A Semantical Approach to Stable Inheritance Reasoning

Craig Boutilier
Department of Computer Science
University of Toronto
Toronto, Canada M5S 1A4

Abstract

Inheritance reasoning has frequently been characterized by algorithms designed to operate on inheritance networks, or collections of links. While the intended meaning of links in a network is understood, formal semantic accounts of such networks are somewhat troublesome, as are semantic accounts of the inference process. We suggest that links be interpreted as sentences in the conditional logic E, providing a formal interpretation for such networks. Furthermore, we develop a semantic characterization of inheritance reasoning based on the technique of minimal (or preferred) models. In the process, we identify a key difference between this characterization of inference in networks and those based on the notion of inferential distance, specifically with respect to stability.

1 Introduction

A number of techniques exist for reasoning with nonmonotonic multiple inheritance systems. While they have drawn the attention of many researchers, consensus on the meaning of networks and the approach to be taken by inheritance reasoners has yet to be achieved. In fact, the number of choices available to the designer of such systems is considerable. In [Touretzky *et al.* 1987], the design space for inheritance reasoners is explicitly mapped out, and the widely varied approaches and implications of existing schemes are made clear. There it is claimed that this divergence represents a number of different, but equally plausible reasoning strategies, suitable for various tasks. We conjecture that a comparison of such techniques is made difficult by a lack of a semantic account of inheritance reasoning, and without such, the intuitive appeal and adequacy of these systems cannot be accurately evaluated. A semantic characterization of inheritance reasoning can more readily capture our intuitions. Furthermore, such a semantics can be used as a yardstick with which to measure the adequacy of syntactic, algorithmic techniques.

The goal of this paper is to provide a semantic characterization of inheritance reasoning. This will be achieved by, first, providing a semantic interpretation of network links, and second, utilizing the concept of minimal (or preferred) models to account for the nonmonotonic nature of inference in inheritance systems. We will also show how this approach differs from traditional methods in a crucial respect.

1.1 Definition of a Network

Inheritance systems are essentially devices which perform (possibly nonmonotonic) inference in a computationally feasible manner by restricting the form "sentences" in the "language" can take. Although many definitions of inheritance networks exist, the following is similar in spirit to the one given in [Touretzky 1986]. The details will be much simpler though, since we require less machinery to elucidate the relevant concepts in this paper. In particular, while inheritance networks generally deal with both classes of objects and individuals, we will not distinguish individuals in order to simplify the treatment. As well, the only relationship between classes permitted is the ISA (or subclass) relation.

Definition An *inheritance network* over a finite set of classes Γ is an acyclic set of ordered triples (*links*) of the form $\langle sign\, x, y\rangle$ where $sign \in \{\#, +, -\}$ and $x, y \in \Gamma$.

The intended interpretation of the links in these networks is as follows:

- $\langle + A, B\rangle$ means "members of A are members of B".
- $\langle - A, B\rangle$ means "members of A are not members of B".
- $\langle \# A, B\rangle$ means "members of A may or may not be members of B".

For instance, $\langle +\, Apple, RedThing\rangle$ represents the assertion that apples are red. The third type of link is not meant to be true when it is unknown whether A's are B's; rather, it is intended to state that both of the other links are false. In other words, the link is true exactly when the relationship of A to B is known, and that relationship is (something like) "A's may or may not be B's with roughly equal likelihood" (e.g. $\langle \# \; American, Republican\rangle$).

Networks are often represented as directed graphs in the obvious manner (see, e.g., [Touretzky 1986]). A link $\langle + \; A, B\rangle$ will be also written as $A \longrightarrow B$, and $\langle - \; A, B\rangle$ as $A \not\longrightarrow B$.

