$.

There are several points worth noting about this definition:

(1) The set $\text{CLOSED-DEFAULTS}(\Delta)$ will in general be countably infinite so that $\text{CLOSED}(\Delta)$ is a closed default theory with countably infinitely many defaults. Since the entire preceding development for closed default theories (Sections 2–6) makes no finiteness assumption about the number of defaults, these results apply equally to $\text{CLOSED}(\Delta)$. In particular then, the definition of an extension for Δ as being that for $\text{CLOSED}(\Delta)$ makes sense.

(2) Initially we began with an unskolemized default theory defined over a first order language L_A whose alphabet A contains a set of function letters F . In Skolemizing this theory, we introduced a new set of function letters Σ , the Skolem functions

of the theory, and $\text{CLOSED}(\Delta)$ is a closed default theory over the first order language $L_{A \cup \Sigma}$. Thus the extensions for Δ consist of closed wffs of the extended language $L_{A \cup \Sigma}$, not of the original language L_A .

(3) In view of the observation in (2), the notion of an extension for a general default theory Δ is not quite a generalization of this notion for closed default theories. For if Δ is closed to begin with, then its extensions will be subsets of $L_{A \cup \Sigma}$ whereas in the definition of an extension for closed theories given in Section 2 these extensions are subsets of L_A . This is not a substantive issue, however. We can easily modify the definition of extension given in Section 2 so that these are subsets of $L_{A \cup \Sigma}$ rather than of L_A , without affecting the subsequent theory. The effect of this modification is merely to admit into extensions certain wffs containing Skolem terms of the original closed default theory.

7.2. Normal default theories: proof theory

Before defining a proof procedure for general normal default theories we shall modify the top down default proofs for closed normal theories of Section 5 to a form that is more appropriate for generalization to the case of open normal defaults. The principal change is that we shall require the notion of a linear resolution proof of an arbitrary wff where this wff is indexed by some set of defaults.

Definition 6. A linear resolution proof of (β, D'_0) from some set S of indexed clauses has the form of Fig. 4, where

- (1) The top indexed clause (R_0, D'_0) is such that R_0 is a clause of $\neg\beta$.
- (2) For $1 \leq i \leq n$, (R_i, D'_i) is a resolvent of the indexed clauses (R_{i-1}, D'_{i-1}) and (C_{i-1}, D_{i-1}) .

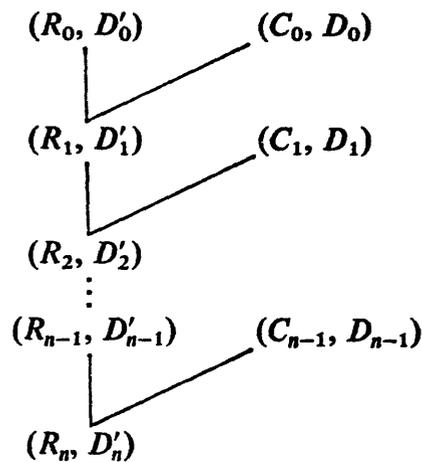


FIG. 4.

(3) For $0 \leq i \leq n-1$, $(C_i, D_i) \in S$ or C_i is a clause of $\neg\beta$ and $D_i = D'_0$, or (C_i, D_i) is (R_j, D'_j) for some $j < i$.

(4) $R_n = \square$.

As before, we say that such a proof *returns* D_n . The clauses (C_i, D_i) are called the *side clauses* of the proof. For $1 \leq i \leq n$, the clauses (R_i, D'_i) are called *centre clauses*.

Suppose L is a linear resolution proof from $\text{CLAUSES}(\Delta)$. If $(C, \{\delta\}) \in \text{CLAUSES}(\Delta)$ is a side clause of L , then the default δ is said to have been *introduced into* L .

Definition 7. A *top down default proof* of β with respect to a closed normal default theory $\Delta = (D, W)$ is a sequence of linear resolution proofs L_0, \dots, L_k such that

(1) L_0 is a linear resolution proof of $(\beta, \{ \})$ from $\text{CLAUSES}(\Delta)$,

(2) For $0 \leq i \leq k$, L_i returns D_i ,

(3) For $1 \leq i \leq k$, let $D^{(i-1)}$ be the set of defaults introduced into L_{i-1} . Then L_i is a linear resolution proof of $(\text{PREREQUISITES}(D^{(i-1)}), D_{i-1})$ from $\text{CLAUSES}(\Delta)$.

(4) The set of defaults introduced into L_k is empty.

(5) $W \cup \text{CONSEQUENTS}(D_k)$ is satisfiable.

Example 7.1. In Example 5.2, Fig. 2(b) represented a top down default proof of D . Fig. 5 is the corresponding proof under the current modified notion of a top down default proof.

Example 7.1 should make it clear that there is no essential difference between the present notion of a top down default proof and that of Section 5. The only distinction is that in the current notion, each subproof L_i carries along with it all of the defaults introduced into the subproofs L_0, \dots, L_{i-1} , so that the final subproof L_k returns D_k , the set of all of the defaults introduced into all of the subproofs. For closed default theories this modified notion is of no consequence. In fact the earlier notion of Section 5 was rather more transparent. However, for the purpose of generalizing a proof theory to the case of open normal defaults, the modified version of top down default proofs is just what is needed. We now proceed with this generalization.

