Tractable Theories of Multiple Defeasible Inheritance In Ordinary Nonmonotonic Logics

Brian A. Haugh
Martin Marietta Laboratories
1450 South Rolling Road
Baltimore, Maryland 21227

Abstract

A suggestion by John McCarthy for general formulations of multiple defeasible inheritance in ordinary nonmonotonic logic is examined and found to suffer from a variety of technical problems, including 1) its restriction to object/class/property networks, 2) unintuitive results in "Nixon diamond"-type networks, 3) unnecessary closed-world assumptions, and 4) susceptibility to unintended models when generalized. A family of theories is presented that substantially revises McCarthy's formulation to avoid these problems and restrictions. Finally, an inference control strategy for computing the theory is identified whose tractability is ensured by a variety of techniques including incremental computation of abnormalities and truth maintenance.

1. Introduction

Marvin Minsky's challenge to proponents of logic-based reasoning in AI to formalize the inheritance of prototypical properties and their exceptions [Minsky, 1975] has resulted in a whole new field of formalized nonmonotonic reasoning including general purpose nonmonotonic logics [McCarthy, 1980; Reiter, 1980; McDermott and Doyle, 1980], applications of such logics to inheritance hierarchy problems [McCarthy, 1986; Etherington, 1987; Sandewall, 1986], and special formalisms just for inheritance reasoning [Touretzky, 1986; Horty et al., 1987]. While these formulations of inheritance have demonstrated the adequacy of formal systems for such commonsense reasoning problems, each suffers from some substantial deficiency, e.g., lack of computationally tractable implementation techniques, absence of any general formalization procedure, limited expressive power, or use of nonstandard formalisms that are difficult to extend or modify. In this paper, we develop essential revisions to a suggestion by McCarthy [1986], creating a new family of logical theories for formalizing general inheritance reasoning that suffers none of these deficiencies.

2. McCarthy's Formulation

2.1. Object/Class/Property Hierarchies

McCarthy [1986] has developed two fairly general methods for representing multiple inheritances with exceptions — a technique using prioritized circumscription (pp. 105-107) and a "class-level" approach using ordinary circumscription that reifies classes and properties (pp. 99-100). Our work has focused on the latter approach because it has an appealing simplicity, and because no general procedure for translating inheritance networks into the prioritized approach has yet appeared.

McCarthy reifies "classes"[1] and properties of objects by assigning first-order variables and constants to them. Inheritance relations between classes are expressed by wffs of the form $c1 \leq c2$, stating that class $c1$ ordinarily inherits from class $c2$, while *in(x, c)* asserts that an object x is a member of a class c. Default properties are expressed by wffs of the form *ordinarily(c, p)*, meaning that objects in class c ordinarily have property p, while *ap(P, x)* states that a predicate P applies to an object x. An abnormality predicate *ab(aspect1(c1,c2,p))* expresses the abnormality of members of class c1 with respect to inheriting property p from class $c2$. Default inheritance of properties by classes is expressed by:

M1. $[\text{ordinarily}(c2,p)\ \&\ c1 \leq c2\ \&\ \neg\, \text{ab}(\text{aspect1}(c1,c2,p))] \supset \text{ordinarily}(c1, p)$.

Cancellation of such inheritance is formulated by:

M2. $[c1 \leq c2\ \&\ c2 \leq c3\ \&\ \text{ordinarily}(c2, \text{not}(p))] \supset \text{ab}(\text{aspect1}(c1,c3,p))$.

Transitivity of inheritance of class membership is asserted by:

M3. $[c1 \leq c2\ \&\ c2 \leq c3] \supset c1 \leq c3$.

Axiom M3 entails the fundamental limitation of McCarthy's theory: class membership relations are not defeasible, i.e., they cannot be cancelled.[2] We refer to McCarthy's approach as an "object/class/property inheritance system," adapting Touretzky's classification scheme [Touretzky, 1986].

