A Modest, but Semantically Well Founded, Inheritance Reasoner

Fahiem Bacchus*
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
N2L-3G1

Abstract

A modest exception allowing inheritance reasoner is presented. The reasoner allows restricted, but semantically well founded, defeasible property inheritance. Furthermore, it gives a well defined and easily understood semantic interpretation to all of the assertions encoded in it.

The semantics allows a knowledge engineer to decide what knowledge can be encoded in the system, and gives him understandable formal guarantees about the quality of the conclusions that will be generated. For this reason the system is a more practical, usable inheritance reasoner than others that have appeared in the literature. The system has been fully implemented in a short (< 75 lines) Prolog program which executes all the examples presented, among others.

Furthermore, although the system performs a restricted form of inheritance reasoning it can still represent and solve most of the inheritance "puzzles" that have appeared in the literature, including the recent heterogeneous inheritance problems.

1 Introduction

This paper presents an exception allowing inheritance reasoner which uses a semantically well defined notion of defeasible typicality. The system is heterogeneous, i.e., it allows both strict and defeasible assertions and has a inference engine which takes into account the differing semantic properties of these two types of assertions. The inferences that it can generate are carefully divided onto deductive and inductive inferences. It is shown that the deductive inferences are sound with respect to the given semantics. The inductive inferences cannot, of course, be shown to be sound (else they would be deductive!); however, a reasonable and intuitive semantic justification can be given for these inferences. Hence, there is some global guarantee imparted to all conclusions generated by the system.

*This research was supported by a grant from the University of Waterloo, and by NSERC grant #41848

The system gives defeasible typicality assertions a particular semantic interpretation. In order to justify this interpretation we start with a brief discussion of the nature of typicality assertions. Then we describe the system itself, giving the syntax of the expressions that we can encode in the system and their formal semantic interpretation. After this the inference procedure is presented. This procedure is broken into two parts, a deductive component, which we show to be sound, and an inductive component for which we give a semantic justification. We end the description of the system with some examples of its operation. We make reference to, and draw comparisons with, other inheritance reasoners in various places.

2 Types of Typicality

Many different inheritance reasoners have appeared in the literature, e.g., [1, 2, 3, 4, 5, 6, 7, 8]. This work has addressed the issue of exception allowing inheritance reasoning. For example, the defeasible inference that Clyde is gray since he is an elephant and elephants are typically gray. As has been discussed in the literature, such typicality assertions cannot be modeled as universally quantified assertions. There may be particular elephants, or even entire subclasses of elephants, that are not gray. These exceptional cases would falsify a universal but do not invalidate the typicality assertions.

Clearly then, typicality assertions do not have the same semantics as universally quantified assertions. A natural question is then: what is the semantics of these typicality assertions? For example, what does the assertion "elephants are typically gray" mean about the relation between the set of elephants and the set of gray objects?

Such defeasible typicality assertions are a subclass of a wider class of assertions known as generics. Specifying the exact meaning of generics is a complex problem and work on this problem has appeared (e.g., Schubert and Pelletier [9, 10], Carlson [11]). This work has revealed that such statements can possess rather complex intensional as well as extensional meaning. That is, the meaning of such statements cannot be *fully* captured simply by a relationship between the sets of objects that are the extensions of the predicates.

This poses a difficulty for designers of defeasible inheritance systems. Without a well defined semantics for the typicality assertions it is very difficult, maybe even impossible, to design an inheritance reasoner which will give *justifiable* inferences in all cases. Note, we do not mean that the system should give correct or sound inferences in all cases. This is clearly impossible—we are dealing with defeasible inference. In some cases these inferences will be wrong. However, even if we may occasionally be wrong we would still like to have some global justification for all of the inferences that the system makes. That is, we want some reason to believe that the inferences made are reasonable, or rational. It is very difficult to see how we can give any global justification to our inferences without knowing exactly what our knowledge means.

A symptom of this problem has already been noted by Touretzky et al. [12]. They noted that in many inheritance systems seemingly reasonable inference procedures produced questionable conclusions in certain cases. It can be argued that these anomalies are a result of using the same inference procedure on assertions which have different meanings. That is, it would seem that these systems can deal with certain types of typicality assertions but not with other types.