Let $\Delta = (D, W)$ be a normal default theory, assumed Skolemized. With no loss in generality, assume W is a set of clauses. If $\delta(\mathbf{x}) = (\alpha(\mathbf{x}) : M w(\mathbf{x}) / w(\mathbf{x})) \in D$, suppose $C_1(\mathbf{x}), \dots, C_r(\mathbf{x})$ are all of the clauses of $w(\mathbf{x})$. Then for $1 \leq i \leq r$ the ordered pair $(C_i(\mathbf{x}), \{\delta(\mathbf{x})\})$ is called a *consequent clause* of the default $\delta(\mathbf{x})$. Define

$$\text{CLAUSES}(\Delta) = \{(C(\mathbf{x}), \{\delta(\mathbf{x})\}) \mid \delta(\mathbf{x}) \in D \text{ and } (C(\mathbf{x}), \{\delta(\mathbf{x})\}) \text{ is a consequent clause of } \delta(\mathbf{x})\} \cup \{(C, \{ \}) \mid C \in W\}$$

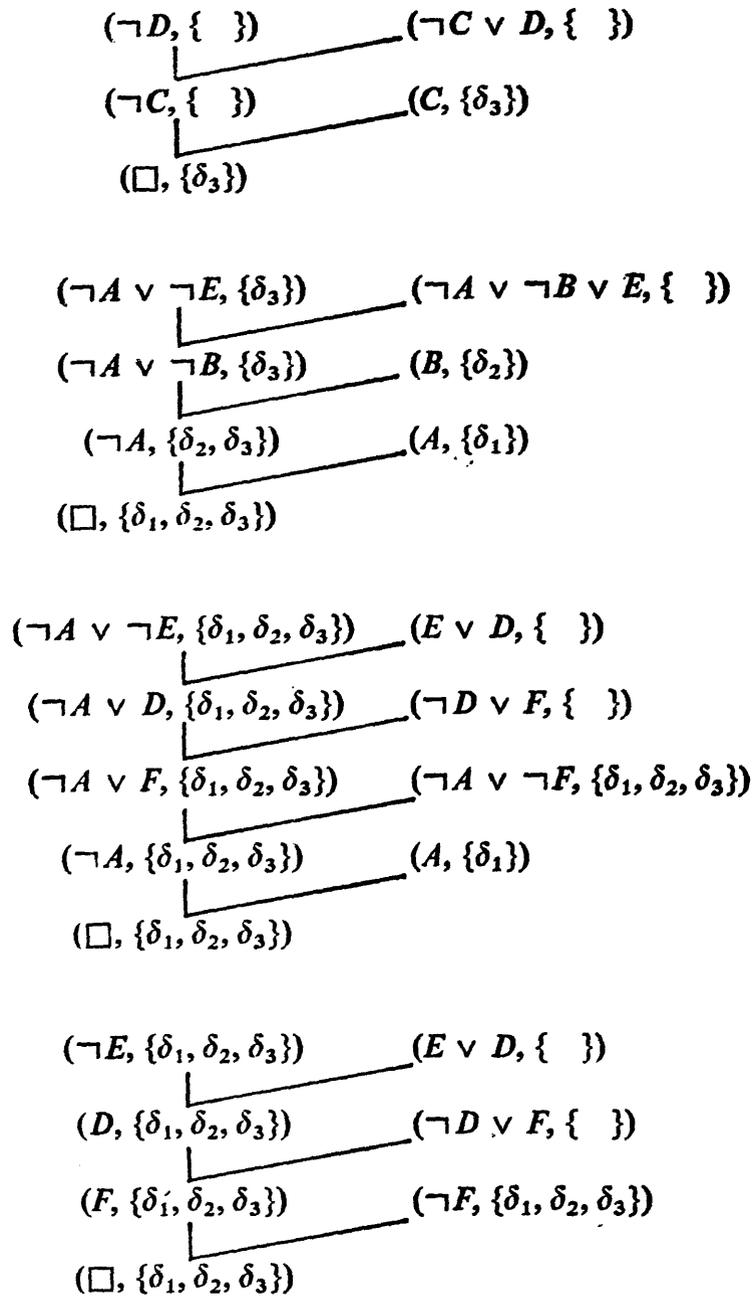


FIG. 5.

As before, an ordered pair (C, D) where C is a clause and D a set of defaults is called an *indexed clause*. If (C_1, D_1) and (C_2, D_2) are indexed clauses, define their resolvent as follows:

Let (C'_1, D'_1) and (C'_2, D'_2) be obtained by uniform renamings⁶ of the variables occurring in (C_1, D_1) and (C_2, D_2) respectively such that (C'_1, D'_1) has no variable in common with (C'_2, D'_2) . Suppose R is an ordinary resolvent of C'_1 and C'_2 under most general unifier (mgu) σ . Then $(R, (D'_1 \cup D'_2)\sigma)$ is a *resolvent* of the indexed clauses (C_1, D_1) and (C_2, D_2) .

With this notion of the resolvent of two indexed clauses the definition of a linear resolution proof of (β, D'_0) from a set S of indexed clauses is the same as before. The only difference is that here the defaults taking part in the proof have free variables and these will become instantiated during the course of the proof.

Consider a linear resolution proof of the form of Fig. 4. Assume that (R_i, D'_i) and (C_i, D_i) are variable disjoint, and that their resolvent (R_{i+1}, D'_{i+1}) is formed under mgu σ , so that $D'_{i+1} = (D'_i \cup D_i)\sigma$. If $\delta(t) \in D'_i$ for some terms t , then $\delta(t)\sigma \in D'_{i+1}$. This occurrence of $\delta(t)\sigma$ in D'_{i+1} is said to be a *descendant* of the occurrence of $\delta(t)$ in D'_i . Similarly, if $\delta(t) \in D_i$, then again $\delta(t)\sigma \in D'_{i+1}$ and this occurrence of $\delta(t)\sigma$ in D'_{i+1} is a *descendant* of the occurrence of $\delta(t)$ in D_i . Finally, for $i < j < k$, if an occurrence of $\delta(t)$ in D'_k is a descendant of an occurrence of $\delta(s)$ in D'_j which in turn is a descendant of an occurrence of $\delta(r)$ in D'_i or in D_i , then $\delta(t)$ is a *descendant* of the occurrence of $\delta(r)$ in D'_i or D_i .