While McCarthy's treatment of inheritance of properties by classes provides foundations adequate for a broad range of inheritance systems, it cannot represent the cancellations of inter-class relations that are found in many more general systems (e.g., [Touretzky, 1986; Horty et al., 1987]).[3]

[1] He observes that these classes are not extensional. We may conceive of them either as intensional classes or as names of the predicates which pick out the members of these classes (taking a meta-level view).

[2] I thank Jeff Horty for pointing this out to me.

[3] We should note that McCarthy presented this approach merely as a promising possibility, not as a comprehensive solution.

2.2. Problems

2.2.1. A Natural Extension

The most natural extension of McCarthy's formalism to handle exceptions to class membership inheritance treats them in the same manner that he treats exceptions to property inheritance, by replacing his axiom M3 as follows:

M3a. $[c1 \leq c2 \ \& \ c2 \leq c3 \ \& \ \neg ab(aspect1(c1,c2,c3))]$
$\supset c1 \leq c3$

M3b. $[c1 \leq c2 \ \& \ c2 \leq c3 \ \& \ c2 \leq not(c4)]$
$\supset ab(aspect1(c1,c3,c4))$

This leads to the intended results in simple class inheritance cancellation cases, such as that shown in Figure 1, although it will admit unintended models in slightly more complex networks if we simply minimize abnormalities (as McCarthy does).

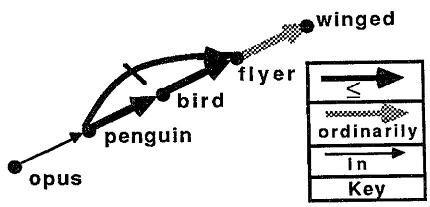

Figure 1. Simple class cancellation

2.2.2. Unintended Models With Gratuitous Links

In some inheritance networks, there will be models of this extended theory which are minimal in abnormalities but achieve that condition by admitting entirely new direct links that cancel the inheritance of intended abnormalities. For example, the network of Figure 2 will have minimal models with a gratuitous explicit link $A \leq not(C)$, which creates the unintended abnormality *ab(aspect1(A, B, C))* while blocking the intended inheritance of the dual abnormalities *ab(aspect1(A, D, F))* and *ab(aspect1(A, D, E))*. Thus, simply minimizing abnormalities will block many of the intended results (e.g., $A \leq C$) in such cases.

This problem is characteristic of the general scheme of directly minimizing abnormalities in abnormality-based meta-level general inheritance systems and is independent of our particular formulation of the axioms. When abnormalities can be inherited, and that inheritance can be cancelled, there will be models of many networks that will have fewer abnormalities if an unexplained cancellation of inheritance holds. Thus, we conclude that simple abnormality minimization will not provide the intended results in <u>any</u> such theories.

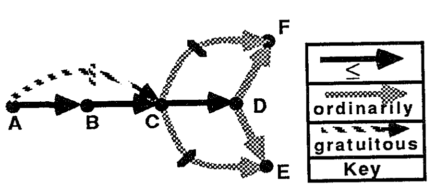

Figure 2. Unintended models

2.2.3. Excessive Closed-World Assumptions

McCarthy's approach of simply minimizing abnormalities cannot escape a wide variety of closed-world assumptions which violate common sense in many cases. Minimizing the abnormalities of individual objects, for example, entails that no objects with any abnormalities (e.g., penguins) exist unless they can be proven to exist. This pervasive problem arises in McCarthy's object-level theories of inheritance [McCarthy, 1986] as well as in his meta-level theories, and in all other previous logic-based object-level formulations (e.g., [Lifschitz, 1985; Etherington, 1987]).

