SOME REPRESENTATIONAL ISSUES IN DEFAULT REASONING

by

Raymond Reiter
and
Giovanni Criscuolo[1]

TECHNICAL REPORT 80-7

1980 AUGUST

DEPARTMENT OF COMPUTER SCIENCE
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA V6T 1W5

ABSTRACT

Although most commonly occurring default rules are normal when I viewed in isolation, they can interact with each other in ways that lead to the derivation of anomalous default assumptions. In order to deal with such anomalies it is necessary to re-represent these rules, in some cases by introducing non-normal defaults. The need to consider such potential interactions leads to a new concept of integrity, distinct from the conventional integrity issues of first order data bases.

The non-normal default rules required to deal with default interactions all have a common pattern. Default theories conforming to this pattern are considerably more complex than normal default theories. For example, they need not have extensions, and they lack the property of semi-monotonicity.

Current semantic network representations fail to reason correctly with defaults. However, when viewed as indexing schemes on logical formulae, networks can be seen to provide computationally feasible heuristics for the consistency checks required by default reasoning.

This paper was written with the financial support of the National Science and Engineering Research Council of Canada under grant A7642.

[1] Present address: Istituto di Fisica Teorica
University of Naples
Naples 80125
Italy

SOME REPRESENTATIONAL ISSUES IN DEFAULT REASONING

by

Raymond Reiter
and
Giovanni Criscuolo

1. INTRODUCTION

In an earlier paper [Reiter 1980a] one of us proposed a logic for default reasoning. The objective there was to provide a representation for, among other things, common sense facts of the form "Most A's are B's", and to articulate an appropriate logic to characterize correct reasoning using such facts.[1] One such form of reasoning is the derivation of default assumptions: Given a particular A , conclude that "This particular A is a B" . Because some A's are not B's this conclusion must be treated as a default assumption or belief about the world since subsequent observations in the world may yield that "This particular A is not a B". The derivation of the belief that "This particular A is a B" is a form of plausible reasoning which is typically required whenever conclusions must be drawn from incomplete information about a world.

It is important to note that not all senses of the word "most" lead to default assumptions. One can distinguish two such senses:

1. A purely statistical connotation, as in "Most voters prefer Carter." Here, "most" is being used exclusively in the sense of "the majority of".

---

[1] Other closely related work with much the same motivation is described in [McCarthy 1980], [McDermott 1980] and [McDermott and Doyle 1980].

This setting does not lead to default assumptions: given that Maureen is a voter one would not want to assume that Maureen prefers Carter. Default logic makes no attempt to represent or reason with such statistical facts.

2. A prototypical sense, as in "Most birds fly." There is a statistical connotation here - the majority of birds do fly - but there is also the sense that a characteristic of a prototypical or normal bird is being described. Given a bird Polly, one is prepared to assume that it flies unless one has reasons to the contrary.[1] It is towards such prototypical settings that default logic is addressed.

The concept of a prototypical situation is central to the frames proposal of [Minsky 1975] and is realized in such frame inspired knowledge representation languages as KRL [Bobrow and Winograd 1977] and FRL [Roberts and Goldstein 1977]. That these are alternative representations for some underlying logic has been convincingly argued in [Hayes 1977a]. Default logic presumes to provide a formalization of this underlying logic.

The approach taken by default logic is to distinguish between prototypical facts, such as "Typically mammals give birth to live young", and "hard" facts about the world such as "All dogs are mammals." The former are viewed as rules of inference, called <u>default rules</u>, which apply to the latter "hard" facts. The point of view is that the set of

---

[1] One way of distinguishing between these two senses of "most" is by replacing its setting using the word "typically". Thus, "Typically voters prefer Carter" sounds inappropriate, whereas "Typically birds fly" feels correct. In the rest of this paper we shall use "typically" whenever we are referring to a prototypical situation.

all "hard" facts will fail to completely specify the world - there will be gaps in our knowledge - and that the default rules serve to help fill in those gaps with plausible but not infallible conclusions. A default theory then is a pair (D,W) where D is a set of default rules applying to some world being modelled, and W is a set of "hard" facts about that world. Formally, W is a set of first order formulae while a typical default rule of D is denoted $\frac{\alpha(\vec{x}) \ : \ M\beta_1(\vec{x}),\ldots,M\beta_n(\vec{x})}{w(\vec{x})}$ where $\alpha(\vec{x})$, $\beta_1(\vec{x}),\ldots,\beta_n(\vec{x})$, $w(\vec{x})$ are all first order formulae whose free variables are among those of $\vec{x} = x_1,\ldots,x_m$. Intuitively, this default rule is interpreted as saying "For all individuals $x_1,\ldots,x_m$, if $\alpha(\vec{x})$ is believed and if each of $\beta_1(\vec{x}),\ldots,\beta_n(\vec{x})$ is consistent with our beliefs, then $w(\vec{x})$ may be believed." The set(s) of beliefs sanctioned by a default theory is precisely defined by a fixed point construction in [Reiter 1980a]. Any such set is called an extension for the default theory in question, and is interpreted as an acceptable set of beliefs that one may entertain about the world being represented.

It turns out that the general class of default theories is mathematically intractable. Accordingly, many of the results in [Reiter 1980a] (e.g. that extensions always exist, a proof theory, conditions for belief revision) were obtained only for the class of so-called normal default theories, namely theories all of whose defaults have the form $\frac{\alpha(\vec{x}) \ : \ Mw(\vec{x})}{w(\vec{x})}$. Such defaults are extremely common; for example "Typically dogs bark." : $\frac{DOG(x) \ : \ M\ BARK(x)}{BARK(x)}$

"Typically American adults own a car.":

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

Many more examples of such normal defaults are described in [Reiter 1980a]. Indeed, the claim was made there that all naturally occurring defaults are normal. Alas, this claim appears to be true only when interactions involving default rules are ignored. For normal default theories such interactions can lead to anomalous conclusions.

