# A Local Formalization of Inheritance: Preliminary Report

Matthew L. Ginsberg
Computer Science Department
Stanford University
Stanford, California 94305

## Abstract

Existing work on formalizing inheritance suffers from two significant drawbacks. On the one hand, the formalization is often nonlocal, in that the translation from the inheritance hierarchy into the relevant declarative language requires a complete examination of the topology of the graph being considered. On the other hand, the formalization frequently uses an inference technique or axiomatization that is carefully tailored to problems of this sort and cannot be applied outside of this very restricted area.

In this paper we present a formalization that avoids these difficulties. It is local and relies on a very simple generalization of existing approaches to default reasoning. The generalization is based on a formalization of the notion of causality.

# 1 Introduction

Existing formalizations of inheritance reasoning suffer from two serious technical problems.

The first of these is that many of the translation schemes are *nonlocal*. By this we mean that if we have a declarative representation of some inheritance diagram, and we add a new arc to such a diagram, there is no way to tell what the declarative translation of the new arc should be without examining the remainder of the diagram. This problem eliminates one of the principal advantages of declarative approaches to reasoning – that they allow the user of a system to encode his domain knowledge in a modular way. Approaches to inheritance reasoning that suffer from this difficulty include the following:

1. Etherington and Reiter's description of inheritance in terms of seminormal defaults [1]. Here, the problem is exceptionally clear, since the seminormal defaults used need to explicitly indicate the possible exceptions to the default arcs they represent.

2. Prioritization-based approaches, such as those using prioritized circumscription [11], path length [2], partial orders on truth values [10] or inferential distance [14]. In all of these approaches, the prioritizations are determined from a global analysis of the inheritance diagram in question.

3. The autoepistemic approach developed by Gelfond in [3]. Here, a global analysis of the diagram is needed to identify conflicts between potentially competing default rules.

Other approaches avoid this difficulty by constructing reasoning techniques that can be used to solve inheritance problems but nothing else; typical here is the work of Horty and his colleagues [9, 15]. The difficulty with this approach is that there are many problems that cannot be phrased in the restricted language that Horty et. al use; an example is the one involving political motivation that is discussed in Section 4 of this paper.[1]

Our intention in this paper is to provide a new formalization of inheritance reasoning that avoids these difficulties. We will do this by attempting to capture in as direct a form as possible the causal arguments that underlie our intuitive understanding of the problem.

In order to do this, we begin in the next section by presenting a very simple formalization of nonmonotonic reasoning that we extend to deal with causal arguments of the form that arise in reasoning about inheritance; this extension is used to formalize inheritance itself in Section 3. In Section 4, we examine the results of applying the formalization to a variety of inheritance problems, showing its similarities to and differences from earlier work. Concluding remarks and suggestions for future research are included in Section 5.

# 2 Nonmonotonic reasoning

The basic formalization of nonmonotonic reasoning with which we will be dealing is the simplest one possible. We imagine that we have some base theory $T$, and some set of defaults $D$. What we'd like to do is simply to accept as true as many of the defaults as possible.[2]

**Definition 2.1** *A* default theory *is a pair $(T, D)$, where $T$ and $D$ are sets of sentences. The elements of $T$ are called the* facts *of the default theory, and the elements of $D$ are called the* defaults.

When we say that we'd like to accept as true as many elements of $D$ as possible, we are saying that we are interested in situations in which a maximal subset of $D$ is valid. We will refer to such a situation as an *extension* of our default theory:

**Definition 2.2** *An* extension *$E$ of a default theory $(T, D)$ is a subset of $D$ such that:*

*1. $T \cup E$ is satisfiable, and*

*2. If $E \subset E' \subseteq D$, then $T \cup E'$ is not satisfiable.*

[1]Gelfond's more recent work with Przymusinska [4] is also of this form. Although set in the framework of autoepistemic reasoning, the axiomatization is so complex that it seems that these authors are describing a computational procedure instead of drawing on legitimate intuitions about inheritance generally.

[2]This idea is far too simple to be new. It appears in [6, 5, 12], and can surely be found in earlier references as well.