2.2.4. Unintuitive Results In "Nixon Diamonds"

McCarthy's theory encounters a variety of difficulties when applied to certain kinds of inheritance networks which I call "Nixon diamonds," after the original example of this type developed by Reiter and illustrated by Figure 3. More general "Nixon diamonds" consist of pairs of arbitrary length, multi-link paths between two nodes, where the final links are contrary — also referred to as "conflicting multi-link paths." McCarthy's theory properly handles the original Nixon diamond, in which "Nixon" refers to an individual, provided the implicit axiom $ap(p,x) \supset \neg ap(not(p),x)$ is assumed. Difficulties arise, however, in "generic Nixon diamonds," in which the root (*Nixon*) is a class (e.g., Nixon's family) instead of an individual, and in "extended Nixon diamonds," where the multi-link paths are longer than two links (as in Figure 4).

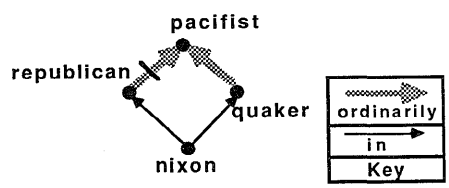

Figure 3. Original Nixon diamond

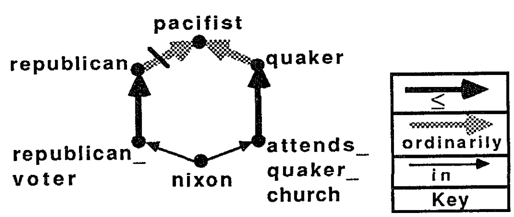

Figure 4. Extended Nixon diamond

In extended Nixon diamonds, McCarthy's theory entails that some abnormality holds, blocking one of the conflicting paths (as expected), but will be indifferent between all the possible link cancellations. Commonsense reasoning, however, preferring the least disruption to its default beliefs, tends to conclude that only the last links in conflicting multi-link paths are suspect, allowing the intermediate conclusions to stand.

3. Excluding Unintended Models

3.1. Alternative Methods

From the problems with unintended models that we identified for the natural extension of McCarthy's theory, it is apparent that common sense does not prefer a simple minimization of his abnormalities, but insists that no unexplained abnormalities should be admitted in place of expected ones. Commonsense use of default relations assumes that there are no abnormalities interfering with prima facie default conclusions unless they can be shown to follow from explicitly known relations using basic rules of default cancellation. Once this general principle is recognized, its formulation proceeds quite naturally by definition of an abnormality predicate *ab* in terms of the rules that generate abnormalities. Then, minimizing the explicit knowledge (direct links) required by those rules to generate abnormalities will restrict all such abnormalities appropriately. This solution comes at the prima facie cost of a general closed-world assumption that entails the falsehood of any general relations that are not provable from such theories. We will show, however, that such assumptions may be fully relaxed while retaining their benefits in excluding just the unintended abnormalities.

The other major alternative for avoiding unintended models within this type of theory involves minimizing what I call "potential abnormalities," i.e., those that would hold if there were no other anormalities blocking them. We do not present our theory for minimizing potential abnormalities here because it is more complex than minimizing direct links, and offers no clear advantages for inheritance reasoning.

3.2. Comparison to Temporal Reasoning

The two identified alternatives for eliminating unintended models of general inheritance theories are examples of general techniques of nonmonotonic reasoning that have been applied previously in temporal-causal reasoning to exclude unintended models identified for temporal persistence theories [Hanks & McDermott, 1986]. Minimizing types of causal relations was used by Lifschitz [Lifschitz, 1987] to eliminate models with unexplained changes, just as minimizing types of explicit inheritance relations here eliminates unexplained inheritance cancellations. Minimizing "potential causes" was used by Haugh [Haugh, 1987] to eliminate spurious particular potential causes of change (which would be actual causes if their preconditions held), just as minimizing potential abnormalities in inheritance theories will exclude spurious inheritance cancellations. Thus, our new results here suggest a broad potential for application of these techniques to other nonmonotonic reasoning issues.