Given a network, an inheritance reasoner is charged with the task of deciding which statements (or links) should be inferred from those given. If classes were interpreted as monadic predicates and links as universally-quantified statements, there would exist an obvious translation of networks into first-order logic. The inheritance reasoner then could simply be a first-order theorem prover to act on these new sentences; the semantics of inheritance networks would be clear. In fact, for inheritance networks without exceptions, this interpretation is acceptable. The problem is most natural subclass relationships are not universal. Exceptions are the rule, so to speak. Therefore, the treatment here will be of *inheritance networks with exceptions.* In this type of network, all links will be viewed as *exception-allowing.* If $\langle + A, B\rangle$ is a link in the network, it may be the case that a particular A is not a B. This suggests the links be given a normative interpretation. So $\langle + \; A, B\rangle$, in an exception-allowing network, stands for "Normally A's are B's".

In exception-allowing networks, the semantic import of the links is not as clear as in the case of nets which prohibit exceptions. Attempts have been made to interpret links in terms of autoepistemic logic ([Touretzky 1986]) and default logic ([Etherington and Reiter 1983]) with limited success; and while these may provide a loose semantic interpretation for the links themselves, little attempt has been made to account for the nature of inference in exception-allowing networks. While exceptionless networks are (relatively) unproblematic, classical semantics will not adequately reflect exception-allowing networks, due to their nonmonotonic nature.

A survey of the traditional approaches to inheritance reasoning can be found in [Touretzky *et al.* 1987]. For reference to specific systems of inheritance, see, for example, [Touretzky 1986], [Horty *et al.* 1986], and [Sandewall 1986].

2 A Minimal Model Approach to Inheritance

2.1 Interpretation of Links

As mentioned in the previous section, links in an inheritance network can be reasonably interpreted as asserting normative statements. Obviously, such statements are not necessarily universally true regarding members of their constituent classes; but they should be subjected to some normative attribution. Some objections to this denotation are raised by Touretzky [1986, pp.6–7]; however, these are not serious and are not considered here (see [Boutilier 1989]).

In [Boutilier 1988], the conditional logic E is presented for reasoning about default or prototypical properties[1]. The language of E is that of classical propositional logic augmented with the connective $\Rightarrow$. The intended interpretation of a sentence $A \Rightarrow B$ is "If A holds, then in the normal course of events, B holds as well". The logic is given a possible worlds semantics that provides the desired intuitive characterization of such exception-allowing sentences. Given a set of possible worlds and a reflexive, transitive, forward-connected (if xRy and xRz, then either yRz or zRy) relation on that set (representing accessibility of less-exceptional worlds), $A \Rightarrow B$ is true at some world if there is some accessible world such that A holds, and $A \supset B$ holds at all worlds accessible from that point.

Since E has a well-developed semantics, and we have argued that links in inheritance networks be given a normative interpretation, it seems that translating inheritance networks into sentences of the logic E would provide an adequate logical account of the "connective" $\longrightarrow$. In fact, it is precisely this device which will be used to provide a semantic account of links in a network. The following translation is suggested:

1. $\langle + \; A, B\rangle$ is translated to $A \Rightarrow B$, with the interpretation that A's are normally B's.

2. $\langle - \; A, B\rangle$ is translated to $A \Rightarrow \neg B$, interpreted as A's are normally not B's.

3. $\langle \# \; A, B\rangle$ is translated to $\neg(A \Rightarrow B) \wedge \neg(A \Rightarrow \neg B)$. This states that neither of the other two alternatives hold.

Now the semantic import of a network is clear: a network is a collection of sentences in the language of E, and as such, has associated with it the semantics of the logic E.

While the meaning of networks themselves has been specified, the logic E cannot be used alone to identify nonmonotonic consequences, and hence, cannot account for inference in inheritance networks. We must establish new conditions with which to capture the nonmonotonic nature of inheritance.

A number of algorithms characterizing inference have been presented in the literature[2], but these characterizations yield incredibly divergent results [Touretzky *et al.* 1987]. It is conjectured that this state of affairs exists because a clear semantic account of inference in inheritance reasoners has not been proposed. A yardstick is needed which can measure the intuitive

[1] E is an extension of the logic N found in [Delgrande 1987] that deals with nested conditionals.