The descendants of an occurrence of a default in the tree of Fig. 4 are simply the successive instances of that default created under successive resolution operations. If a default occurs at node (C_i, D_i) , then it will have descendants at each successor node $(R_{i+1}, D'_{i+1}), \dots, (R_n, D'_n)$. Likewise if a default occurs at node (R_i, D'_i) .

We are now almost in a position to propose the necessary generalization to open defaults of the top down default proofs for closed normal default theories. The following notion comes very close to what is needed.

Definition 8. Let $\Delta = (D, W)$ be a normal default theory, assumed Skolemized, and suppose Σ is the set of Skolem functions of Δ . A sequence of linear resolution proofs L_0, \dots, L_k is an *admissible proof sequence* for β with respect to Δ iff

- (1) L_0 is a linear resolution proof of $(\beta, \{ \})$ from $\text{CLAUSES}(\Delta)$.
- (2) For $0 \leq i \leq k$, L_i returns D_i .
- (3) For $1 \leq i \leq k$, let $D^{(i-1)}$ be the set of those defaults in D_{i-1} which are descendants of the defaults introduced into L_{i-1} . Then L_i is a linear resolution proof of $(\text{PREREQUISITES}(D^{(i-1)}), D_{i-1})$ from $\text{CLAUSES}(\Delta)$.
- (4) The set of defaults introduced into L_k is empty.

⁶ By a uniform renaming of the variables of (C, D) we mean that if x is renamed as y , then every occurrence of x in C as well as in D is replaced by y .

(5) There is a ground substitution γ over $H(F \cup \Sigma)$ such that $D_k\gamma$ has the following two properties:

- (i) Each default instance occurring in $D_k\gamma$ is a ground instance over $H(F \cup \Sigma)$, i.e. if $\delta(\mathbf{g}) \in D_k\gamma$, then each component g_i of the tuple of ground terms \mathbf{g} is an element of $H(F \cup \Sigma)$.⁷
- (ii) $W \cup \text{CONSEQUENTS}(D_k\gamma)$ is satisfiable.

We call the set of ground defaults $D_k\gamma$ the *default support* of the admissible proof sequence. Notice that by footnote 7 the default support is a subset of $\text{CLOSED-DEFAULTS}(\Delta)$.

Example 7.2. Consider the normal default theory Δ with defaults

$$\delta_1(x) = \frac{P(x): \mathbf{M}(\mathbf{E}y)Q(x, y) \wedge R(x, y)}{(\mathbf{E}y)Q(x, y) \wedge R(x, y)}$$

or, in Skolemized form,

$$\delta_1(x) = \frac{P(x): \mathbf{M}(\mathbf{E}y)Q(x, y) \wedge R(x, y)}{Q(x, f(x)) \wedge R(x, f(x))};$$

$$\delta_2(x, y) = \frac{Q(x, y): \mathbf{M}(z)S(x, y, z)}{(z)S(x, y, z)}$$

or, in Skolemized form,

$$\delta_2(x, y) = \frac{Q(x, y): \mathbf{M}(z)S(x, y, z)}{S(x, y, z)};$$

$$\delta_3(x, y) = \frac{R(x, y): \mathbf{M}U(x)}{U(x)}; \quad W = \{P(a)\}.$$

Fig. 6(a) is an admissible proof sequence for $(\mathbf{E}x)(\mathbf{E}y)(z).S(x, y, z) \wedge U(x)$. The negation of this is $(x)(y)(\mathbf{E}z). \neg S(x, y, z) \vee \neg U(x)$ which, after introducing a Skolem function g for the existentially quantified z , leads to the clause $\neg S(x, y, g(x, y)) \vee \neg U(x)$. This is the top clause of Fig. 6(a). For this example γ is the empty substitution so the default support for this proof sequence is $\{\delta_3(a, f(a)), \delta_2(a, f(a)), \delta_1(a)\}$.

Let us extend the notion of a descendant to admissible proof sequences L_0, \dots, L_k . For $1 \leq i \leq k$ an occurrence of a default in the top clause of L_i is a *descendant* of its occurrence in the bottom clause of L_{i-1} . And for $h \leq i \leq j$ if an occurrence of $\delta(\mathbf{t})$ in L_j is a descendant of an occurrence of $\delta(\mathbf{s})$ in L_i which in turn is a descendant of an occurrence of $\delta(\mathbf{r})$ in L_h , then $\delta(\mathbf{t})$ is a *descendant* of the occurrence of $\delta(\mathbf{r})$ in L_h .

If one imagines the successive trees L_0, \dots, L_k placed vertically, each beneath its predecessor, as in Fig. 6(a) then the descendants of a default occurrence are simply its successive instances in all of the left hand nodes below the node in which

⁷ Notice that by this property $\delta(\mathbf{g})$ is a closed normal default, and that $\delta(\mathbf{g}) \in \text{CLOSED-DEFAULTS}(\Delta)$.

it occurs. It follows that if L_0, \dots, L_k is an admissible proof sequence then every default occurrence has a descendant in (\square, D_k) , the bottom clause of L_k . Suppose $\delta(t)$ is a default occurrence whose descendant in D_k is $\delta(s)$. Then $\delta(s)\gamma \in D_k\gamma$, the default support of the admissible proof sequence. We call $\delta(s)\gamma$ the *terminal descendant* of the default occurrence $\delta(t)$. By the nature of γ (see footnote 7 associated with the Definition 8) this terminal descendant is a ground instance of $\delta(t)$ and is an element of $\text{CLOSED-DEFAULTS}(\Delta)$.