For example, Brachman [13] has pointed out that there is a difference between prototypical properties, which are characteristic of a kind, and properties which typically apply to instances of a kind. For example, "birds lay eggs" is a prototypical property of birds, but we would not want to assume by default that a given bird was an egg-layer. The property "birds fly," on the other hand, is one that we can reasonably assume is possessed by a given bird.

Most work in inheritance systems has ignored this issue. Instead it has been assumed that the typicality assertions represented in them will be of the "right" type. That is, it is assumed that no one will encode assertions like "birds lay eggs" in the inheritance net. If they did they would end up with the conclusion that Tweety lays eggs once it is asserted that Tweety is a bird. This conclusion is clearly not reasonable unless we have some other information that makes it reasonable to assume that Tweety is a female bird.

But how is a user of these systems to know if the typicality assertions he wishes to encode in the system are of the "right" type? Unfortunately, this is not so easy since these systems do not give any precise definition of what is the "right" type. That is, these systems do not give any understandable semantic interpretation to the typicality assertions.[1] Hence, the user has to be content with possible "clashes of intuition."

The system proposed in this paper takes a conservative approach. Given that the full semantics of typicality assertions is very complex and as yet not fully understood, we choose to capture only a particular type of typicality assertion, a type which can be given a clear semantics. And we will examine what kinds of inferences can be justified from this semantics.

As a result there will be typicality assertions that cannot be dealt with by the proposed system; we are willing to be less general in order that we can be more confident in our conclusions. However, it turns out that the system is still able to perform most of the inheritance reasoning that has been put forward as reasonable for an exception allowing inheritance system. This is perhaps not too surprising since, as noted above with the "birds lay eggs" example, inheritance reasoning seem only to apply to a limited subset of typicality assertions anyway.

The system gives a very precise and understandable interpretation to all of the knowledge encoded in it. Using this formal interpretation (semantics) we can give clear cut justifications to the conclusions generated by the system. Hence, a user can decide whether or not the knowledge for his application fits the particular interpretation given by the system, and if it does he will be able to use the system and will have certain guarantees on the reasonableness of the conclusions generated by the system.

The system interprets the defeasible typicality assertions as being statistical assertions. For example, it interprets the assertion "birds fly" as meaning that most birds fly.[2] It will be shown how such a statistical fact can be used to justify the inference that a given bird can fly if we do not have any knowledge about what type of bird it is. Clearly many typicality assertions do have a statistical interpretation. This was noted in early work by Rieter and Criscuolo [6], and also in work on generics which indicates that a statistical interpretation is part of the meaning of such assertions [9]. Although the statistical interpretation was considered and rejected by Rieter and Criscuolo, and also later by Sandewall [4], it would seem that this rejection was premature, since this system can perform a wide range of inheritance reasoning.

Geffner and Pearl [1] as well as Neufeld and Poole [8] have both considered probabilistic versions of inheritance. However, neither has used the statistical majority interpretation used here. Geffner and Pearl use probabilities infinitesimally close to 1 and 0. This means that their semantics provides no guidance to a user in deciding if his knowledge fits the interpretation used by their system. Clearly in the real world no properties are actually related via infinitesimal probabilities. Neufeld and Poole interpret typicality assertions as meaning that the unconditional probability is less than the conditional probability. These semantics are can be understood in terms of what it asserts about the world, but it seems to be more related to prototypical assertions, e.g., their

[1]Some systems possess no semantics at all, while others give purely formal semantics, e.g., lattice based semantics [2]. The problem with such purely formal semantics is that it gives no guidance for deciding if our intuitive understanding of a particular typicality assertion matches its formal semantic interpretation.

[2]Here, I take this to be a simple statistical fact about birds. Nutter [14] gives an unconvincing argument that during nesting season there are more non-flying birds than flying birds. Clearly, not every species of bird nests at the same time; so I rather doubt that there is any time of the year when there are more non-flying birds than flying birds. However, her argument does point out that there are various temporal dependencies which can play a role. We do not deal with temporal considerations here, but we can recognize that this is an important direction for future research.

system will sanction the inference "Tweety lays eggs."

Now we present the details of the system.