As an example, let us consider the Nixon diamond, first introduced in [13]:

$$q$$
$$r$$
$$q \wedge \neg \mathtt{ab}_R \supset p$$
$$r \wedge \neg \mathtt{ab}_P \supset \neg p$$

Nixon is a Quaker ($q$) and a Republican ($r$). Any Quaker who is not religiously abnormal is a pacifist; Any Republican who is not politically abnormal is not. The set of defaults is given by

$$D = \{\neg \mathtt{ab}_P, \neg \mathtt{ab}_R\},$$

since we would like to assume if possible that Nixon is both politically and religiously normal.

There are two extensions here. In one we have $\mathtt{ab}_P$; in the other, we have $\mathtt{ab}_R$. Since the two default rules conflict, but either is consistent with $T$ on its own, these are the only extensions in this example.

In order to reason about causality, we will extend our declarative language by adding a binary connective $>$, where $p > q$ will mean that, "$p$ causes $q$." We now make the following definition:

**Definition 2.3** *Let $(T, D)$ be a default theory, and let $E \subseteq D$ be a subset of the defaults in the theory (note that $E$ is not necessarily an extension). We will say that a causal formula* $p > q$ holds *in $E$ if and only if the following conditions are satisfied:*

1. $p \in E$
2. $T \cup E \models q$
3. $T \cup E - \{p\} \not\models q$

In other words, $p$ causes $q$ just in case $q$ is a consequence of $E$ *because $E$ includes $p$*. Note that we have not attempted to give a semantics to causal statements whose antecedents are not defaults.

In the Nixon diamond, the extension in which Nixon is a hawk has $\neg \mathtt{ab}_P > \mathtt{ab}_R$ – Nixon is religiously abnormal because he is politically normal. This is in accord with our intuitions.

Given that Definition 2.3 defines conditions under which a sentence of the form $p > q$ holds in an extension $E$, we can obviously extend these ideas to determine whether compound sentences such as

$$(p > q) \wedge (r > s)$$

hold.

The reason this causal notion is of interest to us is that we will formalize the observation that rules about subclasses take precedence over rules about superclasses by saying that if a normality assumption about a superclass *causes* an abnormality with regard to a subclass, then the original normality assumption must have been in error. Making this informal remark precise is the principal technical objective of this paper.

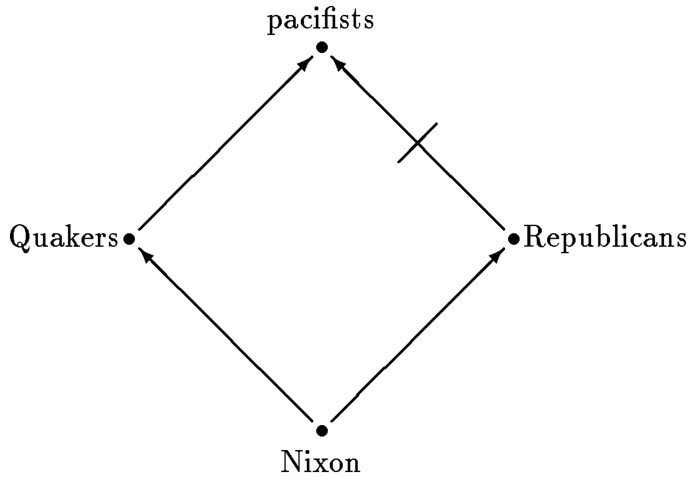


Figure 1: The Nixon diamond

# 3 Formalizing inheritance

Authors such as Horty et. al [9] work directly with inheritance diagrams such as that in Figure 1, as opposed to a declarative translation of such a diagram. Figure 1 corresponds to the Nixon diamond; the crossed arrow indicates that Republicans are in general *not* pacifists.

In order to treat diagrams such as this within our framework, we need to provide a means of translating the diagram into our declarative language, and to formalize the notions of inheritance and specificity, and the prioritization associated with them.