It is the purpose of this paper to describe a variety of settings in which interactions involving defaults are important, and to uniformly generalize the notion of a normal default theory so as to correctly treat these interactions. The resulting semi-normal default theories will then be seen to have some interesting properties: for example they need not have extensions, and they lack the semi-monotonicity property which all normal theories enjoy. We shall also see that the interactions introduced by default rules lead to a new concept of data base integrity, distinct from the integrity issues arising in first order data bases. A final objective of this paper is to analyze current network representations with respect to their ability to correctly reason with defaults. On this count such representations will be found deficient. However, when viewed as indexing schemes on logical formulae, networks will be seen to redeem themselves; they can provide computationally feasible heuristics for the consistency checks required by default reasoning.

2. INTERACTING NORMAL DEFAULTS

In this section we present a number of examples of default rules which, in isolation, are most naturally represented as normal defaults but whose interaction with other defaults or first order formulae leads to counterintuitive results. In each case we show how to "patch" the representation in order to restore the intended interpretation. The resulting "patches" all have a uniform character, which will lead us in Section 4.2 to introduce the notion of a semi-normal default theory.

2.1 "Typically" is not Necessarily Transitive

Consider:

"Typically A's are B's": $\frac{A(x) : MB(x)}{B(x)}$ (2.1)

"Typically B's are C's": $\frac{B(x) : MC(x)}{C(x)}$ (2.2)

These are both normal defaults. Default logic then admits the conclusion that "Typically A's are C's" in the following sense: If a is an individual for which A(a) is known or believed, and if ~B(a) and ~C(a) are not known or believed, then C(a) may be derived. In other words, normal default theories impose transitivity of "typically". But this need not be transitive, for example:

"Typically high school dropouts are adults."
"Typically adults are employed."
(2.3)

From these one would not want to conclude that "Typically high school dropouts are employed."[1] Transitivity must be blocked. This can be done in general by replacing the normal default (2.2) by the non-normal default

$$\frac{B(x) \ : \ M(\sim A(x) \wedge C(x))}{C(x)} \tag{2.4}$$

To see why this works, consider a prototypical individual a which is an A i.e. A(a) is given. By (2.1) B(a) can be derived. But B(a) cannot be used in conjunction with (2.4) to derive C(a) since the consistency condition $\sim A(a) \wedge C(a)$ of (2.4) is violated by the given A(a) . On the other hand, for a prototypical individual b which is a B (i.e. B(b) is given) (2.4) can be used to derive C(b) since presumably nothing is known about b's A-ness - we do not know that A(b) - so that the consistency condition of (2.4) is satisfied.

The introduction of non-normal defaults like (2.4) is a particularly unpleasant solution to the transitivity problem, for as we shall see in Section 4.2, the resulting non-normal default theories lack most of the desirable properties that normal theories enjoy. For example, they sometimes fail to have an extension, they lack semi-monotonicity, and their proof theory appears to be considerably more complex than that for normal theories. Accordingly, to the extent that it can be done, we would prefer to keep our representations "as normal as possible."

----

[1] Nor would we want to conclude that "Typically high school dropouts are not employed." Rather we would remain agnostic about the employment status of a typical high school dropout.

Fortunately transitivity can be blocked using normal defaults whenever it is the case that in addition to (2.1) and (2.2) we have "Typically B's are not A's". This is the case for example (2.3): "Typically adults are not high school dropouts". Under this circumstance, the following normal representation blocks transitivity:

$$\frac{A(x) : MB(x)}{B(x)} \qquad (2.5)$$

$$\frac{B(x) : M \sim A(x)}{\sim A(x)} \qquad (2.6)$$

$$\frac{B(x) \wedge \sim A(x) : MC(x)}{C(x)} \qquad (2.7)$$

Notice how, when given that $B(a)$, a simple back-chaining interpreter would establish the goal $C(a)$. Back-chaining into (2.7) yields the subgoal $B(a) \wedge \sim A(a)$. This splits into the subgoal $B(a)$, which is given and hence solved, and the subgoal $\sim A(a)$. This latter back-chains into (2.6) yielding the subgoal $B(a)$ which is solved. There remains only to verify the consistency requirements associated with the defaults (2.6) and (2.7) entering into the proof i.e. to verify that $\{C(a), \sim A(a)\}$ is consistent with all of the first order formulae in force. Such a back-chaining default reasoner is an incomplete realization of the complete proof procedure of [Reiter 1980a]. The reader might find it instructive to simulate this back-chaining interpreter for the case that $A(a)$ is given, in order to see how a derivation of $C(a)$ is prevented.