4. Closed-World Inheritance

4.1. General Notation

To enable minimization of explicit relations in our closed-world theories, we distinguish them from other, derived, relations by representing them with distinctive predicates. Explicit default network relations (or links) are of the form *isa_x(x,p)*, stating that *x*'s are normally *p*'s, where *x* can refer to either an individual (or object-level term) or a class (or object-level predicate), and *p* refers only to either a class or property. General network relations, both explicit and derived, can be represented by *isa(x,p)*.

A sorted logic is used in which upper case A, B, C, D, A1, B1, ... are variables referring to individual objects, while upper case M, N, O, P, Q, R, S, T, M1, N1, ... are variables for classes/predicates, and upper case letters from the end of the alphabet — U, V, W, X, Y, Z, U1, V1, ... are variables ranging over objects and classes. Corresponding lower case terms are used for constants in the same categories, along with other mnemonic lower case constant names (e.g., *elephant*) whose category should be obvious in context.

To reason about exceptions to default links, we use an abnormality predicate modeled after McCarthy's use of a similar predicate, although our syntax is somewhat simplified. Where McCarthy uses a predication of the form *ab(aspect1(X,P,Q))* [McCarthy, 1986] to represent abnormalities, we use *ab(X,P,Q)* to mean that object/class *X* is abnormal with respect to inheriting any existing default relation between P and Q. For example, *ab(royal-elephant, elephant, gray)* asserts that royal-elephants are abnormal with respect to inheriting any *isa* relation between elephant and gray, or, less formally, royal-elephants are abnormal elephants with respect to being gray. While McCarthy uses different aspects (*aspect1* and *aspect2*) to distinguish between particular and generic abnormalities, we allow these distinctions to be determined by the sorts of terms appearing in *ab* predications.

4.2. Core Inheritance Axioms

4.2.1. Network Relations

Using the notation just presented, and axioms for generating derived relations from explicitly asserted ones, we can formulate a broad range of inheritance theories. We identify a general family of theories which share three core axioms and a simple minimization technique. The first axiom defines all of the default relations derivable from a network as:

A1. $\mathrm{isa}(X,Q) \equiv [\mathrm{isa\_x}(X,Q) \vee (\exists P)[\mathrm{isa}(X,P) \ \& \ \mathrm{isa\_x}(P,Q) \ \& \ \neg \mathrm{ab}(X,P,Q)]]$

which states that *X*'s are (normally) *Q*'s if and only if either there is an explicit network link asserting this, or there is an intermediate node *P* in the network such that X's are (normally) *P*'s, there is an explicit link asserting *P*'s to be *Q*'s (normally), and *X*'s are not abnormal with respect to *P*'s being *Q*'s. The abnormalities referred to are restricted to four primitive types:

A2. $\mathrm{ab}(X,P,Q) \equiv [\mathrm{ab\_d}(X,P,Q) \vee \mathrm{ab\_i}(X,P,Q) \vee \mathrm{ab\_c}(X,P,Q) \vee \mathrm{ab\_x}(X,P,Q)]$.

Explicit abnormalities (*ab_x*) are a type of cancellation link that are explicitly asserted, while the other abnormalities are derived from conflicting *isa_x* relations.

4.2.2. Direct Abnormalities

Direct abnormalities are created by the direct override by a single explicit link of a prima facie multi-link path in a network. The example network of Figure 1 illustrates the direct abnormality created by the cancellation of the path from *penguin* through *bird* to *flyer* by the direct contrary link from *penguin* to *not(flyer)*.[4]

Our third core axiom characterizes the general conditions under which direct abnormalities arise as:

A3. ab_d(X,P,Q) ≡
[isa(X,P) & isa_x(P,Q) & isa_x(X,not(Q))].