We translate the diagram using the two predicates is-a and subclass. Briefly, $\texttt{is-a}(x, c)$ will mean that $x$ has the property $c$, so that we might write

$$\texttt{is-a}(\texttt{Nixon}, \texttt{Quaker})$$

to indicate that Nixon is a Quaker. By $\texttt{subclass}(c, d, t)$ we will mean that $c$'s tend to be $d$'s; by $\texttt{subclass}(c, d, f)$ we will mean that $c$'s tend *not* to be $d$'s.

The declarative translation of a diagram such as that in Figure 1 is now given by the following interpretation of each arc in the diagram:

1. An uncrossed arrow originating at an object $x$ and terminating at a class $c$ will be translated as $\texttt{is-a}(x, c)$.

2. A crossed arrow originating at an object $x$ and terminating at a class $c$ will be translated as $\neg\texttt{is-a}(x, c)$.

3. An uncrossed arrow originating at a class $c$ and terminating at a class $d$ will be translated as $\texttt{subclass}(c, d, t)$. If the arrow corresponds to *monotonic* inclusion, we translate it as $\texttt{subclass}(c, d, t) \wedge \forall x. \neg\texttt{ab}(x, c, d)$.

4. A crossed arrow originating at a class $c$ and terminating at a class $d$ will be translated as $\texttt{subclass}(c, d, f)$. As above, if the arrow corresponds to a monotonic relation, we translate it as $\texttt{subclass}(c, d, f) \wedge \forall x. \neg \texttt{ab}(x, c, d)$.

Given this, the declarative translation of Figure 1 is as follows:

$$\begin{array}{l} \texttt{is-a}(\texttt{Nixon}, \texttt{Quaker}) \\ \texttt{is-a}(\texttt{Nixon}, \texttt{Republican}) \\ \texttt{subclass}(\texttt{Quaker}, \texttt{pacifist}, t) \\ \texttt{subclass}(\texttt{Republican}, \texttt{pacifist}, f) \end{array}$$

In order to progress further, we need to describe the connection between **subclass** and **is-a**. We in fact do this by formalizing the notion of a *class* and an intermediate predicate **member**, where $\texttt{member}(x, c)$ indicates that the object $x$ is a member of the class $c$.

Classes are collections of properties, where each property can appear positively or negatively. Positively-appearing properties are those required of members of the class; negatively-appearing ones are those that members of the class are required *not* to have. Thus the class

$$\langle (\texttt{Quaker}, t), (\texttt{pacifist}, f) \rangle$$

is intended to denote the class whose members *are* Quakers and are *not* pacifists. Formally, we define **member** recursively as:

$$\begin{array}{c} \texttt{member}(x, \langle \rangle) \\ \texttt{member}(x, \langle (c_0, t), \ldots, (c_n, v_n) \rangle) \equiv \texttt{is-a}(x, c_0) \wedge \texttt{member}(x, \langle (c_1, v_1), \ldots, (c_n, v_n) \rangle) \\ \texttt{member}(x, \langle (c_0, f), \ldots, (c_n, v_n) \rangle) \equiv \neg \texttt{is-a}(x, c_0) \wedge \texttt{member}(x, \langle (c_1, v_1), \ldots, (c_n, v_n) \rangle) \end{array}$$

Note that the above axioms define the **member** predicate completely. For a single property $p$, we will often abuse notation and abbreviate the class $\langle (p, t) \rangle$ to simply $p$, so that we have

$$\texttt{is-a}(x, p) \equiv \texttt{member}(x, p). \tag{1}$$

We now define $\texttt{subclass}(c, d, v)$ to be a syntactic abbreviation for the fact that $c$'s are generally $d$'s if $v = t$, and a syntactic abbreviation for the fact that $c$'s are generally *not* $d$'s if $v = f$. In other words:

$$\texttt{subclass}(c, d, t) \equiv [\texttt{member}(x, c) \wedge \neg \texttt{ab}(x, c, d) \supset \texttt{member}(x, d)] \tag{2}$$

$$\texttt{subclass}(c, d, f) \equiv [\texttt{member}(x, c) \wedge \neg \texttt{ab}(x, c, d) \supset \neg \texttt{member}(x, d)] \tag{3}$$

The appearance of the abnormality predicate here allows us to inherit properties from one class to another by assuming that as few things as possible are abnormal.