Example 7.2 (continued). Suppose, in the admissible proof sequence for $(\text{Ex})(\text{Ey})(z)S(x,y,z) \wedge U(x)$ of Fig. 6(a) we do the following to each indexed clause (C,D) : if $\delta \in D$, instantiate the variables of δ so that it becomes identical with its terminal descendant, and instantiate the corresponding variables of C in the same way. The result is Fig. 6(b), which is a top down default proof of $(\text{Ex})(\text{Ey})(z)S(x,y,z) \wedge U(x)$ with respect to $\text{CLOSED}(\Delta)$, i.e. $(\text{Ex})(\text{Ey})(z)S(x,y,z) \wedge U(x)$ is in some extension for $\text{CLOSED}(\Delta)$ and hence by definition is in an extension for Δ .

This example indicates the basic intuition behind the notion of an admissible proof sequence. Unfortunately, this notion does not quite give us what we want.

Example 7.3. Δ has one default

$$\delta(x) = \frac{P(x) \vee Q(x): \text{MR}(x)}{R(x)} \quad W = \{P(a) \vee Q(b)\}$$

Fig. 7(a) is an admissible proof sequence for $(\text{Ey})R(y)$. But the only extension for this theory is $\text{Th}(W)$ and this does not contain $(\text{Ey})R(y)$. Notice however that if we replace each default occurrence of Fig. 7(a) by its terminal descendant, as we did for Example 7.2, we obtain Fig. 7(b) which is not a top down default proof of $(\text{Ey})R(y)$ with respect to $\text{CLOSED}(\Delta)$. The basic reason for this is that the terminal descendant of the occurrence of $\delta(x)$ in the clause $(\neg P(x), \{\delta(x)\})$ in Fig. 7(a) is $\delta(a)$, and this is different than $\delta(b)$, the terminal descendant of $\delta(x)$ in $(\neg Q(x), \{\delta(x)\})$ despite the fact that $\neg P(x)$ and $\neg Q(x)$ are both of the clauses of $\neg \text{PREREQUISITES}(\{\delta(x)\})$.

To obtain a proof theory for open defaults it is necessary to restrict the class of admissible proof sequences in order to exclude situations like that of Example 7.3.

Definition 9. An admissible proof sequence L_0, \dots, L_k for β with respect to a normal default theory Δ is a *top down default proof* of β with respect to Δ iff:

(1) Suppose, for $1 \leq i \leq k$, that L_i has top clause (C, D_{i-1}) and a side clause (C', D_{i-1}) where C and C' are both clauses of the negation of $\text{PREREQUISITES}(D^{(i-1)})$. Then the terminal descendants of the defaults of D_{i-1} in the top clause are identical with the terminal descendants of the defaults of D_{i-1} in the side clause.

(2) Suppose, for $0 \leq i \leq k$, that (C, D) is a centre clause of L_i which also occurs as a side clause of L_i . Then the terminal descendants of the defaults of D in the centre clause are identical with the terminal descendants of the defaults of D in the side clause.

Theorem 7.1. *Let $\Delta = (D, W)$ be a normal default theory, and let $\beta \in L_{A \cup \Sigma}$ be a closed wff where Σ is the set of Skolem functions of Δ . If there is a top down default proof of β with respect to Δ , then Δ has an extension which contains β .*

Proof. Suppose L_0, \dots, L_k is a top down default proof of β with respect to Δ . Our plan is to construct a top down default proof L'_0, \dots, L'_k of β with respect to $\text{CLOSED}(\Delta)$. The theorem will then follow by the completeness of top down default proofs for closed normal theories, and the definition of an extension of Section 7.1.

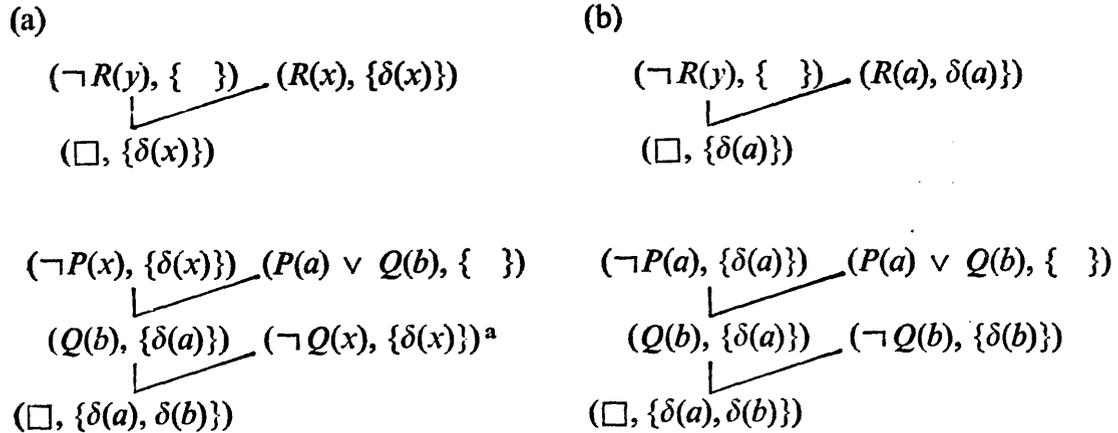


FIG. 7.

^a $\neg Q(x)$ is the other clause of the negation of $\text{PREREQUISITES}(\{\delta(x)\})$.

To begin, notice that the defaults in any occurrence of an indexed clause in L_0, \dots, L_k are functioning as ‘answer literals’ (Green (1969)). The purpose served by the introduction of a default $\delta(x)$ into a proof is to record the fact that $\delta(x)$ is required in the proof, as well as to provide a history of the instances assumed by the free variables x during the course of the proof.