4.2.3. Inherited Abnormalities

Our other types of abnormalities admit of a broad range of alternative interpretations, none of which has achieved any consensus as yet. Here, we present the simplest version of inherited abnormalities. It allows what Touretzky calls "off-path preemptions" [Touretzky et al., 1987], in accord with the theories in [Sandewall, 1986] and [Horty et al., 1987], and is formulated as:

A4. ab_i(X,Q,R) ≡ (∃P)[isa(X,P) &
[ab_d(P,Q,R) v ab_c(P,Q,R) v ab_x(P,Q,R)]].

Such inherited abnormalities exist for every descendant X of a node P that has some direct or conflicting abnormality (e.g., *ab_d(P,Q,R)*). An example is illustrated by Figure 1, wherein *opus* inherits an abnormality with respect to *birds* being *flyers* from *penguin.*

Under this conception of inherited abnormality, an abnormality requires only a path (*isa(X,P)*) from the inheriting node (X) to the original abnormality node (P) to be inherited. One plausible variation of this would require every inherited abnormality *ab_i(X,Q,R)* to have a path (*path(X,P,Q)*) all the way from the inheriting node X, through the original abnormality node P, to the base node Q of the conflict link. This stricter conception could be expressed by only a minor variation of our axiom A4:

A4′. ab_i(X,Q,R) ≡ (∃p)[path(X,P,Q) &
(ab_d(P,Q,R) v ab_c(P,Q,R) v ab_x(P,Q,R)]

where

A4″. path(X,P,Q) ≡ [isa(X,P) &
[[isa_x(P,Q) & ¬ab(X,P,Q)] v
(∃R)[path(X,P,R) & isa_x(R,Q) & ¬ab(X,R,Q)]]].

This variation appears to correspond to the treatment in [Horty et al., 1987] when combined with ambiguity-blocking skepticism in conflicting paths as described below. Our intuitions in discriminating examples examined thus far favor our initial formulation, although further investigations are indicated before this issue can be considered settled.

4.2.4. Conflicting Path Abnormalities

A variety of incompatible alternatives for handling conflicting multi-link paths (or Nixon diamonds) have been proposed in the literature, e.g., [Touretzky et al., 1987; Horty et al., 1987]. Approaches have been categorized as either "skeptical" or "credulous" with regard to their willingness to draw conclusions in these cases. Credulous theories insist that one of the two conflicting terminal links must apply to the root node in such cases (e.g., *isa(nixon, pacifist)*, *isa(nixon, not(pacifist))* for Figures 3 and 4), while skeptical theories support no such conclusions (i.e., the conflicting links cannot be used to make any conclusions about the base node).

4.2.4.1. Skeptical Theories

Skeptical approaches, as distinguished in [Horty et al., 1987], possess the considerable computational advantage of having unique extensions, which obviates any need to examine alternative extensions during derivations. Skeptical theories can be formalized within abnormality theories such as ours by asserting the root of a general diamond to be abnormal with respect to both of the final links to its top, i.e.,

Ambiguity-Blocking Abnormality:

A5. ab_c(X,P,Q) ≡
(∃R)[isa(X,P) & isa_x(P,Q) &
isa(X,R) & isa_x(R,not(Q)) &
¬ab_dix(X,P,Q) & ¬ab_dix(X,R,not(Q))]

where

A6. ab_dix(X,P,Q) ≡
[ab_d(X,P,Q) v ab_x(X,P,Q) v
(∃S)[isa(X,S) & [ab_d(S,P,Q) v ab_x(S,P,Q)]].[5]

This first set of conflicting-path abnormality axioms accords with Horty's interesting results for "nested Nixon diamonds," as illustrated by Figure 5. In this example, these theories block any conclusions about the relations between X and R in the nested diamond, thereby blocking any positive path from X to Q in the larger diamond, leaving the path from X through P to *not(Q)* unopposed.