3 Syntax and Semantics of the Encoded Knowledge

3.1 Syntax

We allow two types of user definable symbols, constant symbols (a, 6, *Tweety,* ...) and predicate symbols (P, R, *elephant,* ...). Along with these are the following logical symbols, '=>' (all are), —► ' (most are), and '-»' (negation). From these symbols we can generate formulas in the following manner:

1. If 'c' is a constant symbol and 'P' is a predicate symbol, then 'c => P' and 'c => -P' are both valid formulas.[3] The first formula corresponds to the atomic assertion that 'c' has property 'P', e.g., *clyde => elephant* while the second corresponds to its negation.
2. If 'P_1' and 'P_2' are predicate symbols, then *'L1 => Li'*, and *'L\ —► L2',* are valid formulas, where *Li* is either Pi or -Pi.

3.2 Semantics

A model, .M, of the inheritance knowledge consists of the following triple:

$$\mathcal{M} = \langle \mathcal{D}, \mathcal{R}, \mu \rangle,$$

where D is a set of individuals, and R is a collection of sets of individuals. Each set in R represents a set of individuals which share a certain property, e.g., the set of birds, or the set of flying objects. Finally, u represents a probability distribution over the field of subsets generated from the collection of sets in R.[4]

3.3 Semantics of Formulas

Given some model, *M.,* and some set of user defined symbols we define an interpretation function, a, that maps the symbols onto semantic entities and assigns truth values to the formulas. In particular, *a* maps every constant symbol 'c' to an element in D, c^a, every predicate symbol 'P' to an element in P, P^a, i.e., a set of individuals, and every negated predicate symbol '(-.P)' to the complement of 'P^a',i.e., *V - Pa.*

We use the notation *Li* to denote any predicate symbol *Pi* or its negation -Pi. With *a* defined on the symbols we can assign truth to the formulas as follows:

1. **$c \Rightarrow L$ is true iff $c^\sigma \in L^\sigma$.**
2. **$L_1 \Rightarrow L_2$ is true iff $L_1^\sigma \subseteq L_2^\sigma$. So '$\Rightarrow$' represents set containment (all are).**
3. **$L_1 \rightarrow L_2$ is true iff $\mu(L_1^\sigma \cap L_2^\sigma) \div \mu(L_1^\sigma) \geq$ c, where is some constant greater than 0.5. So '$\rightarrow$' represent the fact that "most" of the set L_1^σ is in the set L_2^σ where "most" means at least 100c% of the measur of L_1^σ.**

The system's behavior does not depend on the actual value of c, as long as it is greater strictly than 0.5.

4 Inference

Deduction

The inference performed by the system falls into two parts, deductive and inductive. Given the semantics of the formulas there is a large amount of monotonic deductive inference that can be performed. The conclusions generated by deductive inference have the advantage that they are guaranteed to be true if the original knowledge encoded in the system is true. That is, the deductive inferences are sound. The following rules specify the deductive component, where c is any constant symbol and the Li's are any predicate symbols or their negations:

1. **We can deduce $L \Rightarrow L$ for any predicate symbol or its negation. Similarly, we can deduce $\neg\neg L \Rightarrow L$ and $L \Rightarrow \neg\neg L$.**
2. **From $\{c \Rightarrow L_1, L_1 \Rightarrow L_2\}$ deduce $c \Rightarrow L_2$. Similarly, from $\{L_1 \Rightarrow L_2, L_2 \Rightarrow L_3\}$ deduce $L_1 \Rightarrow L_3$.**
3. **From $\{L_1 \Rightarrow \neg L_2\}$ deduce $L_2 \Rightarrow \neg L_1$.**
4. **From $\{L_1 \rightarrow L_2, L_2 \Rightarrow L_3\}$ deduce $L_1 \rightarrow L_3$.**

As an example of the inferences that can be performed by these rules suppose we have *"royal..elephant => elephant"* as part of our knowledge base. By rule 1 and rule 2 we can deduce *royal.elephant* => ¬¬*elephant*, then by rule 3 we can deduce ¬*elephant* => ¬rot/a/*_elephant.* That is, the rules include the rule of contraposition.

Theorem 4.1 *These are sound rules of inference. That is, under any interpretation if the premises are true so must be the conclusions.*

Proof 1. Obvious.