[2] These can apply to sentences of E as well as to other syntactic entities, or links, they were designed for.

appeal of syntactic characterizations and enhance our understanding of the entire process.

The approach we will take is one suggested by Shoham [1986]. To identify the nonmonotonic consequences of a theory, we restrict our attention to those models of the theory which are in some sense preferred. We will associate a reflexive, transitive *preference* relation $\leq$ with the set of models of a network, where $M_1 \leq M_2$ is intended to mean M_1 is as prefable as M_2. The preferred models are those minimal in this relation. The minimal model approach seems to be able to capture in a natural fashion the ideas behind nonmonotonic reasoning, therefore we will enlist it to aid in providing a semantic characterization of inheritance reasoning.

Before discussing the concept of a preferred model, we must define a model for an inheritance network. Links in a network have been associated with sentences in the language of E. Therefore, under this plausible normative interpretation, E-models could be considered as models of inheritance networks, an E-model satisfying a network exactly when it satisfies the translation of its links. Preference criteria could then be defined for E-models such that minimal E-models of the network determine the nonmonotonic consequences (or extensions) of that network.

While in principle this could be achieved, it is not the tack taken here. In dealing with the full generality of the language of E complications and subtleties may arise which are not entirely relevant to the discussion at hand. Thus, a simpler notion of a model will be provided, allowing full attention to be paid to the relevant aspects of the preference relation.

Definition An *I-model* (over a finite set of classes Γ) is a set of links such that exactly one of $\langle +\, A, B \rangle$, $\langle -\, A, B \rangle$, $\langle \#\, A, B \rangle$ is in the set for each $A, B \in \Gamma, A \neq B$, and the $ISA/ISNOTA$-subgraph of this set is acyclic.

Definition An I-model I *satisfies* a network Φ iff for each link $\lambda \in \Phi$, $\lambda \in I$.

2.2 The Assumption of Redundancy

In [Touretzky 1986], Touretzky presents his notion of inferential distance, the use of which has become a hallmark of contemporary inheritance reasoners. A class B is defined to be "closer" to a class A than is a class C in a network if there exists a path from A to C which passes through B. If a conflict arises with respect to a property that class A should inherit, then this conflict is resolved by inheriting the property from the closest of the conflicting superclasses.

The use of inferential distance in this manner yields intuitive results from a number of networks. For instance, the network in Figure 1 asserts that people who work hard (H) are productive (P), and that most researchers (R) are quite diligent, yet fail to make significant progress (for the sake of argument!). Now Sarah (S)

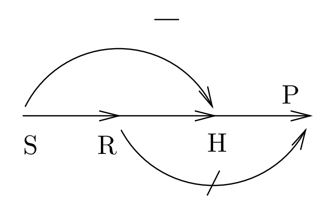

Figure 1: Inferential Distance

should inherit the characteristic of being productive by virtue of being a hard worker, while the fact that she's a researcher suggests she's not productive at all. Reasoning based on inferential distance will solve this conflict by determining that Sarah is unproductive because researchers are a specific subclass of hard workers (i.e. R is closer to S than is H).

This example shows the intuitive appeal of inferential distance. Why is it, though, that such an arbitrary criterion as "inherit from the closest superclass" provides intuitive results in this case? Suppose Sarah is a hard worker for reasons altogether independent of her being a researcher (say, her editorial work), and is quite productive because of this. In such a case, inferential distance does not provide intuitive results. The fact that R is closer to S than is H is no reason to inherit properties from R instead of H. However, this counterexample is intuitively unacceptable. Given just the information in the network, the conclusion that Sarah is unproductive should be forthcoming. The reasoner should assume Sarah is a hard worker *because* she is a researcher (since the an independent reason makes the use of inferential distance quite unreliable). In other words, the link $S \longrightarrow H$ is "redundant", due to the presence of the links $S \longrightarrow R$ and $R \longrightarrow H$. Without assuming such redundancy, the justification for inferential distance simply doesn't exist.