We are now in a position to derive sentences such as

$$\neg \texttt{ab}(\texttt{Nixon}, \texttt{Quaker}, \texttt{pacifist}) \supset \texttt{is-a}(\texttt{Nixon}, \texttt{pacifist}) \tag{4}$$

from the diagram in Figure 1. From $\mathtt{is\text{-}a}(\mathtt{Nixon}, \mathtt{Quaker})$ we derive $\mathtt{member}(\mathtt{Nixon}, \mathtt{Quaker})$ and thence from $\mathtt{subclass}(\mathtt{Quaker}, \mathtt{pacifist}, t)$ we obtain

$$\neg\mathtt{ab}(\mathtt{Nixon}, \mathtt{Quaker}, \mathtt{pacifist}) \supset \mathtt{member}(\mathtt{Nixon}, \mathtt{pacifist}).$$

In light of (1), this becomes (4).

Note that the above axiomatization already extends conventional work on inheritance by allowing us to discuss classes such as "red birds" or "nonflying birds" without inventing new nodes in a particular diagram.

In order to discuss specificity, we introduce a new predicate **path**. Informally,

$$\mathtt{path}(x, \langle c_0, \ldots, c_n \rangle)$$

will mean that there is a path for the object $x$ from the class $c_0$ to the class $c_n$ for which none of the inheritance links is blocked by an abnormality.[3]

As with **member**, **path** is defined recursively. For $n = 0$, we have:

$$\mathtt{path}(x, \langle c_0 \rangle) \equiv \mathtt{member}(x, c_0), \tag{5}$$

while for longer paths, the definition is:

$$\mathtt{path}(x, \langle c_0, \ldots, c_n \rangle) \equiv \mathtt{member}(x, c_n) \wedge \mathtt{path}(x, \langle c_0, \ldots, c_{n-1} \rangle) \wedge \mathtt{subclass}(c_{n-1}, c_n, t). \tag{6}$$

Finally, we need an axiom saying that defaults about superclasses cannot cause defaults about subclasses to fail:

$$\mathtt{path}(x, \langle c_0, \ldots, c_n \rangle) \wedge [\neg\mathtt{ab}(x, c_n, d_n) > \mathtt{ab}(x, c_0, d_0)] \supset \mathtt{ab}(x, c_n, d_n). \tag{7}$$

In other words, if $c_0$ is a subclass of $c_n$ for an object $x$, and a normality assumption about $c_n$ causes an abnormality to occur for the class $c_0$, then $x$ must have been an abnormal member of the class $c_n$ after all.

As an example, suppose that we use this formalism to analyze the Nixon diamond appearing in Figure 1. As already noted, the declarative translation of this diagram is:

$$\begin{array}{l} \mathtt{is\text{-}a}(\mathtt{Nixon}, \mathtt{Quaker}) \\ \mathtt{is\text{-}a}(\mathtt{Nixon}, \mathtt{Republican}) \\ \mathtt{subclass}(\mathtt{Quaker}, \mathtt{pacifist}, t) \\ \mathtt{subclass}(\mathtt{Republican}, \mathtt{pacifist}, f) \end{array}$$

Since there are no paths connecting Quakers and Republicans, there continue to be two extensions. In one, Nixon is an abnormal Republican in that he is a pacifist; in the other, he is abnormal Quaker in that he is not.

[3]Note that this commits us to **path** being a 2-place relation, so that a path may fail to exist for a particular object $x$ because one of the links is blocked for $x$. If we wish to have **path** be a 1-place relation, we need simply drop the clause $\mathtt{member}(x, c_n)$ from (5) and (6). Note also that the extent of the **path** predicate cannot be determined by a local examination of the diagram in question. But this is not at all what we mean by our translation being a local one – given the axiomatization provided by (1) through (7), the translation of any particular inheritance diagram into our declarative system *is* a purely local operation.

