More On Inheritance Hierarchies with Exceptions
Default Theories and Inferential Distance

David W. Etherington[1]
Artificial Intelligence Principles Research Department
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974-2070
ether%allegra@btl.csnet

Abstract

In Artificial Intelligence, well-understood reasoning systems and tractable reasoning systems have often seemed mutually exclusive. This has been exemplified by nonmonotonic reasoning formalisms and inheritance-with-exceptions reasoners. These have epitomized the two extremes: the former not even semidecidable, the latter completely *ad hoc*.

We previously presented a formal mechanism for specifying inheritance systems, and minimal criteria for acceptable inheritance reasoning. This left open the problem of realizing an acceptable reasoner. Since then, Touretzky has developed a reasoner that appears to meet our criteria. We show that his reasoner *is* formally adequate, and explore some of the implications of this result *vis-à-vis* the study of nonmonotonic reasoning.

1. Introduction

Nonmonotonic reasoning formalisms have been the subject of much interest lately (*c.f.* [AI 1980], [AAAI 1986]). They provide principles for representing and reasoning with rules that generally hold but are subject to exceptions. Although the ability to reason with such rules appears to be a central facet of intelligence, the formalisms developed to this point have been intractable. For example, in the general case default logic is not even semidecidable.

Motivated by a need to build systems with good computational properties, many researchers have sacrificed formal precision. Faced with the worst-case intractability of formal systems, they have despaired of formalism altogether. While this has sometimes led to very fast "inference" mechanisms, there has often been little more than vague intuitions about exactly *what* these mechanisms infer. A canonical example of this has been the use of inheritance reasoning in AI systems.

Inheritance reasoners represent a system's knowledge as a connected set of nodes. The nodes represent classes and/or individuals, with associated sets of properties. The connections indicate the flow (or *inheritance*) of properties from "more general" to "less general" nodes. Such systems frequently make provision for exceptions to inheritance, allowing "peculiar" individuals or classes to preempt the normal flow of properties.

In the absence of adequate semantic characterizations of inheritance systems, correct inference has typically been defined (to the extent it has been defined at all) in terms of intuitions and the behaviour of particular systems. This has lead to anomalous results, including mismatches between intuition and system performance (see [Etherington 1987b] or [Touretzky 1986] for examples).

In earlier work [Etherington & Reiter 1983; Etherington 1987b], we presented a formal mechanism for specifying inheritance systems, and minimal criteria for acceptable inheritance reasoning. This left open the problem of realizing an acceptable reasoner. Since then, Touretzky has developed a reasoner that appears to meet our criteria. We show that his reasoner is formally adequate, and explore some of the implications of this result *vis-à-vis* the study of nonmonotonic reasoning.

2. The Inheritance Language

For the purposes of this paper, we adopt Touretzky's [1986] network representation, which differs from that in [Etherington 1987b].[2] This representation has four link types, shown in Figure 1. Each link has one interpretation if it originates from an individual-node, and another if from a class-node. Relational links have a third interpretation when they connect two individual-nodes. (We use upper and lower case letters for classes and individuals, respectively.)

[1] Parts of this work were done at the University of British Columbia, and supported in part by an I.W. Killam Predoctoral Scholarship and by NSERC grant A7642.

[2] Specifically, strict links and exception links are not treated, and relational links have been added.

Example 1

We can illustrate IS-A and ISN'T-A links with an example from [Fahlman *et al* 1981]. Consider the following facts about invertebrates:

> Molluscs are normally shell-bearers.
> Cephalopods are Molluscs,
>   but normally are *not* shell-bearers.
> Nautili are Cephalopods and *are* shell-bearers.
> Fred is a Nautilus.

Our network representation of these facts is given in Figure 2.

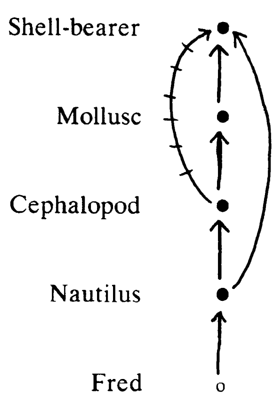

Figure 2 — Network representing facts about Molluscs.