Notice also that the representation (2.5), (2.6) and (2.7) yields a very interesting prediction. Given an individual a which is simultaneously an instance of A and B , nothing can be concluded about its C-ness. This prediction is confirmed with respect to example (2.3): Given that John is both a high school dropout and an adult, we do not want to assume that John is employed. Notice that the non-normal representation (2.1) and (2.4) yields the same prediction. We shall have more to say about defaults with common instances of their prerequisites in Section 2.3.[1]

A somewhat different need for blocking transitivity arises when it is the case that "Typically A's are not C's" i.e. in addition to (2.1) and (2.2) we have

$$\frac{A(x) : M \sim C(x)}{\sim C(x)} \qquad (2.8)$$

For example,

"Typically university students are adults."
"Typically adults are employed."
"Typically university students are not employed."
(2.9)

Under these circumstances, consider a prototypical instance a of A . By (2.1) and (2.2) C(a) can be derived. But by (2.8) ~ C(a) can be derived. This means that the individual a gives rise to two different extensions for the fragment default theory (2.1), (2.2) and

---

[1] If $\frac{\alpha(\vec{x}) : M\beta_1(\vec{x}), \ldots, M\beta_n(\vec{x})}{w(\vec{x})}$ is a default rule then $\alpha(\vec{x})$ is its <u>prerequisite</u>.

(2.8). One of these extensions - the one containing $C(a)$ - is intuitively unacceptable; only the other extension - the one containing $\sim C(a)$ - is admissible. But a fundamental premise of default logic is that <u>any</u> extension provides an acceptable set of beliefs about a world. The problem then is to eliminate the extension containing $C(a)$. This can be done by replacing the normal default (2.2) by the non-normal (2.4), exactly as we did earlier in order to block the transitivity of "typically". Now, given $A(a)$, $B(a)$ can be derived from (2.1), and $\sim C(a)$ from (2.8). $C(a)$ cannot be derived using (2.4) since its consistency requirement is violated. On the other hand, given a prototypical instance b of B , $C(b)$ can be derived using (2.4).

Once again a non-normal default has been introduced, something we would prefer to avoid. As before, a normal representation can be found whenever it is the case that "Typically B's are not A's". This is the case for example (2.9): "Typically adults are not university students". Under this circumstance the following normal representation will do:

$$\frac{A(x)\ :\ MB(x)}{B(x)}$$

$$\frac{B(x)\ :\ M \sim A(x)}{\sim A(x)}$$

$$\frac{B(x) \wedge \sim A(x)\ :\ MC(x)}{C(x)}$$

$$\frac{A(x)\ :\ M \sim C(x)}{\sim C(x)}$$

Notice that this representation predicts that any individual which is simultaneously an instance of A and B will be an instance of not C , rather than an instance of C . This is the case for example (2.9): Given that Maureen is both a university student and an adult one wants to assume that Maureen is not employed.

Figure 2.1 summarizes and extends the various cases discussed in this section. The first three entries of this table are unproblematic cases which were not discussed, and are included only for completeness.

2.2 Interactions Between "All" and "Typically"

Phenomena closely related to those stemming from the non-transitivity of "typically" arise from interactions between normal defaults and certain universally quantified first order formulae. Consider

"All A's are B's". $$(x).A(x) \supset B(x) \qquad (2.10)$$

"Typically B's are C's". $$\frac{B(x) : MC(x)}{C(x)} \qquad (2.11)$$

Default logic forces the conclusion that "Typically A's are C's" in the sense that if a is a prototypical A then it will also be a C . But this conclusion is not always warranted, for example:

"All 21 year olds are adults."
"Typically adults are married." } (2.12)

Given that John is a 21 year old, we would not want to conclude that he is married. To block the unwarranted derivation, replace (2.11) by

$$\frac{B(x) : M(\sim A(x) \wedge C(x))}{C(x)} \qquad (2.13)$$

Figure 2.1

Typically A's are B's. Typically B's are C's.	Default Representation
No A is a C .	$(x) \,.\, A(x) \supset \sim C(x)$ $\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, MC(x)}{C(x)}$
All A's are C's.	$(x) \,.\, A(x) \supset C(x)$ $\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, MC(x)}{C(x)}$
Typically A's are C's.	$\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, MC(x)}{C(x)}$
It is not the case that A's are typically C's. Transitivity must be blocked.	$\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, M(\sim A(x) \wedge C(x))}{C(x)}$
Typically B's are not A's. It is not the case that A's are typically C's. Transitivity must be blocked.	$\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, M \sim A(x)}{\sim A(x)}$ $\frac{B(x) \wedge \sim A(x) \,:\, MC(x)}{C(x)}$
Typically A's are not C's .	$\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, M(\sim A(x) \wedge C(x))}{C(x)}$ $\frac{A(x) \,:\, M \sim C(x)}{\sim C(x)}$
Typically B's are not A's. Typically A's are not C's .	$\frac{A(x) \,:\, MB(x)}{B(x)}$ $\frac{B(x) \,:\, M \sim A(x)}{\sim A(x)}$ $\frac{B(x) \wedge \sim A(x) \,:\, MC(x)}{C(x)}$ $\frac{A(x) \,:\, M \sim C(x)}{\sim C(x)}$

As was the case in Section 2.1 the introduction of this non-normal default can be avoided whenever it is the case that "Typically B's are not A's"[1] by means of the representation (2.10) together with

$$\left.\begin{array}{c} \dfrac{B(x)\ :\ M \sim A(x)}{\sim A(x)} \\[2ex] \dfrac{B(x) \wedge \sim A(x)\ :\ MC(x)}{C(x)} \end{array}\right\} \quad (2.14)$$

Notice that this representation, as well as the representation (2.10) and (2.13) predicts that no conclusion is warranted about the C-ness of any given common instance of A and B .

A related problem arises when it is the case that "Typically A's are not C's" so that, in addition to (2.10) and (2.11) we have

$$\frac{A(x)\ :\ M \sim C(x)}{\sim C(x)} \qquad (2.15)$$

For example:

"All Québecois are Canadians."