Now consider any linear resolution proof L_i in the sequence L_0, \dots, L_k . For any occurrence of an indexed clause (C, D) in L_i , let σ be that substitution such that for each default $\delta(t) \in D$, $\delta(t)\sigma$ is the terminal descendant of this occurrence of $\delta(t)$. Define $(C, D)\sigma$ to be the g -instance of this occurrence of (C, D) in L_i . (Thus to obtain the g -instance of (C, D) , instantiate the free variables of each default of D such that it becomes identical with its terminal descendant, and instantiate the corresponding variables of C in the same way.) Now replace each indexed clause of L_i by its g -instance. The result is a tree L'_i . For example Fig. 6(b) is obtained from Fig. 6(a) by replacing each indexed clause by its g -instance.

We first show that L'_i is in fact a linear resolution proof, i.e. that successive clauses in the tree L'_i are indexed resolvents of their immediate predecessors. Now L_i returns a set of defaults D_i which is the ‘answer clause’ of the deduction L_i . Moreover, the set of terminal descendants of the defaults of D_i is an instance of D_i ,

i.e. this set is an instance of the 'answer clause' of L_i . It then follows from the results in Luckham and Nilsson (1971) that the tree L'_i represents a linear resolution proof.

To prove that L'_0, \dots, L'_k is a top down default proof of β with respect to $\text{CLOSED}(\Delta)$ it is necessary to prove:

- (1) L'_0 is a linear resolution proof of $(\beta, \{ \})$ from $\text{CLAUSES}(\text{CLOSED}(\Delta))$.
- (2) For $0 \leq i \leq k-1$, if $D^{(i)}$ is the set of defaults introduced into L'_i and if L'_i returns D'_i then L'_{i+1} is a linear resolution proof of $(\text{PREREQUISITES}(D^{(i)}), D'_i)$ from $\text{CLAUSES}(\text{CLOSED}(\Delta))$.
- (3) The set of defaults introduced into L'_k is empty.
- (4) If L'_k returns D'_k then $W \cup \text{CONSEQUENTS}(D'_k)$ is satisfiable.

Property (3) follows immediately from the corresponding property for L_k . To prove (4) notice that by the construction of L'_0, \dots, L'_k it is the case that D'_k is precisely the default support of the admissible proof sequence L_0, \dots, L_k so by the definition of an admissible proof sequence, $W \cup \text{CONSEQUENTS}(D'_k)$ is satisfiable.

We now prove (1). We begin by proving that L'_0 is a linear resolution proof of $(\beta, \{ \})$ from $\text{CLAUSES}(\text{CLOSED}(\Delta))$. We know that L_0 is a linear resolution proof of $(\beta, \{ \})$ from $\text{CLAUSES}(\Delta)$ and L'_0 is obtained from L_0 by replacing each indexed clause by its g -instance. If, for some clause C_β of $\neg\beta$ ($C_\beta, \{ \}$) is a top or side clause of L_0 , then it is its own g -instance and $(C_\beta, \{ \})$ is the corresponding top or side clause of L'_0 . Similarly, if $(C, \{ \}) \in \text{CLAUSES}(\Delta)$ (so that $C \in W$) is a side clause of L_0 then it is a side clause of L'_0 and $(C, \{ \}) \in \text{CLAUSES}(\text{CLOSED}(\Delta))$. Suppose $(C(\mathbf{x}), \{\delta(\mathbf{x})\}) \in \text{CLAUSES}(\Delta)$ (so that $C(\mathbf{x})$ is a clause of the consequent of $\delta(\mathbf{x})$) is a side clause of L_0 . Then if $\delta(\mathbf{g})$ is the terminal descendant of this occurrence of $\delta(\mathbf{x})$, then $(C(\mathbf{g}), \{\delta(\mathbf{g})\})$ is the g -instance of $(C(\mathbf{x}), \{\delta(\mathbf{x})\})$ and this is the corresponding side clause of L'_0 ; clearly $(C(\mathbf{g}), \{\delta(\mathbf{g})\}) \in \text{CLAUSES}(\text{CLOSED}(\Delta))$. Finally, suppose (C, D) is a side clause of L_0 which is also a centre clause of L_0 . Then in L'_0 the corresponding side clause is $(C, D)\sigma$, the g -instance of the occurrence of (C, D) as a side clause of L_0 . Also, in L'_0 the corresponding centre clause is $(C, D)\sigma'$, the g -instance of the occurrence of (C, D) as a centre clause of L_0 . But by condition (2) of Definition g , $\sigma = \sigma'$ so that in L'_0 the side clause $(C, D)\sigma$ is also a centre clause. This completes the proof of (1).

It remains to prove (2), i.e. that L'_{i+1} is a linear resolution proof of $(\text{PREREQUISITES}(D^{(i)}), D'_i)$ from $\text{CLAUSES}(\text{CLOSED}(\Delta))$. We already know that L'_{i+1} is a linear resolution proof. Hence it is sufficient to prove:

- (a) Whenever $(C(\mathbf{x}), \{ \}) \in \text{CLAUSES}(\Delta)$ is a side clause of L'_{i+1} then its g -instance, which is identical to itself, is in $\text{CLAUSES}(\text{CLOSED}(\Delta))$. This is immediate.
- (b) Whenever $(C(\mathbf{x}), \{\delta(\mathbf{x})\}) \in \text{CLAUSES}(\Delta)$ is a side clause of L'_{i+1} then its g -instance $(C(\mathbf{g}), \{\delta(\mathbf{g})\})$ is in $\text{CLAUSES}(\text{CLOSED}(\Delta))$. This is immediate.