2. A member of a subset is a member of the superset. The relation '$\subseteq$' is transitive.
3. The complement of a subset must include everything outside of of the superset. That is, the complement of the superset is a subset of the complement of the subset.
4. L_2^σ is a subset of L_3^σ. Hence, any member of L_1^σ that is in L_2^σ must also be in L_3^σ. The relative measure of these elements of L_1^σ is at least c. Therefore the relative measure of the elements of L_1^σ that are in L_3^σ must also be at least c. By the semantic definition we have that $L_1 \rightarrow L_3$ must be true under σ.

Induction

Deductive, monotonic, inference is not enough to generate conclusions like "Clyde is gray" from information like "Clyde is an elephant" and "most elephants are gray." Here we are concluding that an individual, Clyde,

[3]Note, 'c —► P' and 'c —► - P ' are not valid formulas. That is, the base set of properties that c possesses are assumed to be known with certainty.

[4]This field of subsets is the smallest collection of sets which contains 72., is closed under intersection, union, and complementation with respect to D, and contains *V.* Such a field is the minimum structure over which a probability distribution can be denned.

is a member of a set, gray, based on information that Clyde is a member of another set, elephant. If we knew that all elephant were gray this conclusion would be deductively sound. However, since there are some non-gray elephants it is quite possible that Clyde is not gray. Hence, this inference is not sound, i.e., it does not guarantee preservation of truth.

There is however a reasonable justification that can be given to this inference if we know that most elephants are gray. If all that we know is that Clyde is an elephant then we can reasonably assume that to the best of our knowledge Clyde was selected at random. That is, we have no reason to believe that Clyde is anyway special, he could be any elephant from the set of elephants. If Clyde was selected at random it is more likely than not that he would be gray, since most elephants are gray. This is the basis for the first inductive rule.

Randomization From {c => *L1, L1* → *L2)* infer c —>L_1 *L2,* which we read as "*c* is defeasibly an L_2 based on c being a random L1."[5]

There are many situations, however, when we know more about the individual. We may know that Clyde is an African elephant. If we do not know anything about the proportion of African elephants that are gray we can inherit the statistical information from the superset, all elephants. This corresponds to assuming that the property "African" gives no further information about gray once we know "elephant." This is a reasonable assumption to make in the face of no further information since most properties in the world are not correlated, and it is the statistical analog of property inheritance assumptions made in more traditional non-monotonic approaches. On the other hand we may know that Clyde is a royal elephant, and we may have information that the proportion of royal elephants that are gray is very different from the proportion of all elephants that are gray. In this case we have a preference for the more specific information.

Of course, this specificity, or subset, preference appears in almost every inheritance system. It is interesting, however, to examine this preference in terms of the statistical semantics. The defeasible conclusion is generated by assuming that the individual is, to the best of our knowledge, indistinguishable from any other member of a base set. Say we have Tweety the bird, if we consider him to be indistinguishable from any other bird we have lost some information about him as a particular instance. If however we know that he is a penguin then he is not an arbitrary bird. Instead we know that he is a special kind of a bird, a penguin. So, if we were to consider him to be an arbitrary bird we would lose the information that he is a penguin. If instead we consider him to be an arbitrary penguin we still lose some information about him, but not as much: we still retain the knowledge that he is a penguin. Chosing the base class (the reference class) corresponds in a close way to deciding what knowledge is relevant to the defeasible conclusion. These notions are examined in more detail in [16].

[5]The set *L* has also been called a reference class (Kyburg [15]).

Subset Preference The defeasible inference 'c —$L_{X\,1}$ *L'* supersedes the inferences **'$c \rightarrow_{L_2} P$' and '$c \rightarrow_{L_2} \neg P$' if $L_1 \Rightarrow L_2$, where L is either P or $\neg P$.**

Finally, it may be the case that we can deductively show that an individual has a property. In this case we have the guarantee of soundness. Hence, we need not consider any defeasible inferences.

Certainty Preference The deductive inference **'$c \Rightarrow$** *V* supersedes the inferences '$c \rightarrow_{L_1}$ P' and '$c \rightarrow_{L_1}$ ¬P' for any *L\,* where *L* is either P or -iP.