"Typically Canadians are native English speakers."

"Typically Québecois are not native English speakers."

As in Section 2.1, a prototypical instance a of A will give rise to two extensions for the theory (2.10), (2.11) and (2.15), one containing C(a) ; the other containing ~ C(a) . To eliminate the

---

[1] Note that example (2.12) seems not to have this character. One is unlikely to include that "Typically adults are not 21 years old" in any representation of a world.

extension containing C(a) , replace (2.11) by (2.13).

As before, the introduction of the non-normal default (2.13) can be avoided whenever it is the case that "Typically B's are not A's", by means of the representation (2.10), (2.14) and (2.15).

Figure 2.2 summarizes the cases discussed in this section. The first three entries of this table are unproblematic cases which were not discussed, and are included only for completeness.

2.3 Conflicting Default Assumptions: Prerequisites with Common Instances

In this section we discuss the following pattern, in which a pair of defaults have contradictory consequents but whose prerequisites may share common instances[1]:

$$\left.\begin{array}{c} \dfrac{A(x) \ : \ M \sim C(x)}{\sim C(x)} \\[2em] \dfrac{B(x) \ : \ MC(x)}{C(x)} \end{array}\right\} \qquad (2.16)$$

The problem here is which default assumption (if any) should be made when given an instance a of both A and B i.e. should C(a) be assumed, or ~ C(a) or neither? Two cases have already been considered:
1. If it is the case that all A's are B's, then row 6 and possibly row 7 of Figure 2.2 provide representations; in both ~ C(a) is

---

1 If $\dfrac{\alpha(\vec{x}) \ : \ M\beta_1(\vec{x}), \ldots, M\beta_n(\vec{x})}{w(\vec{x})}$ is a default rule, then $\alpha(\vec{x})$ is its prerequisite and $w(\vec{x})$ its consequent.

Figure 2.2

All A's are B's. Typically B's are C's.	Default Representation
No A is a C.	$(x) . A(x) \supset \sim C(x)$ $(x) . A(x) \supset B(x)$ $\frac{B(x) : MC(x)}{C(x)}$
All A's are C's.	$(x) . A(x) \supset C(x)$ $(x) . A(x) \supset B(x)$ $\frac{B(x) : MC(x)}{C(x)}$
Typically A's are C's.	$(x) . A(x) \supset B(x)$ $\frac{B(x) : MC(x)}{C(x)}$
It is not the case that A's are typically C's Transitivity must be blocked.	$(x) . A(x) \supset B(x)$ $\frac{B(x) : M(\sim A(x) \wedge C(x))}{C(x)}$
Typically B's are not A's. It is not the case that A's are typically C's. Transitivity must be blocked.	$(x) . A(x) \supset B(x)$ $\frac{B(x) : M \sim A(x)}{\sim A(x)}$ $\frac{B(x) \wedge \sim A(x) : MC(x)}{C(x)}$
Typically A's are not C's.	$(x) . A(x) \supset B(x)$ $\frac{B(x) : M(\sim A(x) \wedge C(x))}{C(x)}$ $\frac{A(x) : M \sim C(x)}{\sim C(x)}$
Typically B's are not A's. Typically A's are not C's.	$(x) . A(x) \supset B(x)$ $\frac{B(x) : M \sim A(x)}{\sim A(x)}$ $\frac{B(x) \wedge \sim A(x) : MC(x)}{C(x)}$ $\frac{A(x) : M \sim C(x)}{\sim C(x)}$

derivable whenever A(a) and B(a) are simultaneously given.

2. If it is the case that "Typically A's are B's" then row 6 and possibly row 7 of Figure 2.1 provide representations in both of which $\sim C(a)$ is derivable given A(a) and B(a) .

The problematic setting is when there is no entailment relationship between A and B . For example:

$$\left.\begin{array}{l}\text{"Typically Republicans are not pacifists."}\\ \text{"Typically Quakers are pacifists."}\end{array}\right\} \quad (2.17)$$

Now, given that John is both a Quaker and a Republican, we intuitively want to make no assumptions about his warlike nature. This can be done in the general case by replacing the representation (2.16) by the non-normal defaults

$$\left.\begin{array}{c}\dfrac{A(x)\ :\ M(\sim B(x) \wedge \sim C(x))}{\sim C(x)}\\[2ex] \dfrac{B(x)\ :\ M(\sim A(x) \wedge C(x))}{C(x)}\end{array}\right\} \quad (2.18)$$

This representation admits that a typical A is not a C , a typical B is a C , but a typical A which is also a B leads to no conclusion.

When it is the case that "Typically A's are not B's" and "Typically B's are not A's" the non-normal defaults (2.18) can be replaced by the following normal ones:

$$\frac{A(x)\ :\ M \sim B(x)}{\sim B(x)}$$

$$\frac{B(x)\ :\ M \sim A(x)}{\sim A(x)}$$

$$\frac{A(x) \wedge \sim B(x) : M \sim C(x)}{\sim C(x)}$$

$$\frac{B(x) \wedge \sim A(x) : MC(x)}{C(x)}$$

This appears to be the case for example (2.17):

"Typically, Republicans are not Quakers."

"Typically, Quakers are not Republicans."

It is not always the case that the pattern (2.16) should lead to no default assumptions for common instances of A and B . Consider:

"Typically full time students are not employed."

"Typically adults are employed."