(c) Whenever (C, D) is a side clause of L_{i+1} which is also a centre clause of L_{i+1} then the g -instance of its occurrence as a side clause in L_{i+1} is identical to the g -instance of its occurrence as a centre clause in L_{i+1} . The proof of this is the same as for L_0 .

(d) Whenever (C, D_i) occurs in L_{i+1} , either as the top clause or as a side clause, where C is a clause of $\neg \text{PREREQUISITES}(D^{(i)})$, then its g -instance $(C\sigma, D_i\sigma)$ is such that $C\sigma$ is a clause of $\neg \text{PREREQUISITES}(D^{(i)'})$ and $D_i\sigma = D_i'$.

To prove (d) we distinguish two cases:

Case 1. (C, D_i) is the top clause of L_{i+1} .

Now $D_i\sigma$ is the set of terminal descendants of the default occurrences in (C, D_i) , the top clause of L_{i+1} . But these are the same as the terminal descendants of the default occurrences in (\square, D_i) , the bottom clause of L_i . Since L_i' returns D_i' , this latter clause has g -instance (\square, D_i') . Hence $D_i\sigma = D_i'$. Moreover, $D^{(i)}\sigma$ is the set of terminal descendants of the defaults introduced into L_i and by the construction of L_i' , this is the same as the set of defaults introduced into L_i' , i.e. $D^{(i)}\sigma = D^{(i)'}$. Hence

$\text{PREREQUISITES}(D^{(i)})\sigma = \text{PREREQUISITES}(D^{(i)}\sigma) = \text{PREREQUISITES}(D^{(i)'})$. Since C is a clause of $\neg \text{PREREQUISITES}(D^{(i)})$, $C\sigma$ is a clause of $\neg \text{PREREQUISITES}(D^{(i)'})$.

Case 2. (C, D_i) is a side clause of L_{i+1} .

Then if (B, D_i) is the top clause of L_{i+1} , and $(B\sigma', D_i\sigma')$ its g -instance, then by condition (1) of Definition g $\sigma = \sigma'$ so that $D_i\sigma = D_i\sigma'$. By Case 1, $D_i\sigma' = D_i'$ so $D_i\sigma = D_i'$. By the same argument as in Case 1 $C\sigma$ is a clause of $\neg \text{PREREQUISITES}(D^{(i)'})$.

We shall say, as we did for closed theories, that a default theory is *consistent* iff it has a consistent extension.

Theorem 7.2. *Let $\Delta = (D, W)$ be a consistent normal default theory. If Δ has an extension which contains β then there is a top down default proof of β with respect to Δ .*

Proof. We shall merely sketch a proof. Since Δ has an extension E containing β then E is an extension for the closed normal default theory $\text{CLOSED}(\Delta) = (\text{CLOSED-DEFAULTS}(\Delta), W)$. $\text{CLOSED}(\Delta)$ is consistent since Δ is. Then by the completeness of top down default proofs for closed normal default theories (Theorem 5.1), there is a top down default proof L'_0, \dots, L'_k of β with respect to $\text{CLOSED}(\Delta)$. By using standard lifting techniques from resolution theory (Robinson (1965)) we can lift this proof L'_0, \dots, L'_k to a top down default proof L_0, \dots, L_k of β with respect to Δ . Essentially this is done by first lifting L'_0 to L_0 by replacing, in L'_0 , each clause $\{C(g), \{\delta(g)\}\} \in \text{CLAUSES}(\text{CLOSED}(\Delta))$ by its

uninstantiated general clause $(C(\mathbf{x}), \{\delta(\mathbf{x})\}) \in \text{CLAUSES}(\Delta)$ and using this general clause to form the resulting resolvent. In the case of a side clause of L'_0 which is also a centre clause, this centre clause will already have been lifted and it is this lifted clause which replaces the side clause. In general then, assume L'_0, \dots, L'_i have been lifted to L_0, \dots, L_i respectively. Assume L'_i and L_i return D'_i and D_i respectively, and that $D^{(i)'}$ and $D^{(i)}$ are the defaults introduced into L'_i and L_i respectively. Now to lift L'_{i+1} to L_{i+1} first lift its top clause (C', D'_i) . C' is a clause of $\neg \text{PREREQUISITES}(D^{(i)'})$. Now D'_i is an instance of D_i so that (C', D'_i) is an instance of a clause (C, D_i) where C is a clause of $\neg \text{PREREQUISITES}(D^{(i)})$. Replace (C', D'_i) in L'_{i+1} by (C, D_i) . Similarly, if (C', D'_i) is a side clause of L'_{i+1} replace it by (C, D_i) . To lift the remaining side clauses of L'_{i+1} proceed as for L'_0 .

The resulting lifted proofs L_0, \dots, L_k have the following properties:

- (1) L_0 is a linear resolution proof of $(\beta, \{ \})$ from $\text{CLAUSES}(\Delta)$.
- (2) For $1 \leq i \leq k$, L_i is a linear resolution proof of $(\text{PREREQUISITES}(D^{(i-1)}), D_{i-1})$ from $\text{CLAUSES}(\Delta)$.
- (3) The set of defaults introduced into L_k is empty.

Moreover, since L'_0, \dots, L'_k is a top down default proof of β with respect to $\text{CLOSED}(\Delta)$, $D'_k \subseteq \text{CLOSED-DEFAULTS}(\Delta)$. Since D'_k is a ground instance of D_k there is a ground substitution γ over $H(F \cup \Sigma)$, where Σ is the set of Skolem functions of Δ , such that $D'_k = D_k\gamma$. Finally, since $W \cup \text{CONSEQUENTS}(D'_k)$ is satisfiable, so is $W \cup \text{CONSEQUENTS}(D_k\gamma)$. We have proved that L_0, \dots, L_k is an admissible proof sequence for β with respect to Δ .