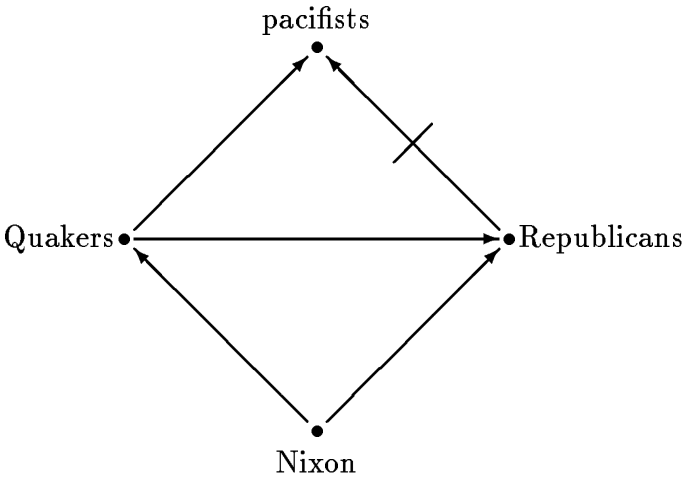


Figure 2: A modified diamond

# 4 Examples

Other examples are more interesting. Suppose that we modify the Nixon diamond by including an arc saying that Quakers tend to be Republicans, as in Figure 2. This allows us to derive

$$\mathtt{path}(\mathtt{Nixon}, \langle \mathtt{Quaker}, \mathtt{Republican} \rangle),$$

so that we can use (7) to conclude

$$[\neg \mathtt{ab}(\mathtt{Nixon}, \mathtt{Republican}, d_1) > \mathtt{ab}(\mathtt{Nixon}, \mathtt{Quaker}, d_0)] \supset \mathtt{ab}(\mathtt{Nixon}, \mathtt{Republican}, d_1),$$

or specifically

$$[\neg \mathtt{ab}(\mathtt{Nixon}, \mathtt{Republican}, \mathtt{pacifist}) > \mathtt{ab}(\mathtt{Nixon}, \mathtt{Quaker}, \mathtt{pacifist})] \\ \supset \mathtt{ab}(\mathtt{Nixon}, \mathtt{Republican}, \mathtt{pacifist}). \tag{8}$$

There are now four possible extensions, depending on the truth or falsity of the two abnormality assumptions appearing in (8).

The candidate extension in which both normality assumptions hold is inconsistent for the usual reason – we conclude both $\mathtt{is\text{-}a}(\mathtt{Nixon}, \mathtt{pacifist})$ and its negation. The extension $E_1$ in which we have $\neg \mathtt{ab}(\mathtt{Nixon}, \mathtt{Quaker}, \mathtt{pacifist})$ remains consistent, satisfying (8) because Nixon is an abnormal Republican in this extension and the conclusion of (8) therefore holds. This observation eliminates the candidate extension in which *neither* normality assumption holds, since it is nonmaximal.

Finally, consider the candidate extension $E_2$ in which Nixon is politically normal. As we have already observed,

$$\neg \mathtt{ab}(\mathtt{Nixon}, \mathtt{Republican}, \mathtt{pacifist}) > \mathtt{ab}(\mathtt{Nixon}, \mathtt{Quaker}, \mathtt{pacifist})$$