Suppose that John is an adult full time student. One would want to assume that he is not employed. So in general, given the setting (2.16) for which the default assumption $\sim C$ is preferred for common instances of A and B , use the following non-normal representation:

$$\frac{A(x) : M \sim C(x)}{\sim C(x)}$$

$$\frac{B(x) : M(\sim A(x) \wedge C(x))}{C(x)}$$

Whenever, in addition, it is the case that "Typically B's are not A's," use the following normal representation:

$$\frac{A(x) : M \sim C(x)}{\sim C(x)}$$

$$\frac{B(x)\ :\ M \sim A(x)}{\sim A(x)}$$

$$\frac{B(x) \wedge \sim A(x)\ :\ MC(x)}{C(x)}$$

3. DEFAULT INHERITANCE IN HIERARCHIES: NETWORK REPRESENTATIONS

We have focused, in Section 2, on certain fairly simple patterns of default rules. Our choice of these patterns was conditioned by their frequent occurrence in common sense reasoning, and by the fact that they are typical of the kinds of default knowledge which various "semantic" network schemes presume to represent and reason with. Most such networks are designed to exploit the natural hierarchical organization of much of our knowledge about the world and rely heavily for their inferencing power upon the inheritance of properties associated with a general class "down the hierarchy" to more restricted classes. Networks usually provide for defaults and their inheritance, although they do not all distinguish in their graphical notation between default rules and exception-free statements about the world.[1,2] In any event those systems which deal with defaults appear to rely exclusively on a shortest path heuristic, embedded in the network interpreter, for

---

[1] So that the representations often appear to be inconsistent. See [Winograd 1980]. Of course, once a proper semantics is defined for the network ([Schubert 1976], [Woods 1975]) the apparent inconsistency evaporates. Advocates of the need to reason from inconsistent information are, in part, confusing default rules with first order facts about a world.

[2] The SNePS system [Bechtel and Shapiro 1976] does make this distinction through the introduction of an "almost-all" "quantifier".

computing default inheritances in hierarchies [Shapiro 1978], [Winograd 1980]. To see what this device is and why it is deemed necessary, consider:

"Typically, students are full time."

$$\frac{\text{STUDENT}(x)\ :\ M\ \text{FULL-TIME}(x)}{\text{FULL-TIME}(x)}$$

"Typically, night students are not full time."

$$\frac{\text{NIGHT-STUDENT}(x)\ :\ M\ \sim \text{FULL-TIME}(x)}{\sim \text{FULL-TIME}(x)}$$

"All night students are students."

$$(x).\ \text{NIGHT-STUDENT}(x) \supset \text{STUDENT}(x)$$

A network representation for these facts might look something like that of Figure 3.1. (We have slightly modified the notation of [Shapiro 1978].) Now suppose that John is a night student. We want to conclude that he is not full time, not that he is full time. But what is to prevent a network interpreter from traversing the MEMBER and ISA link from John to NIGHT-STUDENT to STUDENT and thence via the default PROP link to FULL-TIME? Enter the shortest path heuristic. Basically this says that an individual, e.g. John, will inherit a property P provided there is a path from the node "John" to the node "P" and there is no shorter or equal length path from John to "not P". This is a slightly more precise statement of that in [Winograd 1975]:

"Any property true of a concept in the hierarchy is implicitly true of

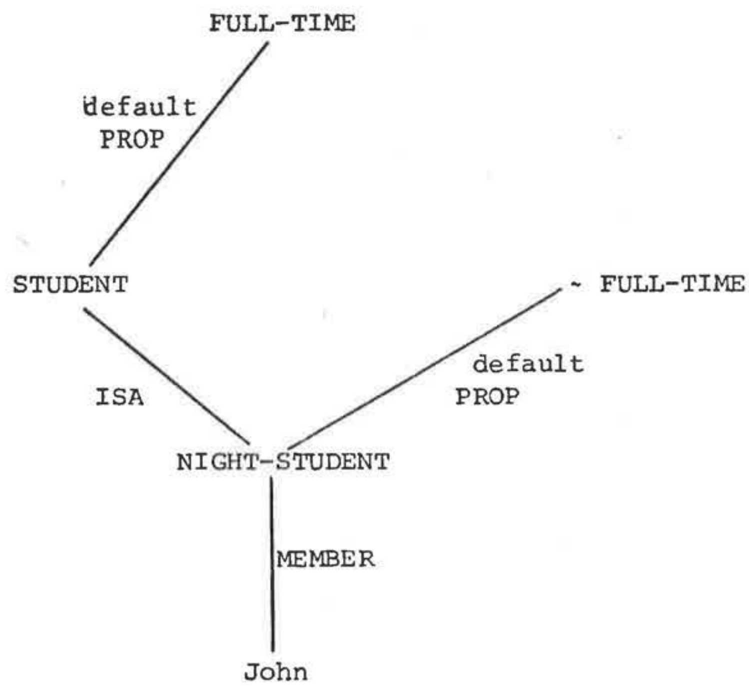
FULL-TIME
default
PROP
STUDENT
~ FULL-TIME
default
PROP
ISA
NIGHT-STUDENT
MEMBER
John

Figure 3.1

anything linked below it, unless explicitly contradicted at the lower level."

It is easy to see that this principle, as applied to Figure 3.1, will prevent the unwarranted assumption that John is full time.

Unfortunately, except in the simplest of cases, the shortest path heuristic is wrong. For example, consider a slightly embellished version of the Quaker-Republican defaults:

"Typically, Quakers are pacifists."
"Typically, Republicans are hawks."
"No hawk is a pacifist."

Suppose that John is a Quaker Republican. Then there is a path from "John" to "PACIFIST" as well as one from "John" to "~ PACIFIST" and the former path is shorter than the latter. The shortest path heuristic would thus predict that John is a pacifist whereas intuitively no default assumption is warranted.