Calling links redundant may give the impression that such links add no information to a network, which is not the intention[3]. Rather than call these links redundant, "independently justified" may be a more appropriate term. In general, a link is considered to be redundant in a network if there exists a set of links that can be construed as the reason for the truth of that link. A more rigorous and comprehensive definition of redundancy can be given after the development of certain definitions in the next subsection.

2.3 The Preference Relation

In the minimal model framework, nonmonotonic consequences are derived by identifying the preferred models of a theory, in this case, a network. In the terminology of other network formalisms, the preferred models should

[3]Indeed, the conclusions reached from a network with a "redundant" link may be different from those with the link missing (e.g. see [Touretzky 1986, pp.10–11], or [Boutilier 1989]).

be those which satisfy permissible "arguments", or *paths.* Rather than speak of arguments, we will use the concept of *support.* For instance, if the links $A \longrightarrow B$ and $B \longrightarrow C$ belong to Γ, then these sentences will support the conclusion $A \longrightarrow C$, unless more specific evidence contradicts this. It remains to be stated what constitutes this specific evidence for inheritance networks.

It will be convenient to use the path notation of traditional reasoners as shorthand.

Definition A *path* $\langle *x_1, x_2, \ldots, x_n \rangle$ *is contained in* a set of links Φ iff each of the links $\langle + x_1, x_2 \rangle$, $\langle + x_2, x_3 \rangle$, $\ldots$, $\langle * \, x_{n-1}, x_n \rangle$ is in Φ (where $* \in \{+, -\}$).

For convenience, $\overline{*}$ will denote the "complement" of $*$, which is any sign not equal to $*$ (e.g. if the value of $*$ is $+$, then $\overline{*}$ can be either of $\#$ or $-$). Notice that paths can only be positive or negative, while links can be neutral as well. If λ is a (positive or negative) link $\langle * \, x, y \rangle$, then $\overline{\lambda}$ denotes either of the two complement links $\langle \overline{*} \, x, y \rangle$[4].

We will now define a reflexive, transitive preference relation on the set of I-models. In general, models which respect the transitivity of subclass relationships should be preferred to those which do not. That is, if a model contains the path $\langle * \, x_1, x_2, \ldots, x_n \rangle$, then it should also contain the link $\langle *x_1, x_n \rangle$. Unsurprisingly however, there are exceptions to this rule, when more specific evidence presents itself. These exceptional conditions reflect redundancy considerations.

Example Assume the following paths are present in a network Γ:

$\langle + \;\; x_1, x_2, \ldots, x_n \rangle, \quad \langle + \;\; x_i, y, x_{i+1} \rangle, \quad$ and $\langle - \, y, x_n \rangle$,

where $1 \leq i < n-1$. The first path indicates support for the assertion $x_1 \longrightarrow x_n$. This is because x_1's are x_2's, which are x_3's, and so on. The second sequence can be interpreted as the reason $x_i \longrightarrow x_{i+1}$ holds. Now, this line of reasoning can be extended to show that the justification for concluding $x_1 \longrightarrow x_n$ is the path $\langle + x_1, \ldots, x_i, y, x_{i+1}, \ldots, x_n \rangle$. However, it is also known that $y \not\longrightarrow x_n$. This makes y an exceptional subclass of x_{i+1} with respect to x_n. Therefore x_1 should inherit properties from y over x_{i+1}, making x_1 an exceptional subclass of x_{i+1}. Just as with inferential distance, since y and x_{i+1} conflict with respect to property x_n, x_1 should inherit the property $x_1 \not\longrightarrow x_n$ from y, since y is closer to x_1 than is x_{i+1}. The concept illustrated in this example is closely related to Touretzky's [1986] notion of preclusion.