The inferences made about relationships between properties are strictly deductive. There is no ambiguity. Let Pi and P2 be the two properties. We can either deduce $L_1 \Rightarrow L_2$, $L_1 \rightarrow L_2$, $L_2 \Rightarrow L_1$, $L_2 \rightarrow L_1$, or no relation.

The inferences about the relation between an individual c and a given property P are more complex. The combination of deductive and inductive inferences can leave us in four different situations. Let *L* be P or -iP.

1. *c* => *L.*
2. We may be left with a set of agreeing defeasible inferences none of which is superseded. That is, a set of inferences *c* —+*Li* *L* with different *L{* but with *L* fixed. In this case we conclude that the knowledge supports the defeasible inference *c* is an *L.*
3. We may have a set of disagreeing defeasible inferences none of which is superseded. That is, a set of inferences which includes c —►Li *P* and c —►£„ -»P for some *L\,* *Li-* In this case we conclude that our knowledge is ambiguous about c's relation to P.
4. We may be unable to derive any relation between *c* and P. In this case we conclude that we have no knowledge about the relation between *c* and P.

One final point about the system is that it does not sanction inheritance down more than one "—►" link. An examination of the semantics shows that such multiple defeasible inheritance can never be justified under this semantic interpretation. For example, we may have 99% of all P_0's being Pi's and 99% of all P^s being P_2's and still have no Po's being P2's. However, as the examples will show this limitation does not stop the system from performing a large amount of useful inheritance reasoning.

5 Examples

Example 5.1 *It is ambiguous whether or not Nixon is a hawk or a dove, but he is probably politically motivated. (Ginsberg **

From the knowledge *{Nixon => republican, Nixon => quaker, republican —► hawk, quaker —♦ dove, hawk => ->dove, hawk => politically .motivated, dove => politically-motivated},* the system can produce the following conclusions, among others.

1. *republican —► ->dove,* (Rule 4). *dove => -^hawk,* (Rule 3). *-^politically-motivated => -*hawk,* (Rule 1, 2, and 3) From *-ihawk* nothing can be inferred about *dove* or *-^dove.*

2. *Nixon —'republican hawk* and *Nixon →quaker ¬hawk* are the only unsuperseded inferences about *Nixon* and *hawk.* That is, based on Nixon being a republican we can defeasibly conclude that he is a hawk, while based on him being a quaker we can conclude that he is not a hawk. Neither inference is superseded by the other. Hence, our knowledge is ambiguous about whether or not Nixon is a hawk. Similarly for Nixon being a dove.

3. *Nixon —republican politically.motivated* and *Nixon —>quaker politically.motivated* are the only unsuperseded inferences about *Nixon* and *politically-motivated.* Since all of the inferences agree we conclude that our knowledge supports the defeasible inference that *Nixon* is *politically-motivated.*

Example 5.2 *Hermann the Pennsylvania Dutch speaker. (Horty and Thomas on [7]).*

From the knowledge

$$\{ \quad Hermann \Rightarrow pdutch\_speaker, \quad pdutch\_speaker \Rightarrow German\_speaker, \quad pdutch\_speaker \rightarrow Pennsylvania\_born, \quad Pennsylvania\_born \Rightarrow usa\_born, \quad German\_speaker \rightarrow \neg usa\_born \quad \},$$

which says that Hermann is a Pennsylvania Dutch speaker, Pennsylvania Dutch speakers are German speakers (since Pennsylvania Dutch is a dialect of German), most Pennsylvania Dutch speakers are born in Pennsylvania, every one born in Pennsylvania is born in the U.S.A., and most German speakers are not born in U.S.A., the system can generate the following conclusions.

1. *¬usaJborn* =>¬ *Pennsylvania-born,* i.e., no one who is not born in the U.S.A. can be born in Pennsylvania.