Despite our criticism of the shortest path heuristic, we nevertheless feel that there is a profound implementation principle lurking here. One of the most serious computational difficulties afflicting default logic is the requirement that one test for the consistency of all of the default assumptions entering into a derivation. For example:

"Typically birds fly except for penguins, ostriches, oil covered birds, dead birds, etc. etc."

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

$$(x)\ .\ PENGUIN(x) \supset \sim FLY(x)$$

$$(x)\ .\ OSTRICH(x) \supset \sim FLY(x)$$

etc.

Now suppose given BIRD (tweety), and nothing else about tweety. Then FLY (tweety) can be derived provided that FLY (tweety) is consistent with all of the first order facts in the data base. One way of establishing consistency is by failing to derive a contradiction from all of the consequences of the formula FLY (tweety). Of course, the detection of an inconsistent set of formulae is undecidable in general, but let's try anyway. From FLY (tweety) one can derive ~ PENGUIN (tweety), ~ OSTRICH (tweety), ~DEAD-BIRD (tweety) etc. etc. So with this method of performing the consistency check, one must consider all of the possible exceptions to the default rule about flying birds! Since the exceptions to flight are legion we are faced with a potentially overwhelming computation. Ideally, we do not want even to entertain the possibility of an exception unless the given facts naturally compel us to do so. The only way of testing consistency which avoids "conscious" consideration of all of the exceptions to flight is to begin with the given fact BIRD (tweety), and using only the first order facts in the data base derive all consequences of this; if ~ FLY (tweety) is not one of these consequences then consistency is guaranteed.

Now consider Figure 3.2 which is a network representation of this same setting. We can tell at a glance that FLY (tweety) is consistent with our knowledge: ~ FLY (tweety) is not derivable because there is no

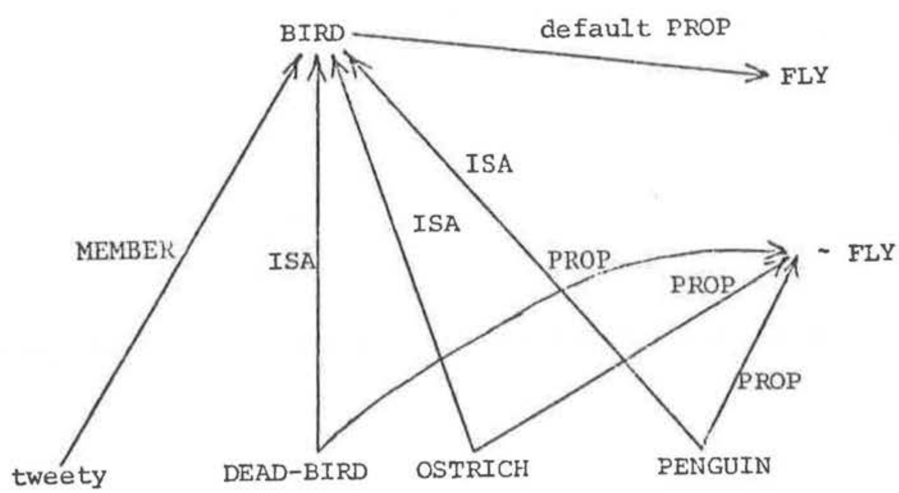
BIRD
default PROP
FLY
ISA
ISA
MEMBER
ISA
PROP
PROP
~ FLY
PROP
tweety
DEAD-BIRD
OSTRICH
PENGUIN

FIGURE 3.2

directed path from "tweety" to "~ FLY". Potential derivation chains in the logical representation are explicit as directed paths in the network representation. Now the consistency check which began with BIRD (tweety) and derived all consequences of this corresponds in the network to an exploration of all paths from "tweety". If there is no such path to "~ FLY" then the consistency of FLY(tweety) is assured. Now computationally the exploration of all directed paths beginning at node "tweety" might not appear very promising since the search will get mired in all of the links in that part of the hierarchy lying above the node "BIRD". But recall that we are testing consistency only with respect to all of the first order facts about the world, not the default rules. Hence no path containing a default PROP link need be considered, and most network links are of this kind. Moreover, hierarchies tend to be shallow. Hence the search for a path from "tweety" to "~ FLY" in the hierarchy above "BIRD" appears feasible. It follows that a good strategy is to perform a unidirectional search from "tweety"; if "~ FLY" is not encountered, then the default assumption FLY (tweety) is acceptable. This strategy has the computationally important consequence that the myriad possible exceptions to flight are never "consciously" entertained.

Now what is really going on here? The answer is apparent from the observation that a semantic network reflects a particular choice of an indexing scheme on formulae.[1] The indexing scheme is such that whenever an entailment relation logically holds between two nodes, then those

---

[1] The fact that networks are notational variants of logical formulae is by now a truism in Artificial Intelligence circles. See [Hayes 1977b], [Schubert 1976].

nodes are connected by a directed path; network paths correspond to derivation chains in the underlying logical representation. The nonexistence of a path in Figure 3.2 from "tweety" to "~ FLY" guarantees that ~ FLY (tweety) cannot be derived i.e. that FLY (tweety) is consistent with the first order formulae of the data base.