Example Assume the following paths are present in a network Γ:

$$x_1 \overbrace{\longrightarrow x_2 \longrightarrow x_3 \ldots x_{n-2} \longrightarrow x_{n-1}}^{\text{Type}-1} \longrightarrow x_n$$
$$x_1 \longrightarrow \underbrace{x_2 \longrightarrow x_3 \ldots x_{n-2} \longrightarrow x_{n-1} \longrightarrow}_{\text{Type}-2} x_n$$

Figure 2: Positioning of intermediaries

$\langle + \;\; x_1, x_2, \ldots, x_n \rangle, \quad \langle + \;\; x_i, y, x_{i+1} \rangle, \quad$ and $\langle - \, x_1, y \rangle$,

where $2 \leq i < n$. The first path indicates support for the assertion $x_1 \longrightarrow x_n$ and the reason for this (if the second path is taken into consideration) is because x_1's are x_2's, which are $\ldots x_i$'s, which are y's, which are x_{i+1}'s, $\ldots$ which are x_n's. However, the third link asserts that x_1's are not y's, thus disabling the entire chain of reasoning. Hence, the conclusion $x_1 \longrightarrow x_n$ should not be drawn. This example illustrates a concept closely related to Touretzky's [1986] notion of contradiction.

These definitions, in the style of [Touretzky 1986], will formalize the concepts required by the preference relation. Let Φ be any set of inheritance links.

Definition A *type-1 intermediary* to a path $\langle *x_1, \ldots, x_n \rangle$ in Φ is any class y such that either

1. $y = x_i \, : \, (1 < i < n)$; or
2. the path $\langle + \, x_i, y_1, \ldots, y_j, \ldots, y_m, x_{i+1} \rangle \in \Phi \, :$ $m \geq 1$, $1 \leq i < n-1$, and $y = y_j$.

Definition A *type-2 intermediary* to a path $\langle *x_1, \ldots, x_n \rangle$ in Φ is any class y such that either

1. $y = x_i \, : \, (1 < i < n)$; or
2. the path $\langle + \, x_i, y_1, \ldots, y_j, \ldots, y_m, x_{i+1} \rangle \in \Phi \, :$ $m \geq 1$, $2 \leq i < n-1$, and $y = y_j$; or
3. the path $\langle * \, x_{n-1}, y_1, \ldots, y_j, \ldots, y_m, x_n \rangle \in \Phi \, :$ $m \geq 1$ and $y = y_j$.

Figure 2 illustrates exactly where an intermediary of either type can be situated along a path. Both type-1 and type-2 intermediaries can be viewed, given the assumption of redundancy of links, as reasons for the support of the conclusion $\langle *x_1, x_n \rangle$ from the appropriate path. In the examples, it was seen that support for this conclusion should be withdrawn if a type-1 y exists such that $y \not\longrightarrow x_n$ holds, or if a type-2 z exists such that $x_1 \not\longrightarrow z$ holds.

Now that the reasons for disabling the support for a conclusion have been discussed, a definition of supported sentences can be presented along with the preference relation for models. Recall that meta-variable $*$ ranges over $\{+, -\}$, while $\overline{*}$ can be any of the three signs. Let Φ be any set of links.

Definition A link $\langle * \, x_1, x_n \rangle$ is *supported in* Φ *by the path* $\phi = \langle * \, x_1, \ldots, x_n \rangle$ iff

[4] E.g. $\langle + \, x, y \rangle$ has two complements, $\langle - \, x, y \rangle$ and $\langle \# \, x, y \rangle$, while the complement of $\langle \# \, x, y \rangle$ is undefined.

1. $\phi \in \Phi$,
2. There exists no type-1 y to ϕ in Φ such that $\langle \overline{*}\, y, x_n \rangle \in \Phi$, and
3. There exists no type-2 z to ϕ in Φ such that $\langle \overline{+}\, x_1, z \rangle \in \Phi$.