Given the intuitive link definitions, above, one can see the correspondence between the facts contained in the English description and the links in Figure 2. What remains to be done is to describe how such structures can be used to retrieve the information that common sense suggests is contained in the example. For example, are Cephalopods Shell-bearers?

3. The Inferential Distance Algorithm

To show how the information in Example 1 is actually encoded in Figure 2, we must briefly describe how conflicting inheritance is resolved. Looking at Figure 2, *Cephalopods* could be *Shell-Bearers* by virtue of the chain of IS-A links through *Mollusc*. On the other hand, they might not be, because of the ISN'T-A link. The usual approach decides conflicting inheritance by choosing the value most closely-connected with the node in question, but this "shortest-path heuristic" can lead to anomalous results [Etherington 1982, 1987b; Touretzky 1986].

A better approach is the inferential distance algorithm,[3] which arbitrates such conflicts by appealing to a topological view of the network. This approach avoids the failings of the shortest-path heuristic, yet remains faithful to the intuitions that make inheritance networks appealing. In particular, more specific facts prevail over those less specific. Essentially, if an individual could inherit property P because she IS-A B, and property $\neg P$ because she IS-A C, then the ambiguity is resolved by considering relationships between B's and C's. If C IS-A B and not *vice versa*, $\neg P$ is inherited; otherwise, if B IS-A C and not *vice versa*, P is inherited; otherwise, neither is inherited.

As an illustration, consider the network of Figure 2. Because *Nautilus* is a subclass of *Cephalopod*, which is a subclass of *Mollusc*, inferential distance gives the desired results: *Nautili*, such as *Fred*, are *Shell-Bearers*, while *Cephalopods* not known to be *Nautili* are not. In the network of Figure 3, however, neither *Republican* nor *Quaker* is a subclass of the other, so inferential distance sanctions no conclusions about whether *Nixon* is a *Pacifist*.

[3] Due to space limitations, we can only present an oversimplified approximation of the algorithm. The interested reader is referred to Touretzky's [1986] dissertation.

(1) IS-A: *A*●———>●*B*: Normally *A*'s are *B*'s, but there may be exceptions.
a○———>●*B*: The individual *a* belongs to the class *B*.

(2) ISN'T-A: *A*●—┼┼┼—>●*B*: Normally *A*'s are not *B*'s.
a○—┼┼┼—>●*B*: *a* is not an *B*.

(3) RELATED: *A*●══R══>●*B*: Normally *A*'s are related by *R* to *B*'s.
a○══R══>●*B*: Normally *a* is related by *R* to *B*'s.
a○══R══>○*b*: *a* is related by *R* to *b*.

(4) UNRELATED: *A*●╫╫R╫╫>●*B*: Normally *A*'s are not related by *R* to *B*'s.
a○╫╫R╫╫>●*B*: Normally *a* is not related by *R* to *B*'s.
a○╫╫R╫╫>○*b*: *a* is not related by *R* to *b*.

Figure 1 — Links and informal semantics.

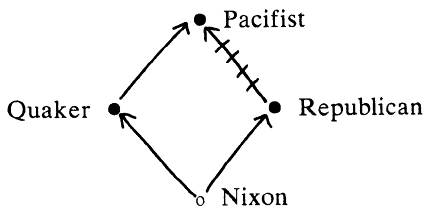

Figure 3— A genuinely ambiguous inheritance net.

Touretzky [1985, 1986] also explores the applications of inferential distance to "inheritable relations". These are relations between classes and/or individuals that, like class-membership, may be inherited by subclasses/instances and may be subject to exceptions. For example, consider Example 2, whose corresponding network is shown in Figure 4.

Example 2

Citizens dislike crooks.
Elected crooks are crooks.
Gullible citizens are citizens.
Gullible citizens don't dislike elected crooks.
Dick is an elected crook.
Fred is a gullible citizen.

In this example, citizens generally dislike crooks, and hence elected crooks. However, Fred, the gullible citizen, doesn't dislike Dick, the elected crook.

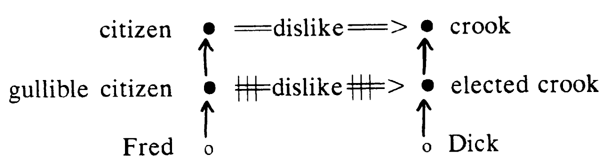

Figure 4— Inheritable relations.