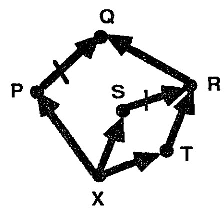

Figure 5. Nested diamonds

The "ambiguity-blocking" skepticism just formulated may be considered unintuitive because the positive path through a nested diamond (e.g., X --> T --> R --> Q in Figure 5) seems to remain a possible conflicting path to its contrary alternative in the larger diamond (e.g., X --> P -/-> Q in Figure 5), as discussed in [Touretzky et al., 1987]. Thus, it may seem that a reasonable skeptic should not conclude *isa(X,not(Q))* in our example, and should propagate the uncertainty of the nested diamond to the larger diamond. Previous attempts to formulate such "ambiguity-propagating" skepticism have not been successful, although it is quite straightforward within

[4] Note that all the directed arcs of our illustrations correspond to *isa_x* links in our theories, while a slash through arcs represents relations to the complement class of the destination node, as *isa_x(penguin, not(flyer))* is represented in Figure 1. Narrow line links correspond to *isa_x* links from individual objects, as *isa_x(opus, penguin)* is represented in Figure 1.

[5] Note that our stricter notion of abnormality inheritance would require a *path(X,S,P)* instead of the simple *isa(X,S)* in this axiom.

our abnormality theories, using the following axioms:

Ambiguity-Propagating Abnormality:

A5′. $ab\_c(X,P,Q) \equiv$
$(\exists R)[isa\_p(X,P) \ \& \ isa\_x(P,Q) \ \& \ isa\_p(X,R) \ \& \ isa\_x(R,not(Q)) \ \& \ \neg ab\_dix(X,P,Q) \ \& \ \neg ab\_dix(X,R,not(Q))]$

where

A7. $isa\_p(X,R) \equiv [isa\_x(X,R) \ \vee \ (\exists T)[isa\_p(X,T) \ \& \ isa\_x(T,R) \ \& \ \neg ab\_dix(X,T,R)]]$.

These axioms function by defining potential *isa* relations between nodes (e.g., *isa_p(X,R)* in Figure 5) that will be genuine relations if they are not cancelled by a conflicting multi-link path. Hence, these axioms enable cancellation of all paths from roots to tips in embedded Nixon diamonds, and yield complete skepticism about all such relations (e.g., between X and R and between X and Q in Figure 5).

4.2.4.2. Credulous Theories

Less skeptical theories are possible if we require only that one or the other of two conflicting paths is blocked, as formalized in [Haugh, 1988].

4.3. Performing the Minimizations

Minimization of explicit relations can be conveniently performed by parallel circumscription of our explicit link predicates *isa_x* and *ab_x*. Alternatively, for non-disjunctive theories, we may define a single explicit link predicate in terms of them and circumscribe it, e.g.:

$$link(X,P,Q) \equiv [ab\_x(X,P,Q) \ \vee \ (isa\_x(X,P) \ \& \ Q = isa\#)]$$

where *isa#* is a reserved constant that simply restricts the range of the extra variable. We can achieve the same effect as circumscription of *link* in a default logic [Reiter, 1980] with the single normal default rule:

$$\frac{: \neg link(a1,c2,c3)}{\neg link(a1,c2,c3)}$$

which asserts that whenever it is consistent for no explicit link to exist between nodes, we may infer that there is none.

4.4. Full Object-Level Interpretations

It is possible to translate all of our meta-level default relations into object-level axioms using ordinary predicates. These translations, however, cannot occur in isolation, since all the relevant abnormalities can be determined only from the structure of the network as a whole. Thus, if we wish an object-level translation, we must determine the relevant abnormalities based upon the structure of the whole net, using our meta-level axioms above, and combine these derived abnormalities with the individual translations of each link.

After all the provable abnormalities are determined from a meta-level theory, translation should proceed as follows:

1) Translate each general abnormality *ab(p,q,r)* as $[p(X) \supset ab(X,q,r)]$ and each particular abnormality *ab(a,q,r)* as itself.
2) Translate every *isa_x(a,p)*, where a is an individual constant, as *p(a)*, and every *isa_x(a, not(p))* as $\neg p(a)$.
3) Translate every explicit default relation *isa_x(p,q)* as $[[p(X) \ \& \ \neg ab(X,p,q)] \supset q(X)]$.
4) Translate every explicit abnormality *ab_x(X,p,q)* as *ab(X,p,q)*.
5) Minimize specific abnormalities (*ab(X,P,Q)*) in the new theory.