To see that L_0, \dots, L_k is a top down default proof of β , notice that if (C', D') is a clause of L' which was lifted to a clause (C, D) of L_i , then the terminal descendants of the defaults of D are precisely the defaults of D' . From this the result follows.

The combination of Theorems 7.1 and 7.2 yields the following:

Theorem 7.3 (Completeness of top down default proofs). *Let Δ be a consistent normal default theory and Σ the set of Skolem functions of Δ . Suppose $\beta \in L_{A \cup \Sigma}$ is a closed wff. Then Δ has an extension containing β iff β has a top down default proof with respect to Δ .*

7.3. Open defaults: other generalizations

In view of the definition of an extension for default theories with open defaults, all of the results of Section 3 continue to hold for normal default theories in general. In particular then, any such theory always has an extension, its extensions are orthogonal, it will be semi-monotonic, and the addition of new normal defaults cannot decrease the number of its extensions.

To generalize the results of Section 6 on belief revision requires only defining, for a top down default proof P_β of β , $\text{DS}(P_\beta)$ to be the default support of P_β , i.e., with reference to Definition 8 in Section 7.2, $\text{DS}(P_\beta)$ is defined to be $D_k\gamma$. In the

case of closed normal default theories $DS(P_\beta)$ is the set of all of the defaults which enter into the proof P_β . In the case of normal default theories in general, the defaults entering into the proof P_β become instantiated during the course of the proof. $DS(P_\beta)$ is the set of all of the final instances of these defaults or, in the unlikely event that these final instances still contain variables, $DS(P_\beta)$ is obtained from these final instances by suitably substituting ground terms of $H(F \cup \Sigma)$ for these variables in such a way that $W \cup \text{CONSEQUENTS}(DS(P_\beta))$ is satisfiable. With this notion of $DS(P_\beta)$ all of the results on belief revision of Section 6 continue to hold for normal default theories in general.

8 Discussion and Further Problems

Many interesting problems related to default logic remain to be explored. The most obvious formal omission in the previous sections is a model theory; the concept of an extension is defined proof theoretically. What one would like is a model theoretic characterization of the extensions of a default theory (D, W) in the following sense:

We want some way of viewing the defaults of D as restricting the models of W such that any resulting restricted set of models of W satisfies all and only the wffs of an extension for (D, W) ; and conversely, any extension for (D, W) will be the set of wffs satisfied by some such restricted set of models of W .

An interesting proposal which involves non-monotonic reasoning is the method of circumscription described in McCarthy (1977). Briefly circumscription is a formalization of the notion of reasoning within the minimal models of some set of first order axioms. There appears to be some relationship between circumscription and default logic. What can be proved is that the minimal models for a sentential set of wffs coincide with the models of the extensions induced on these wffs by the closed world assumption. This result suggests some deeper relationship between closed world default theories and the method of circumscription although I have been unable to discover just what this might be.

We have elsewhere (Reiter (1978a)) discussed some relationship between defaults and the 'negation' operator of Artificial Intelligence programming languages like PROLOG (Roussel (1975), Kowalski (1979)) and PLANNER (Hewitt (1972)). As implemented in these languages, $\text{NOT}(\langle \text{wff} \rangle)$ succeeds iff $\langle \text{wff} \rangle$ is not provable. It turns out that there are a number of difficulties with this notion of a procedural negation (Clark (1978)), not the least of which is specifying just what 'not provable' means in this setting. Accordingly, it might be of some interest to view the proof theory of Section 7 for normal default theories as a possible basis for defining the semantics of negation in such languages.

A major computational problem stems from the ubiquitous satisfiability tests required in the proof theory as well as in the conditions for belief revision. These all have the form of a satisfiability requirement on $W \cup \text{CONSEQUENTS}(D')$ for some set of defaults $D' \subseteq D$. While no algorithm exists for this, a variety of

heuristics are possible. One such heuristic would be to begin with the wffs of $\text{CONSEQUENTS}(D')$ and use these as a basis for successive 'forward chaining' or bottom up deductions into W , in order to derive new consequences of these. If, after some time bound or forward chaining level bound has been exceeded, no contradiction has been derived then one might assume satisfiability of the original set of wffs. While this appears to be the most general such heuristic it is also apt to be the most computationally expensive. Clever schemes should be possible. One such might involve extensive indexing of the wffs of W , perhaps along the lines of Kowalski (1974). Accessibility links in this index scheme could in some instances provide rapid tests for the satisfiability of W with respect to $\text{CONSEQUENTS}(D')$. The whole question of heuristics for testing satisfiability appears to be totally unexplored.

An important pragmatic problem is the specification of techniques for choosing between two or more competing extensions. In the hometown example of Section 1.3 we were faced with choosing between extensions, one in which Mary's hometown is Toronto because her spouse lives there, the other in which her hometown is Vancouver because her employer is located there.

One general heuristic would be to rank order conflicting defaults of this kind in order of their intuitive plausibility, selecting the highest priority default first in attempting a proof for some wff. Only during a process of belief revision would lower priority defaults be invoked. While some such static priority scheme might well be useful in certain settings, it seems not to work for the example of Mary's hometown. There appears to me to be no a priori reason for preferring Toronto over Vancouver, or vice versa here. Instead, what seems to be required is some dynamic scheme for *rationalizing* the apparent discrepancy. For example, having discovered that Mary's hometown might be both Vancouver and Toronto a default reasoning program should seek ways of resolving the conflict by searching out suitable detailed information about Mary. Is she separated from her spouse? If so, opt for her employer's location. Is she some sort of area representative for her employer? Is that area near Toronto? If so, choose her spouse's hometown. This view of extension selection as a process of rationalizing discrepancies has the nice property of avoiding the introduction of fine details, such as the possibility of Mary being an area representative, into the default reasoning process itself. Such details are taken into consideration only if the default reasoner discovers a discrepancy in its derivation of a belief. One can imagine a similar need for rationalization whenever belief revision is called for.