The inheritance mechanism for inheritable relations is similar to that for property inheritance except that, in addition to exceptions to IS-A inheritance, exceptions to relations (such as gullible citizens not disliking elected crooks) must be accounted for.

4. Default Logic

In the spirit of [Etherington & Reiter 1983], we present a translation from Touretzky's inheritance networks to default logic [Reiter 1980]. The proof-theory of default logic then provides minimal criteria that inference algorithms for inheritance systems should satisfy. Unfortunately, this presupposes a familiarity with default logic, which could not be reasonably be presented to the neophyte in the space available. We can only present a sketchy refresher.

For our purposes, a *(normal) default* is a rule of inference, of the form:

$$\frac{\alpha(\vec{x}) : \beta(\vec{x})}{\beta(\vec{x})},$$

which can be interpreted as saying that if $\alpha(\vec{x})$ is known and it is consistent to believe $\beta(\vec{x})$, then it is reasonable to assume $\beta(\vec{x})$. The defaults can be viewed specifying preferred ways of extending one's knowledge about the world.

A *default theory* consists of a set of defaults, D, and a set of first-order facts, W. An *extension* for a default theory is a smallest set that contains W and satisfies the defaults in D. A normal default theory may have one or more extensions.

5. Default Logic and Inheritance Networks

In section 3, we presented a number of links that could be used to create inheritance networks. We now interpret these, using defaults and first-order formulae, as theories of default logic.

Depending on whether it originates from a class A or an individual a, an IS-A link to B is interpreted by:

$$\frac{A(x) : B(x)}{B(x)} \quad \text{or} \quad B(a)$$

respectively. Similarly, ISN'T-A links from classes or individuals are identified, respectively, with:

$$\frac{A(x) : \neg B(x)}{\neg B(x)} \quad \text{or} \quad \neg B(a)\,.$$

The three forms of RELATED link for relation R — class-class, individual-class, and individual-individual — are represented, respectively, by:

$$\frac{A(x) \wedge B(y) : R(x,y)}{R(x,y)},\ \frac{B(y) : R(a,y)}{R(a,y)} \quad \text{and} \quad R(a,b)\,.$$

Finally, the three forms of UNRELATED link respectively yield:

$$\frac{A(x) \wedge B(y) : \neg R(x,y)}{\neg R(x,y)},\ \frac{B(y) : \neg R(a,y)}{\neg R(a,y)} \quad \text{and} \quad \neg R(a,b)\,.$$

These mappings allow Touretzky's inheritance networks to be interpreted as default theories. For example, the default logic representation[4] of the network in Figure 2 is:

$$D = \left\{ \begin{array}{lll} \frac{M(x) : Sb(x)}{Sb(x)}, & \frac{C(x) : M(x)}{M(x)}, & \frac{N(x) : C(x)}{C(x)}, \\ \frac{C(x) : \neg Sb(x)}{\neg Sb(x)}, & \frac{N(x) : Sb(x)}{Sb(x)} & \end{array} \right\}$$

$$W = \left\{ N(Fred) \right\}.$$

[4] (using the obvious abbreviations)

Since the extensions of normal default theories represent orthogonal sets of beliefs that might be justified given the first-order world-description and the defaults, we clearly must require that the conclusions drawn by an inheritance reasoner lie within a single extension of the corresponding default theory. This does not provide a complete characterization of an inheritance reasoner, however, since it does not specify *which* extension should be chosen in cases where there are more than one. Happily, inferential distance also provides a mechanism for choosing among multiple extensions.

6. Relating Inferential Distance and Default Logic

Touretzky [1984] considers the possibility of applying the inferential distance topology to default theories, and gives some examples. The idea is that the notions which lead one link to be preferred to another in a network might also be applicable to help resolve conflicting defaults in default theories, without changing the forms of the defaults themselves (as in [Reiter and Criscuolo 1983]). It is not clear from Touretzky's presentation, however, exactly how the results of this application correspond to the results sanctioned by default logic. In this paper, we begin to explore this question.