A link λ is *supported in* Φ iff there exists a path $\phi \in \Phi$ which supports λ.

Definition The set of *contradicted links* in Φ is

$CONTRA(\Phi) = \{\overline{\lambda} : \overline{\lambda} \in \Phi$ and λ is supported in $\Phi\}$.

In other words, if a link λ is supported in Φ, its complement links are contradicted (if either is in Φ).

Definition An I-model M_1 is *as preferable as* an I-model M_2 (written $M_1 \leq M_2$) iff

$$CONTRA(M_1) \subseteq CONTRA(M_2).$$

M_1 is *preferred to* M_2 (written $M_1 < M_2$) iff $M_1 \leq M_2$ and not $M_2 \leq M_1$.

The preferred models of a network are those models which satisfy the links of the network and are minimal in the relation $\leq$. The nonmonotonic consequences of a theory, or network, are precisely those links true in all preferred models. Essentially, the preferred models of a network are those which satisfy as many supported links as possible, where a link is supported if there is a "chain of reasoning" (path) which leads to that conclusion, and no more specific evidence (intermediaries) inhibit this support.

Given the definitions here, a formal characterization of the notion of redundant link, discussed in the last subsection, can be given. The idea is that a redundant link is one which has some independent justification, consisting of a set of other links, for its truth.

Definition A link $\langle * \, x_1, x_n \rangle$ is *redundant* in a network Φ iff there exists a path $\langle * \, x_1, \ldots, x_n \rangle$ (where $n > 2$) such that the restriction (in the graph-theoretic sense) of Φ to $\{x_1, \ldots, x_n\}$ contains no contradicted links.

3 A Comparison to Existing Techniques

There are a number of different intuitions regarding the exact nature of inheritance reasoning, and these are reflected in a divergent collection of reasoning techniques. The "clash of intuitions" is described in detail in [Touretzky *et al.* 1987], and the space of options available to an inheritance reasoner is explicitly mapped out. There it is claimed that these choices represent a number of different, but equally plausible, reasoning strategies, suitable for varying tasks. We claim here that many of these design options are unintuitive, and while space prohibits a complete discussion of where the minimal model reasoner fits along this scale and a presentation of examples (see [Boutilier 1989] for details), we will show how it differs from other techniques in a crucial respect.

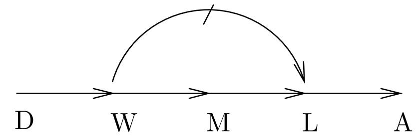

Figure 3: An Unstable Network

In the terminology of [Touretzky *et al.* 1987], our approach is *skeptical*, in that conclusions are reached only when no evidence contradicts them. Unlike the "restricted skepticism" of [Horty *et al.* 1986], this approach is truly skeptical. The reasoner performs *coupling*, it is not *opportunistic*, and it uses *on-path preemption*, all desirable qualities. As well, the reasoner is *stable*, a point to be discussed in greater detail now.

3.1 Stability and Redundancy

In (standard) logics the property of monotonicity is guaranteed by definition. Letting S and T be sets of sentences, α a sentence, and C the consequence operation of some logic, the property of monotonicity can be expressed as

$$\text{If } \alpha \in C(T), \text{ then } \alpha \in C(T \cup S).$$

Of course, this property is too strong for nonmonotonic inference systems. However, a property known as the *Cumulation Property* was proposed in [Gabbay 1985] as being a minimal requirement (among others) of any nonmonotonic reasoning system. Interpreting C as the nonmonotonic "consequence" operation of such a system, this property can be expressed as

$$\text{If } \alpha \in C(T), \text{ then } \beta \in C(T) \text{ iff } \beta \in C(T \cup \{\alpha\}).$$

This condition asserts that if a sentence is derivable from some theory, the addition of that sentence to the theory won't affect the set of consequences[5].