Now there exist far more sophisticated indexing schemes on formulae than any provided in the literature on semantic networks. See, for example, [Kowalski 1974], or the indexing on clauses in PROLOG [Clark and McCabe 1979]. Normally such schemes are used to improve the efficiency of theorem provers although they can be used for the construction of plans in deductive search [Kellog et al. 1978]. The discussion of paths in networks and their relationship to consistency suggests another use of indices on formulae: the path structure of the index scheme can provide a powerful and computationally feasible heuristic for the consistency checks required in default reasoning. An example of such an heuristic is the following, with reference to our bird example:

If node "~ FLY" cannot be found within a sufficiently large radius r of the node "tweety" (i.e. if no directed path of length r or less from "tweety" to "~ FLY" exists in the index structure) then it is a good bet that FLY (tweety) is consistent with the given first order data base.

It seems to us that an heuristic of this kind is precisely the sort of resource limited computation required for common sense reasoning [Winograd 1980]. Moreover, there is a very good theoretical justification for appealing to a resource limitation in this setting; consistency is not even a semi-decidable property of first order theories so that some sort of heuristic must be applied. What is interesting about this formal analysis is that the nature of, and reasons for, at least one form of resource limited computation can be theoretically articulated.

Notice also that this consistency heuristic is simply a path finding procedure for directed graphs. No deductions are performed. Rather, the non existence of a sufficiently long path of a certain form strongly suggests the consistency of some set of formulae.

4. DISCUSSION

In this section we discuss some issues raised by the results of the previous sections. Specifically, we address the question of data base integrity arising from default interactions, as well as some of the formal problems associated with the non-normal default rules introduced to correctly represent these interactions.

4.1 Integrity of Default Theories

A very nice feature of first order logic as an Artificial Intelligence representation language is the extensibility of any theory expressed in this language. That is, provided that some axiomatization of a world has that world as a model (so that the axiomatization faithfully represents certain aspects of that world) then the result of adding a new axiom about the world is still a faithful representation. It is true that specialized deduction mechanisms may be sensitive to such updates (e.g. adding a new "theorem" to a PLANNER-like data base); but semantically there is no problem. Unfortunately, as we have seen, default theories lack this semantic extensibility; the addition of a new default rule may create interactions leading to unwarranted conclusions, even though in isolation this rule appears perfectly correct.

This observation leads to a new concept of data base integrity, one with quite a different character than the integrity issues arising in data base management systems [Hammer and McLeod 1975] or in first order data bases [Nicolas and Yazdanian 1978, Reiter 1980b]. For such systems an integrity constraint specifies some invariant property which every state of the data base must satisfy. For example, a typical integrity constraint might specify that an employee's age must lie in the range 16 to 99 years. Any attempt to update the data base with an employee age of 100 would violate this constraint. Formally one can say that a data base satisfies some set of integrity constraints if the data base is logically consistent with the constraints. The rôle of integrity constraints is to restrict the class of models of a data base to include as a model the particular world being represented. Now the objective of the default representations of Section 2 had precisely this character; we sought representations which would rule out unwarranted default assumptions so as to guarantee a faithful representation of real world common sense reasoning. But notice that there was no notion of an integrity constraint with which the representation was to be consistent.

Indeed, consistency of the representation cannot be an issue at all since any default theory will be consistent provided its first order facts are [Reiter 1980a, Corollary 2.2]. It follows that, while there is an integrity issue lurking here, it has a different nature than that of classical data base theory.

We are thus led to the need for some form of integrity maintenance mechanism as an aid in the design of large default data bases. The natural initial data base design would involve representing all default rules as normal defaults, thereby ignoring those potential interactions of the kind analyzed in Section 2. An integrity maintenance system would then seek out possible sources of integrity violations and query the user as to the appropriate default assumptions to be made in this setting. Once the correct interpretation has been determined, the system would appropriately re-represent the offending normal default rules. For example, when confronted with a pair of default rules of the form (2.16), the system would first attempt to prove that A and B can have no common instance i.e. that $W \cup \{(Ex).A(x) \wedge B(x)\}$ is inconsistent, where W is the set of first order facts. If so, this pair of defaults can lead to no integrity violation. Otherwise the system would ask whether a common instance of A and B is typically a C, a $\sim C$, or neither, and depending on the response would suitably re-represent the pair (2.16), if necessary by non-normal default rules.

4.2 Semi-Normal Default Theories

In Section 2 we had occasion to introduce certain non-normal default rules in order, for example, to block the transitivity of "typically". Inspection of the representations of that section will reveal that all such non-normal default rules share a common pattern; they all have the form $\frac{A(x) : M(\sim B(x) \wedge C(x))}{C(x)}$. Accordingly, it is natural to define a default rule to be <u>semi-normal</u> iff it has the form $\frac{\alpha(\vec{x}) : M(\beta(\vec{x}) \wedge w(\vec{x}))}{w(\vec{x})}$ where α, β and w are formulae of first order logic with free variables among $\vec{x} = x_1, \ldots, x_m$. A default theory is <u>semi-normal</u> iff all of its default rules are semi-normal. Normal default rules are a special case of semi-normal, in which $\beta(\vec{x})$ is the identically true proposition.

[Reiter 1980a] investigates the properties of normal default theories. Among the results obtained there are the following:

1. Every normal theory has an extension.
2. Normal theories are <u>semi-monotonic</u> i.e. if D_1 and D_2 are sets of normal default rules and if E_1 is an extension for the theory (D_1, W) , then the theory $(D_1 \cup D_2, W)$ has an extension E_2 such that $E_1 \subseteq E_2$.

One consequence of semi-monotonicity is that one can continue to maintain one's old beliefs whenever a normal theory is updated with new normal defaults. Another is a reasonably clean proof theory.