Conceptually, the inferential distance algorithm eliminates those extensions that violate the "hierarchical" nature of the representation, then draws those conclusions that hold in the remaining extensions. This approach captures the semantic intuition that properties associated with subclasses should override those associated with superclasses, which is the fundamental *raison d'être* for inheritance representations. That it also avoids the pitfalls of incorrect behaviour that curse shortest-path inference algorithms is shown by the following theorem.[5]

Theorem

> In the absence of "no-conclusion" links, the ground facts returned by the inferential distance algorithm lie within a single extension of the default theory that corresponds to the inheritance network in question. ∎

The theorem begins to determine the connections between Touretzky's work and default logic, by showing that ground facts returned by inferential distance — *e.g.*, "Clyde is an elephant", or "Clyde loves Fred" — belong to a common extension of the corresponding default theory. However, inferential distance also sanctions normative conclusions, such as "Albino-elephants are [typically] herbivores". We have begun to explore the relationship such statements inferred under inferential distance bear to the underlying default theory, but our results are only preliminary.

Touretzky also allows "no-conclusion" links, which allow inheritance to be blocked without explicit cancellation. Default logic has no analogue for the no-conclusion link, and we have not considered them here. It appears straightforward to add a similar capacity to the logic, assuming such links prove useful. The proof of the theorem suggests that its generalization to networks with no-conclusion links *vis-à-vis* such an extended logic would present no problems.

Touretzky [1986] explores the properties of inferential distance inheritance reasoning in detail. He also provides a constructive mechanism for determining the 'grounded expansions' (analogous to extensions) of a network. Many of his results bear a superficial similarity in form and proof to the corresponding results for default logic. We speculate (as has Touretzky) that this is no accident. In the next section, we suggest that the two approaches are so closely related that an inferential distance reasoner can be viewed as a restricted default logic theorem-prover.

7. Tractability

As we mentioned earlier, tractability is the rock on which formalism founders. Logic-based approaches in AI tend to be semi-decidable or worse, and so do not lend themselves to implementation. Conversely, informal systems often have attractive computational complexities (*e.g.*, $O(N)$ for inheritance in a hierarchy). Furthermore, there has been an expectation that massively-parallel machine architectures could yield a further logarithmic improvement. Still, we argue that it is not particularly useful to do "I-don't-know-what", *very* quickly. Can *principled* commonsense reasoning be done quickly?

The answer appears to be "yes, although perhaps not as quickly as unprincipled reasoning". For example, one proposed parallel architecture for inheritance reasoning involves parallel marker-passing machines [Fahlman 1979]. Touretzky shows that there are networks for which parallel marker-passing algorithms cannot derive the conclusions sanctioned by the inferential distance algorithm. However, he also shows that any network can be "conditioned", by adding logically-redundant links, in such a way that a parallel marker-passing algorithm *can* return correct results. Unfortunately, this conditioning, which must be done each time the network is modified, is expensive (Touretzky [1986] gives a polynomial-time algorithm that adds $O(N^2)$ links in the worst case) and is apparently not amenable to parallel marker-passing

[5] The proof of the theorem is given in [Etherington 1987a].

implementation [Touretzky 1982: personal communication; 1983].

Compared with the worst-case undecidability of default theories, however, a polynomial-time/space update algorithm and a linear-time inference algorithm do not seem entirely unattractive. Given the theorem above, such algorithms correctly determine inheritance in the presence of exceptions. Thus, they can be viewed as fast inference algorithms for reasoning with (what can be seen to be) the tractable class of default theories that correspond to inheritance networks. This result has dual significance. Not only does it provide another formal justification for inheritance reasoners based on inferential distance, it proves that there are interesting classes of theories for which default logic is tractable, and provides an efficient algorithm for computing with such default theories.

8. Discussion

If we are suggesting that inferential distance can be used to reason with some class default theories, we should at least consider how close the relationship between the two approaches is. For example, while we showed that all conclusions reached by inferential distance lie within a single extension of the underlying default theory, those conclusions may actually be more tightly constrained. In the Quaker/Republican example no conclusion is reached about Dick's being a Pacifist. The default theory has 2 extensions, however, one supporting each possibility. Inferential distance, in this case, returns conclusions that lie in the intersection of the extensions.