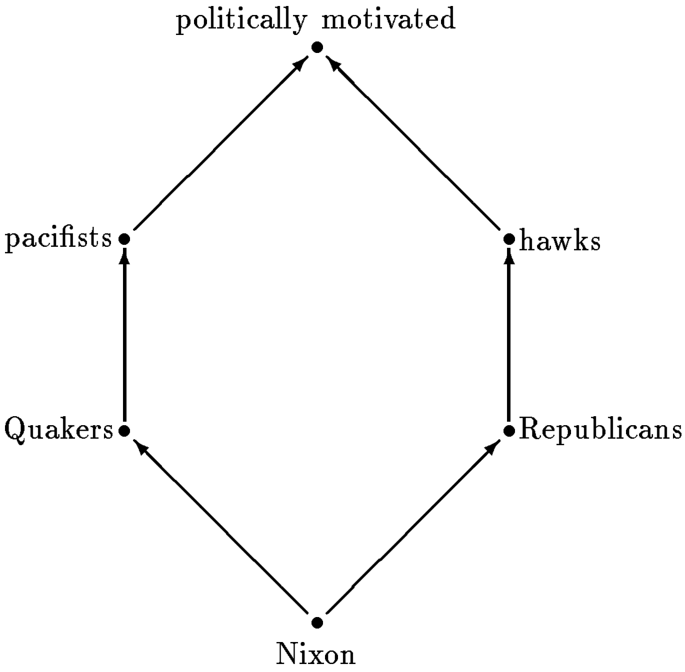


Figure 3: An extended diamond

holds in $E_2$, so it fails to satisfy (8) and is therefore inconsistent. As a result, we are left with the single extension $E_1$ in which Nixon is religiously normal and therefore a pacifist.

The translation we have described embeds the translated diagrams in a full declarative system. As mentioned in the introduction, an advantage of this is that we are able to deal with some situations that are troublesome for previous approaches, such as the diagram in Figure 3.

What we would like to do is to augment this diagram with a rule saying that the sets of hawks and pacifists are disjoint:

$$\texttt{is-a}(x, \texttt{hawk}) \supset \neg\texttt{is-a}(x, \texttt{pacifist}). \tag{9}$$

If we do this, we once again get two extensions, depending upon whether Nixon is a hawk or a pacifist; he is politically motivated in both. Some previous authors cannot handle this problem because the associated inheritance diagram is not acyclic.[4]

Figure 4 shows another distinction that stems from the declarative nature of our approach, based on an example due to Haugh [8]. We would treat the two diagrams identically, since they can be transformed into one another by contraposing the arc in the left-hand diagram indicating that Republicans are not pacifists. In both cases, we conclude that there are three extensions, depending on the point at which we break the inference chain.

[4]The single axiom (9) becomes the pair of axioms $\texttt{subclass}(\texttt{hawk}, \texttt{pacifist}, f)$ and $\texttt{subclass}(\texttt{pacifist}, \texttt{hawk}, f)$ and therefore generates a cycle.

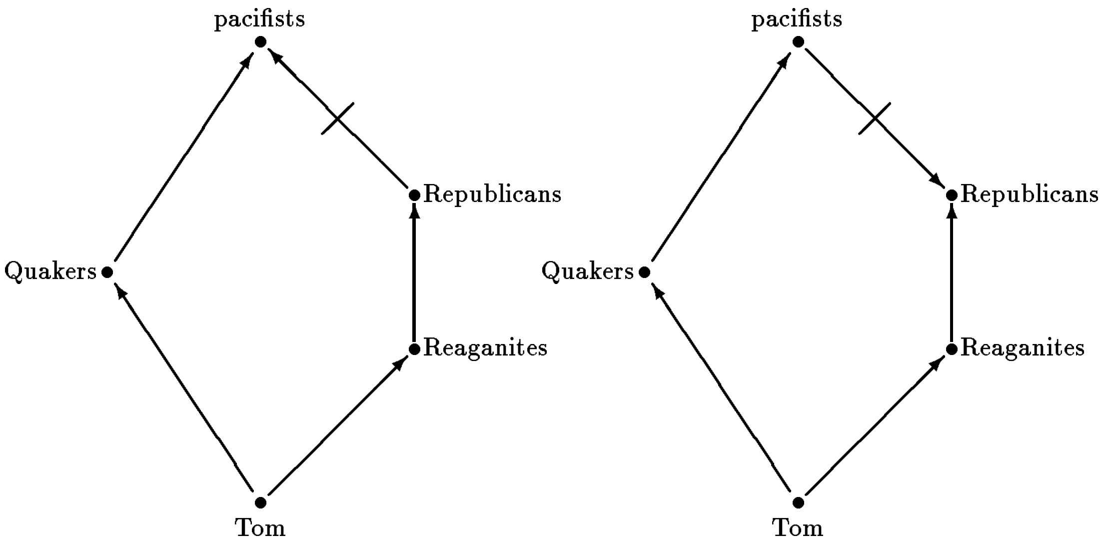


Figure 4: Equivalent diagrams?