2. *Hermann —+German speaker ¬usaJ>orn* and *Hermann →dutch speaker usaJborn* are the inferences that relate *Hermann* to *usaJborn.* However, we also have *pdutch speaker => German speaker]* hence by the subset preference, we are only left with *Hermann —>pdutch-»pcaker usaJborn.* Therefore we conclude that our knowledge supports the defeasible inference that *Hermann* is *usaJborn.*

Horty and Thomason's system [7] also generates this last conclusion (although not the first). However, they give no semantic reason why their system sanctions this inference, rather it comes about from the way they designed their inferential calculus. The essential difficulty is that defeasible links are not given any semantic interpretation in their system. This system, on the other hand, gives an easy justification for this conclusion. If *Hermann* was considered to be an arbitrary *Germanspeaker* he would likely be *-^usaJborn.* However, we know more about *Hermann,* namely that he is a speaker of a special dialect of German, *pdutch]* so it is not reasonable to consider him to be an arbitrary *Germanspeaker.* We know nothing more about *Hermann* beyond the fact that he is a *pdutch speaker]* so in the face of a lack of any other knowledge it is reasonable to consider him to be an arbitrary *pdutchspeaker,* in which case it is likely that he is *usaJborn.*

Example 5.3 *Most birds are not penguins.*

Prom the knowledge *{bird —♦ flier, penguin => bird, penguin => -• flier},* the system can conclude *bird —♦ -^penguin.* This is another example of the surprising power of this simple reasoner.

6 Conclusions

We have presented a very simple inheritance reasoner that has a number of important features. Foremost among these features is a clear semantic commitment to a particular interpretation of defeasible typicality. On the surface it would appear that such a commitment restricts the generality of the system. However, it turns out that because this particular interpretation can be treated so completely the resulting system is in some ways more general that other inheritance reasoners.

The only reason that we can give such a complete treatment of the reasoning possible under our particular interpretation is that clear formal semantics have been provided. These semantics have been used to guide the creation of a inferential calculus, rather than the much more difficult opposite approach of inventing a calculus and then searching for meaning. The success of the system raises questions about exactly what kinds of inferences other more complex inheritance reasoners are trying to capture. And indicates that these systems might benefit from being more explicit about the meaning of the knowledge represented in them.

Acknowledgements

Thanks to Len Schubert, Henry Kyburg, and Scott Goodwin for useful comments on earlier drafts.

References

[1] Hector Geffner and Judea Pearl. Sound defeasible inference. Technical Report CSD870058, Cognitive Systems Laboratory, U.C.L.A., Los Angeles, CA. 90024-1596, U.S.A., 1987.

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

[3] John F. Horty, Richmond H. Thomason, and David S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. In *AAAI-87,* pages 358-363, 1987.

[4] Erik Sandewall. Nonmonotonic inference rules for multiple inheritance with exceptions. *Proceedings of the IEEE,* 74(10):1345-1353, October 1986.

[5] David W. Etherington. Formalizing nonmonotonic reasoning systems. *Artificial Intelligence,* 31:41-85, 1987.

[6] R. Reiter and G Criscuolo. On interacting defaults. In *IJCAI-81,* pages 270-276, 1981.

[7] J.F. Horty and R.H. Thomason. Mixing strict and defeasible inference. In *AAAI-88,* pages 427-432, 1988.

[8] E. Neufeld and D. Poole. Probabilistic semantics and defaults. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence,* 1988.

[9] L.K. Schubert and F.J Pelletier. Problems in the representation of the logical form of bare plurals, generics, habituals and mass terms. In E. LePore, editor, *New Directions in Semantics.* Academic Press, 1988.

10] L.K. Schubert and F.J Pelletier. Generically speaking. In Chierchia and Turner, editors, *Property Theory, Type Theory, and Semantics.* Reidel, 1988.

11] Greg Carlson. Generic terms and generic sentences. *Journal of Philosophical Logic,* 11:145-182, 1982.

12] David S. Touretzky, John F. Horty, and Richmond H. Thomason. A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. In *IJCAJ-87,* pages 476-482, 1987.

13] Ronald J. Brachman. I lied about the trees. *AI Magazine,* 6(3):80-93, 1985.

14] Jane Terry Nutter. Uncertainty and probability. In *Proceedings of the 10th IJCAI,* pages 373-379, 1987.

15] Henry E. Kyburg, Jr. The reference class. *Philosophy of Science,* 50(3):374-397, September 1983.

16] Fahiem Bacchus. *Representing and Reasoning With Probabilistic Knowledge.* MIT-Press, Cambridge, Massachusetts, 1989 (forthcoming).