5. Re-Opening A Closed-World

5.1. Dual Predicate Classes

5.1.1. Meta-Level Theories

Our closed-world theories yield just the right results regarding abnormalities, but make more assumptions than necessary about what isn't provable. Ideally, we would like to keep the abnormality results, yet no longer insist that all the excluded primitive links are false. This can be accomplished by dividing our predicates into two parallel classes, one in which the closed-world minimizations are performed, and another, general all-inclusive class, in which there are no explicit minimizations, but only some restrictions to prevent interference with the results of the closed-world minimizations.

In particular, let us create new predicates by appending a "*" to all of our previous predicate names, e.g., *isa_x** instead of *isa_x*, and *ab_x** instead of *ab_x*. Then, we may rewrite our axioms using the new predicates, creating a new theory H* whose general axioms correspond to to the old theory H. All of the links of particular inheritance theories will be expressed in H*, and will be minimized as above. But, now our minimized H* theory is only a subset of a larger theory H** which includes H and H*, wherein the predicates of H* are considered as instances of those of H, formalized:

$$isa^*(X,P) \supset isa(X,P)$$
$$ab^*(X,P,Q) \supset ab(X,P,Q), \text{ etc.}$$

Thus, the minimizations of H* will no longer entail closed world assumptions with respect to H**, since explicit links that are absent from H* may still appear in H. Finally, H** will need one more axiom to exclude any new relations in H that could otherwise cancel conclusions of the closed-world theory H*.

$$[isa^*(X, Q) \ \& \ \neg isa\_x^*(X, not(Q))] \supset \neg isa(X, not(Q)).$$

This asserts that no unopposed network relation of the original closed-world theory H* is ever opposed by a contrary relation in the larger theory. This excludes unwanted models while allowing any other relations that don't conflict with the original theory. We might also wish to restore the law of the excluded middle to our meta-level theory for properties of individual objects, i.e.,

$$isa(A,P) \supset \neg isa(A,not(P)).$$

With this formalization of open-world inheritance, we may selectively specify any degree of closure assumptions we like for particular predicates.

5.1.2. Object-Level Interpretations

Object-level predicates would also come in pairs, and network relations would be translated into starred predicates (e.g., *isa_x*(p,q)* translates into *[[p*(x) & ¬ ab*(x,p,q)] ⊃ q*(x)]*). The additional axioms relating starred and unstarred predicates would also be required (e.g., *[C1*(x) ⊃ C1(x)]*) for all object-level predicates. After translation, object-level open-world theories would proceed to minimize the provable direct abnormalities (ab*), thereby permitting any other abnormalities that did not disagree with the positive conclusions of the closed-world theories. Additional closed-world assumptions may also be added for particular classes, as desired, providing the fullest flexibility in specifying the intended assumptions.

5.2. Auto-Epistemic Interpretations

Our use of dual predicates in open-world theories is highly suggestive of an interpretation in auto-epistemic theories. Our starred explicit relations are quite clearly just those that can be proven to hold, i.e., those that are "known" by the system. Thus, it would be natural to translate the entire theory into an auto-epistemic logic in which the starred relations are translated into statements of knowledge, according to the following type of schema:

isa_x*(X,P) => Knows isa_x(X,P)
ab_x*(X,P) => Knows ab_x(X,P)

Object-level translations would be analogous, and semantically revealing:

isa_x*(P,Q) =>
[[Knows P(x) & ¬Knows ab(x,P,Q)]
⊃ Knows Q(x)]

Thus, a default relation between P and Q can be interpreted as asserting that if something is known to be P and is not known to be abnormal with respect to being Q, then it is known to be Q.