Unfortunately, semi-normal default theories enjoy none of these nice properties. For example, the following theory has no extension:

$$\frac{: M(A \wedge B)}{B} \qquad B \supset \sim A$$

To see that semi-monotonicity may fail to hold for semi-normal theories consider the theory

$$\frac{: M(A \wedge B)}{B}$$

This has unique extension Th({B}) where, in general, Th(S) is the closure of the set of formulae S under first order theoremhood. If the new default rule $\frac{: M \sim A}{\sim A}$ is added to this theory a new theory is obtained with unique extension Th({~ A}) and this does not contain Th({B}) .

Most of the formal properties of semi-normal default theories remain unexplored. Two problems in particular require solutions: Under what conditions are extensions guaranteed to exist, and what is an appropriate proof theory?

5. CONCLUSIONS

Default theories are complicated. Unlike theories represented in first order logic, default theories lack extensibility. Whenever a new default rule is to be added to a representation its potential interactions with the other default rules must be analyzed. This can lead to a re-representation of some of these defaults in order to block certain unwarranted derivations. All of which leads to a new concept of data base integrity, distinct from the integrity issues arising in first order data bases. These observations also suggest the need for a default integrity maintenance system as a tool for aiding in the design of large default data bases. Such a system would seek out potentially interacting defaults during the data base design phase and query the designer about the consequences of these interactions.

Default theories are computationally intractable in principle because of the consistency checks required by their proof methods. Semantic networks provide an indexing scheme on first order formulae, but many other schemes are possible. An important role of indexing is the provision of an efficient heuristic for consistency checking without the need to perform deductions. Such consistency checks are prime examples of the kind of resource limited computations required in common sense reasoning.

Semi-normal default theories are complicated. They have none of the nice properties that make normal theories so appealing. Most of their formal properties are totally unexplored. At the very least a proof theory is needed, as well as conditions under which extensions are guaranteed to exist.

Acknowledgements

Many of the results and examples of this paper were inspired by a conversation that one of us (R.R.) had with N. Sridharan one afternoon in Vancouver. It was he who first pointed out to us the need for non normal defaults in common sense reasoning.

This work was done with the financial assistance of the National Science and Engineering Research Council of Canada grant A7642. G. Criscuolo was sponsored by the Council for a six month research position with the Department of Computer Science at the University of British Columbia.

REFERENCES

Bechtel, R. and Shapiro, S.C. (1976), A logic for semantic networks, Dept. of Computer Science, Indiana University, Technical Report No. 47.

Bobrow, D.G. and Winograd, T. (1977), An overview of KRL-0, a knowledge representation language, Cognitive Science 1 (1) (January).

Clark, K.L. and McCabe, F.G. (1979), IC-PROLOG reference manual, CCD Report 79/7, Imperial College, London.

Hammer, M.M. and McLeod, D.J. (1975), Semantic integrity in a relational data base system, PROC. of the Int. Conf. on Very Large Data Bases, Framington, Mass., Sept., 25-47.

Hayes, P.J. (1977a), The logic of frames, Technical Report, Dept. of Computer Science, University of Essex, Colchester, UK.

Hayes, P.J. (1977b), In defence of logic, Proc. IJCAI-77, MIT, Cambridge, Mass., Aug. 22-25, 559-565.

Kellog, C., Klahr, P. and Travis, L. (1978), Deductive planning and pathfinding for relational data bases, in: H. Gallaire and J. Minker (Eds.), Logic and Data Bases (Plenum Press, New York).

Kowalski, R.A. (1974), A proof procedure using connection graphs, J. ACM 22 (October), 572-595.

McCarthy, J. (1980), Circumscription - a form of non-monotonic reasoning, Artificial Intelligence 13, 27-39.

McDermott, D. (1980), Non-monotonic logic II, Research Report No. 174, Dept. of Computer Science, Yale University, New Haven, Conn.

McDermott, D. and Doyle, J. (1980), Non-monotonic logic I, Artificial Intelligence 13. 41-72.

Minsky, M. (1975), A framework for representing knowledge, in: P. Winston (Ed.), The Psychology of Computer Vision (McGraw-Hill, New York).

Nicolas, J.M. and Yazdanian, K. (1978), Integrity checking in deductive data bases, in: H. Gallaire and J. Minker (Eds.) Logic and Data Bases (Plenum Press, New York).

Reiter, R. (1980a), A logic for default reasoning, Artificial Intelligence 13, 81-132.

Reiter, R. (1980b), Data bases: a logical perspective, ACM SIGART, SIGMOD, SIGPLAN Proc. Workshop on Data Abstraction, Data Bases and Conceptual Modelling, Pingree Park, Colorado, June 23-26, to appear.

Roberts, R.B. and Goldstein, I. (1977), The FRL Manual, A.I. Memo No. 409, MIT, Cambridge, MA.

Schubert, L.K. (1976), Extending the expressive power of semantic networks, Artificial Intelligence 7 (2), 163-198.

Shapiro, S.C. (1978), Path-based and node-based inference in semantic networks, PROC. TINLAP II, Urbana, Ill., July 25-27, 219-225.

Winograd, T. (1975), Frame representations and the declarative/ procedural controversy, in: D.G. Bobrow and A. Collins (Eds.), Representation and Understanding, (Academic Press, New York).

Winograd, T. (1980), Extended inference modes in reasoning by computer systems, Artificial Intelligence 13, 5-26.

Woods, W.A. (1975), What's in a link? in: D.G. Bobrow and A. Collins (Eds.), Representation and Understanding (Academic Press, New York).