An inheritance reasoner is said to be *stable* iff it satisfies the Cumulation Property. In other words, a reasoner is stable if, for any network, the addition of a link representing a consequence of that network will not invalidate any consequences of the original network, nor introduce new consequences. While stability is a seemingly incontrovertible property of inheritance reasoners, reasoners which use inferential distance (ID-reasoners), such as those of [Touretzky 1986] or [Horty *et al.* 1986], do not exhibit this attribute.

Consider the network in Figure 3. In this network an ID-reasoner will conclude that D's are not L's, and hence make no conclusion about D's with respect to A's.

[5] Actually, Gabbay proposed two separate rules, *Weak Monotonicity* and *Cut*, which were combined in [Besnard 1988] as the resultant Cumulation Property.

Notice that the conclusion that M's are A's is also warranted. Now consider the network augmented with "M's are A's" (sanctioned in the original network), via the explicit link $M \longrightarrow A$. In contrast to its determination in the original network, an ID-reasoner will now assert that D's are indeed A's. Even though a link has been added which was previously deducible from the network, it significantly altered the consequences of the network. The minimal model characterization of inheritance, however, is stable, almost by definition. The link $D \longrightarrow A$ is not true in all minimal models of either network, thus $D \longrightarrow A$ is not concluded in either case. The question remains whether stability makes the minimal model reasoner a more plausible approach than ID-reasoners.

It is suggested that this sort of instability may be desirable, giving inheritance reasoners an added flexibility by allowing "a sensitivity to the structure of arguments that is difficult to achieve in deductive systems"[Horty *et al.* 1986, p.21]. This claim is supported by the following interpretation of the networks. Let D stand for Moby, W for Whales, M for Mammals, L for Land-Dwellers, and A for Air-Breathers. Given the information in the first network, the reasoner cannot (and should not) conclude that Moby is an air breather. In the second network, the direct link from mammals to air breathers changes things, it is argued. While whales are still not land-dwellers, they are mammals and mammals are *directly* linked to air breathers (hence, they are air breathers independently of being land dwellers[6]). Therefore, it is claimed, the instability exhibited by the reasoner is actually quite useful because the desired conclusion is reached, namely, that Moby is an air breather.

Clearly, this argument is incompatible with the notion of redundancy assumed by the minimal model reasoner. A simple change in the interpretation of this network will perhaps illustrate the problem clearly. Let all nodes be interpreted as before except for A, which will now stand for Walking-Things. ID-reasoners will determine that Moby the whale can walk because of the explicit link from M to A. In other words, they assume the link is not redundant (i.e. it is independently justified). On the other hand, the minimal model reasoner will not conclude anything about whales (in particular, Moby) walking, since the link from M to A is assumed to be redundant. That is, mammals can walk *because* they are land-dwellers.

The key difference between the reasoners is in their treatment of redundant links. One may say that since each "works" (gives intuitive conclusions) on a different example, each is useful for different purposes. The minimal model reasoner assumes that $M \longrightarrow A$ is redundant, useful in one instance, while the others do not, helpful in the other circumstance. However, it seems that the minimal model reasoner is more principled in its choice. It assumes that all links with a certain "topology" are redundant. If it turns out that a particular network has this "shape" but its link is not redundant (e.g. $M \longrightarrow A$), then the reason for the minimal model reasoner's lack of intuitive conclusions is obvious: it wasn't designed to work on this network (which does not fit its definition of an inheritance network).

In contrast, the reasoners of [Touretzky 1986] and [Horty *et al.* 1986] are a little less clear in this respect. While the use of inferential distance makes a tacit assumption about redundancy, these systems fail to make clear which links will be taken to be redundant. As a result, they are designed to work on specific examples, some of which have redundant links, and some of which have "pseudo-redundant" links. The problem is, looking at an uninterpreted network, we cannot distinguish these types of links. While the minimal model reasoner makes its principles explicit by stating that *all* such links are redundant, the other reasoners exhibit a certain unpredictability.