In all of the extensions, Tom is a Quaker and supports Reagan. In the first extension, $E_1$, he is a pacifist and an abnormal Republican. In $E_2$, he is a Republican, a hawk, and an abnormal Quaker. In $E_3$, he is a pacifist, a Democrat by virtue of being a pacifist, and an abnormal Reaganite.

The situation is very different if we take Horty's approach (or the one suggested by Haugh in [8]). Here, $E_1$ and $E_2$ are possible extensions for the diagram on the left, and $E_2$ and $E_3$ are possible extensions for that on the right. We find our results more intuitive, although others might well disagree.[5]

# 5 Conclusion

We conclude with some very brief remarks about possible extensions to our ideas.

The most important extension involves dropping the condition in Definition 2.3 that the $p$ in $p > q$ be a default as opposed to either another sentence or a compound formula. This can be handled by viewing $>$ as a modal operator in the sense of [7]; the definitions appearing there generalize those of Section 2.

The work in [7] allows us to view the connective appearing in $p > q$ as a function on the truth values of $p$ and $q$ respectively, where these truth values are taken to be functions from the subsets of $D$ into $\{t, f\}$ – in other words, the truth values give the truth or falsity of $p$ and $q$ as a function of extension, and not in any absolute sense.

[5]The reason we find our results more intuitive is that we find contraposition itself intuitively appealing. In those cases where contraposition introduces additional extensions, they can generally be disambiguated using (7).

Viewing $>$ in this way brings out the fact that Definition 2.3 is capturing a semantic notion and not a syntactic one. Even though the description appearing in Section 2 is syntactic, the material in [7] allows us to understand it in semantic way.

The work in [7] also describes an implementation of the ideas we have presented. This implementation has been applied to inheritance using the axiomatization discussed in this paper, and the results are as we have described.

In summary, the approach we have described is different from other work in this area in two fundamental ways:

1. First, it is indeed a purely local translation, and a translation into a full declarative language.

2. Second, the translation is a fairly natural one. The predicates `is-a`, `member`, `subclass` and `path` are all defined as naturally as possible. This is in sharp contrast with much of the other work in this area, which all too often seems to lead the reader into a quagmire of rules and definitions that are poorly motivated and difficult to follow.


# Acknowledgement

This work has been supported by General Dynamics and by NSF under grant number IRI89-12188. I would like to thank Andrew Baker, John Horty, T. Krishnaprasad, Vladimir Lifschitz, Eunok Paek, David Smith and Rich Thomason for many enlightening discussions.


# References


[1] D. W. Etherington and R. Reiter. On inheritance hierarchies with exceptions. In *Proceedings of the Third National Conference on Artificial Intelligence*, pages 104–108, 1983.

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

[3] M. Gelfond. Autoepistemic logic and formalization of common-sense reasoning. In *Proceedings 1988 Non-monotonic Reasoning Workshop*, Munich, 1988. American Association for Artificial Intelligence.

[4] M. Gelfond and H. Przymusinska. Formalization of inheritance reasoning in autoepistemic logic. *Fundamenta Informaticae*, 1990. To appear.

[5] M. L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–79, 1986.

[6] M. L. Ginsberg. A circumscriptive theorem prover. *Artificial Intelligence*, 39:209–230, 1989.

[7] M. L. Ginsberg. Bilattices and modal operators. *Journal of Logic and Computation*, 1, 1990.

[8] B. Haugh. Tractable theories of multiple defeasible inheritance in ordinary nonmonotonic logics. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 421–426, 1988.

[9] J. F. Horty, R. H. Thomason, and D. S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 358–363, 1987.

[10] T. Krishnaprasad, M. Kifer, and D. S. Warren. On the declarative semantics of inheritance networks. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1099–1103, 1989.

[11] J. McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[12] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[13] R. Reiter and G. Criscuolo. On interacting defaults. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 270–276, 1981.

[14] D. S. Touretzky. Implicit ordering of defaults in inheritance systems. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 322–325, 1984.

[15] D. S. Touretzky, J. F. Horty, and R. H. Thomason. A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 476–482, 1987.