Only one same reason is shared by all of us:
we wish to create worlds as real as,
but other than the world that is.

John Fowles, *The French Lieutenant's Woman*

ACKNOWLEDGEMENTS

Keith Clark and Bob Kowalski patiently entertained the many incarnations of my ideas on default reasoning, and often asked the right question at the right time. Alan Bundy and David Warren provided valuable insights into the relationship between default proofs and PROLOG. Susan Frykberg through her speculative metaphysics prevented the enterprise from taking itself too seriously.

I am indebted to the University of British Columbia for granting me the sabbatical leave during which much of this work was done, and to Imperial College, London, and Bob Kowalski in particular, for making available to me the facilities of the Department of Computing and Controls. This paper was written and typed with the financial support of the National Research Council of Canada under grant A7642.

REFERENCES

- Bobrow, D. G. and Winograd, T., An overview of KRL-O, a knowledge representation language, *Cognitive Science* 1 (1) (January 1977).
- Church, A., An unsolvable problem of number theory, *Amer. J. Math.* 58 (1936) 345–363.
- Clark, K., Negation as failure, in: H. Gallaire and J. Minker (Eds.), *Logic and Data Base*. (Plenum Press, New York, 1978) 293–322.
- Doyle, J., *Truth Maintenance Systems for Problem Solving*, MIT Artificial Intelligence Laboratory, Report AI-TR-419 (January 1978).
- Green, C., Application of theorem proving to problem solving, *Proc. First IJCAI* (1969) 219–239.
- Hayes, P. J., The frame problem and related problems in artificial intelligence, in: A. Elithorn and D. Jones (Eds.), *Artificial and Human Thinking* (Jossey-Bass, San Francisco, 1973) 45–49.
- Hayes, P. J., *The Logic of Frames*, Technical Report, Dept. of Computer Science, Univ. of Essex (November 1977).
- Hewitt, C., *Description and Theoretical Analysis (Using Schemata) of PLANNER: A Language for Proving Theorems and Manipulating Models in a Robot*, A.I. Memo No. 251, MIT Project MAC, Cambridge, MA (April 1972).
- Kowalski, R. A., A Proof procedure using connection graphs, *J. ACM* 22 (October 1974) 572–595.
- Kowalski, R. A., *Logic for Problem Solving* (North-Holland Elsevier, New York, 1979).
- Kowalski, R. A. and Kuehner, D., Linear resolution with selection function, *Artificial Intelligence* 2 (1971) 221–260.
- Kramosil, I., A note on deduction rules with negative premises, *Proc. Fourth IJCAI* (1975) 53–56.
- Loveland, D. W., A linear format for resolution, *Proc. IRIA Symp. Automatic Demonstration*, Versailles, France, 1968 (Springer-Verlag, New York, 1970) 147–162.
- Luckham, D., Refinements in resolution theory, *Proc. IRIA Symp. Automatic Demonstration*, Versailles, France, 1968 (Springer-Verlag, New York, 1970) 163–190.
- Luckham, D. and Nilsson, N. J., Extracting information from resolution proof trees, *Artificial Intelligence* 2 (1971) 27–54.
- McCarthy, J., Epistemological problems of Artificial Intelligence, *Proc. Fifth IJCAI* (1977) 1038–1044.
- McCarthy, J. and Hayes, P. J., Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer and D. Michie (Eds.), *Machine Intelligence* 4 (Edinburgh University Press, Edinburgh, 1969) 463–502.
- McDermott, D. and Doyle, J., Non-monotonic Logic I, MIT AI Memo 486 (August 1978).
- Minsky, M., A framework for representing knowledge, in: P. Winston (Ed.), *The Psychology of Computer Vision* (McGraw-Hill, New York, 1975).
- Raphael, B., The frame problem in problem-solving systems, in: N. V. Findler and B. Meltzer (Eds.), *Artificial Intelligence and Heuristic Programming* (Edinburgh University Press, Edinburgh, 1971).

- Reiter, R., Two results on ordering for resolution with merging and linear format, *J.ACM* **18** (4) (1971) 630–646.
- Reiter, R., On reasoning by default, *Proc. Second Symp. on Theoretical Issues in Natural Language Processing*, Urbana, Illinois, July 25–27, 1978a
- Reiter, R., On closed world data bases, in: H. Gaillaire and J. Minker (Eds.), *Logic and Data Bases* (Plenum Press, New York, 1978b) 55–76.
- Reiter, R., Some results on default logics, Technical Report, Dept. of Computer Science, Univ. of British Columbia. Forthcoming.
- Roberts, R. B. and Goldstein, I., *The FRL Manual*, A.I. Memo No. 409, MIT, Cambridge, MA, 1977.
- Robinson, J. A., A machine oriented logic based on the resolution principle, *J.ACM* **12** (1965) 25–41.
- Rogers, H., *Theory of Recursive Functions and Effective Computability* (McGraw–Hill, New York, 1967).
- Roussel, P., *PROLOG, Manuel de Reference et d'Utilisation*, Group d'Intelligence Artificielle, U.E.R. de Marseille, France, 1975.
- Sandewall, E., An approach to the frame problem and its implementation, in: B. Meltzer and D. Michie (Eds.), *Machine Intelligence 7* (Edinburgh University Press, Edinburgh, 1972) 195–204.
- Waldinger, R., *Achieving Several Goals Simultaneously*, Artificial Intelligence Center Technical Note 107, Stanford Research Institute, Menlo Park, CA, 1975.

Received 22 August 1979