3.2 Concluding Remarks

While many syntactic characterizations of inheritance reasoning exist, the semantics of these systems has remained unclear and, as a result, disparate conclusions abound. We have provided a semantic interpretation of links in these networks and a semantic characterization of inference in inheritance systems, based on minimal models. In particular, we have pointed out what appears to be a deficiency in inferential distance based reasoners which causes unstable behavior[7] and a certain lack of predictability. While ID-reasoners work well on a number of networks, the minimal model reasoner seems more principled in its design choices.

The characterization presented here is quite intuitive, and it can be extended in a number of ways. As it stands, the semantic interpretation of links and the semantic account of the inference process are somewhat distinct. The approach may be augmented to account for more "general" inheritance reasoning by extending the the preference relation to E-models, rather than I-models. In this manner, we may account for the full interpretation of networks, and characterize how information in non-network form can be integrated. Also, an account can be given of inheritable properties and relations of classes (see, e.g., [Touretzky 1986]). As ID-reasoners and the approach presented here differ with respect to stability, the minimal model characterization cannot be construed as a semantics for these forms of inheritance. Those algorithms are of no use for computing the consequences of a network, in the sense defined

[6]In fact, this is the case in the real world, which is why the results *seem* so appropriate.

[7]While inferential distance causes a violation of the "if" clause of the Cumulation Property, those reasoners (e.g. [Horty *et al.* 1986]) which use off-path preemption as well, also violate the "only if" clause, resulting in "super-unstable" behavior.

here. While consequences can be derived in a "brute force" fashion, a (clearly desirable) syntactic account of the "traditional" sort has yet to be developed.

Acknowledgements

Thanks to Bart Selman for helpful discussion and Susan McRoy and Gerhard Lakemeyer (and the referees) for commenting on an earlier draft of this paper. This research was supported by an NSERC Postgraduate Scholarship.

References

[Besnard 1988] Philippe Besnard. 'Axiomatizations in the metatheory of nonmonotonic inference systems'. *Proc. of CSCSI*, Edmonton, pp.117–124, 1988.

[Boutilier 1988] Craig Boutilier. *Default Reasoning with the Conditional Logic E*, University of Toronto M.Sc. thesis, Toronto, 1988.

[Boutilier 1989] Craig Boutilier. *The Semantics of Inheritance Reasoning*, forthcoming technical report, University of Toronto, Toronto, 1989.

[Delgrande 1987] James P. Delgrande. 'An approach to default reasoning based on a first-order conditional logic'. *Proc. of AAAI*, Seattle, pp.340–345, 1987.

[Etherington and Reiter 1983] David W. Etherington, Raymond Reiter. 'On inheritance hierarchies with exceptions'. *Proc. of AAAI*, pp.104–108, 1983.

[Gabbay 1985] D.M. Gabbay. 'Theoretical foundations for non-monotonic reasoning in expert systems'. In Krzysztof R. Apt (ed.), *Logics and Models of Concurrent Systems*, pp.439–457, Springer-Verlag, Berlin, 1985.

[Horty *et al.* 1986] John F. Horty, Richmond H. Thomason, David S. Touretzky. *A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks*, Carnegie-Mellon University Technical Report CMU-CS-87-175, 1987.

[Sandewall 1986] Erik Sandewall. 'Nonmonotonic Inference Rules for Multiple Inheritance with Exceptions'. *Proc. of the IEEE* 74(10), pp.1345–1353, 1986.

[Shoham 1986] Yoav Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*, Yale University Technical Report YALEU/CSD/RR#507, New Haven, 1986.

[Touretzky 1986] David S. Touretzky. *The Mathematics of Inheritance Systems*, Pitman, London, 1986.

[Touretzky *et al.* 1987] David S. Touretzky, John F. Horty, Richmond H. Thomason. 'A Clash of Intuitions: The Current State of Nonmonotonic Multiple Inheritance Systems'. *Proc. of IJCAI*, Milan, pp.476–482, 1987.