Although more work is required on the formal details of such auto-epistemic versions of our theories, they offer considerable promise in providing deeper semantic foundations and more coherent integration with general theories of an agent's knowledge and self-reflection.

6. Computing the Theory

Our formulations are very simple nonmonotonic theories (using a single normal default, or a single simple circumscription), which allow use of the simplest nonmonotonic technique of negation by failure to minimize explicit inheritance links, or to minimize abnormalities in the object-level theories. Furthermore, the skeptical versions of our theories have models that are provably unique with respect to abnormalities, so that alternative extensions needn't be examined. Our theories have also been formulated to avoid generating the many unnecessary abnormalities found in McCarthy's original proposal. Thus, the difficulties of computing the consequences of general inheritance theories have already been minimized significantly by the form of our theories. Computational demands are further reduced in our implementation scheme by careful incremental derivation of abnormalities, truth maintenance on them, and inference algorithms tailored to query types. This scheme and its implementation in Prolog is described in depth in a longer report [Haugh, 1988].

A key technique in computing the consequences of our theories is an initial determination of abnormalities which avoids their repeated computation on every query. Truth maintenance is performed on all derived abnormalities, since subsequent changes may undermine the justifications for current abnormalities. With all the current abnormalities kept updated, queries are processed very efficiently by tracing the unblocked paths through a network.

Acknowledgements

I am especially indebted to Jeff Horty for assistance in identifying the limitations of previous work. I also thank Steve Barash, Steve Jameson, Stuart Pearlman, and Donald Perlis for helpful comments.

References

Etherington, D.W. 1987. "Formalizing Nonmonotonic Reasoning Systems," *Artificial Intelligence*, 31(1):41-86.

Hanks, S. and D. McDermott. 1986. "Default Reasoning, Nonmonotonic Logics, and the Frame Problem," in *AAAI-86*, pp. 328-333.

Haugh, B. 1987. "Simple Causal Minimizations for Temporal Persistence and Projection," in *AAAI-87*, pp. 218-223.

Haugh, B. 1988. "Tractable Logical Theories of Multiple Defeasible Inheritance," Martin Marietta Laboratories Technical Report (forthcoming).

Horty, J., R. Thomason, and D. Touretzky. 1987. "A Skeptical Theory of Nonmonotonic Semantic Networks," in *AAAI-87*, pp. 358-363.

Lifschitz, V. 1985. "Computing Circumscription," in *IJCAI-85*, pp. 121-127.

Lifschitz, V. 1987. "Formal Theories of Action (Preliminary Report)," in *IJCAI-87*, pp. 966-972.

McCarthy, J. 1980. "Circumscription — A Form of Non-Monotonic Reasoning," *Artificial Intelligence*, 13:27-39.

McCarthy, J. 1986. "Applications of Circumscription to Formalizing Common-Sense Knowledge," *Artificial Intelligence*, 28(1):89-116.

McDermott, D. and J. Doyle. 1980. "Non-Monotonic Logic I," *Artificial Intelligence*, 13:41-72.

Minsky, M. 1975. "A Framework for Representing Knowledge," In *Psychology of Computer Vision*, P. Winston (ed.), New York: McGraw-Hill, pp. 211-277.

Reiter, R. 1980. "A Logic for Default Reasoning," *Artificial Intelligence* 13:81-132.

Sandewall, E. 1986. "Nonmonotonic Inference Rules for Multiple Inheritance with Exceptions," in *Proceedings of the IEEE*, 74(10):1345-1353.

Touretzky, D. S. 1986. *The Mathematics of Inheritance Systems*, Los Altos: Morgan Kaufmann.

Touretzky, D., J. Horty, and R. Thomason. 1987. "A Clash of Intuitions: The Current State of Nonmonotonic Inheritance Systems," in *IJCAI-87*, pp. 476-482.