It seems that this is a general situation. Certain extensions are ruled out altogether (those corresponding to possible inferences clearly superceded by inferences associated with subclasses). It appears that conclusions are returned that lie in the intersection of those extensions that reflect genuine ambiguities in the network. This remains to be proved.

Since the default theories that represent hierarchies are normal, it would seem that a result analogous to Reiter's [1980] "semimonotonicity" theorem might be expected for networks. This result guarantees that adding new defaults to a normal default theory never causes extensions to "go away", so a reasoner committed to one extension may remain so committed on discovering new default information. Unfortunately, this is not the case. While the set of extensions for the underlying theory does not contract, some that may have been preferred initially may not remain so given new defaults, and *vice versa*. This is not unexpected, given Touretzky's observation that networks must be "reconditioned" after each update.

Another discrepancy between the two approaches is that there are no strict (exception-free) links in Touretzky's scheme. Brachman [1985] and others have argued that this is a serious shortcoming in a knowledge representation system. We have not considered the feasibility of adding such links. It is clearly possible, but we have no idea what the computational cost would be.

9. Conclusions

We have explored a correspondence between default theories and inheritance networks with exceptions. Using the proof-theory of default logic as minimum correctness criteria for inheritance-determination, we showed that Touretzky's inferential distance algorithm is a satisfactory inheritance reasoner. More importantly, we were able to turn our notion of satisfactory around and find that Touretzky's algorithm provides a tractable proof-theory for certain classes of default theories. Such tractable algorithms are welcome not only for their own sake, but because they suggest that intractability may not be the inevitable cost of formal adequacy in commonsense reasoning.

The formality/tractability controversy has long divided AI, with little communication (beyond epithets) between the camps. Recently there has been interest in exploring the terrain between the encampments. Early reports, including this one, suggest that the ground is fertile. Perhaps the natives are even friendly!

Acknowledgements

I am grateful to David Touretzky, Raymond Reiter, and Robert Mercer, for helping to sharpen my insights into these problems.

References

AAAI [1986], *Proc. American Assoc. for Artificial Intelligence-86*, Philadelphia, PA, August 11-15, 406-410.

AI [1980], Special issue on non-monotonic logic, *Artificial Intelligence 13*, North-Holland.

Brachman, R.J. [1985], "I lied about the trees, or defaults and definitions in knowledge representation", *AI Magazine 6(3)*, 80-93.

Etherington, D.W. [1982], *Finite Default Theories*, M.Sc. thesis, Dept. Computer Science, University of British Columbia.

Etherington, D.W. [1987a], *Reasoning from Incomplete Information*, Pitman Research Notes in Artificial Intelligence, Pitman Publishing Limited, London.

Etherington, D.W. [1987b], "Formalizing nonmonotonic reasoning systems", *Artificial Intelligence 31*, North-Holland, 41-85.

Etherington, D.W., and Reiter, R. [1983], "On inheritance hierarchies with exceptions", *Proc. American Assoc. for Artificial Intelligence-83*, Washington, D.C., August 24-26, 104-108.

Fahlman, S.E. [1979], *NETL: A System for Representing and Using Real-World Knowledge*, MIT Press, Cambridge, Mass.

Fahlman, S.E., Touretzky, D.S., and van Roggen, W. [1981], "Cancellation in a parallel semantic network", *Proc. Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B.C., Aug. 24-28, 257-263.

Reiter, R. [1980], "A logic for default reasoning", *Artificial Intelligence 13*, North-Holland, 81-132.

Reiter, R. and Criscuolo, G. [1983], "Some representational issues in default reasoning", *Int'l J. Computers and Mathematics 9*, 1-13.

Touretzky, D.S. [1983], *Multiple Inheritance and Exceptions*, Unpublished Manuscript, Department of Computer Science, Carnegie-Mellon University.

Touretzky, D.S. [1984], "Implicit ordering of defaults in inheritance systems", *Proc. American Assoc. for Artificial Intelligence*, 322-325.

Touretzky, D. [1985], "Inheritable relations: a logical extension to inheritance hierarchies", *Proc. Theoretical Approaches to Natural Language Understanding*, Halifax, 28-30 May, 55-60.

Touretzky, D.S. [1986], *The Mathematics of Inheritance Systems*, Pitman Research Notes in Artificial Intelligence, Pitman Publishing Limited, London.