A Preference-Based Approach to Default Reasoning: Preliminary Report

James P. Delgrande
School of Computing Science,
Simon Fraser University,
Burnaby, B.C.,
Canada V5A 1S6
email: jim@cs.sfu.ca

Abstract

An approach to nonmonotonic inference, based on preference orderings between possible worlds or states of affairs, is presented. We begin with an extant weak theory of default conditionals; using this theory, orderings on worlds are derived. The idea is that if a conditional such as "birds fly" is true then, all other things being equal, worlds in which birds fly are preferred over those where they don't. In this case, a red bird would fly by virtue of red-bird-worlds being among the least exceptional worlds in which birds fly. In this approach, irrelevant properties are correctly handled, as is specificity, reasoning within exceptional circumstances, and inheritance reasoning. A sound proof-theoretic characterisation is also given. Lastly, the approach is shown to subsume that of conditional entailment.

Introduction

In any approach to nonmonotonic reasoning there are several principles that one would want to hold. For example, suppose that we are given that birds fly, birds have wings, penguins necessarily are birds, and penguins don't fly. We can write this as:

$$\{B \rightarrow F,\ B \rightarrow W,\ P \Rightarrow B,\ P \rightarrow \neg F\}. \qquad (1)$$

According to the principle of *specificity*, a more specific default should apply over a less specific default. Thus if we were given that P is true, we would want to conclude $\neg F$ by default, since being a penguin is a more specific notion than that of being a bird. This would be the case even if we were given that penguins normally are birds, $P \rightarrow B$. Second, one should obtain *inheritance* of properties by default. So, given that P is true, one would also want to conclude (assuming no information to the contrary) that W was true, and so penguins have wings by virtue of being birds. Third, *irrelevant* properties should be properly handled and so, all other things being equal, we would want to conclude that a green bird flies.

Unfortunately, it has proven difficult to specify an approach that achieves just the right balance of properties. In the last few years much attention has been paid to weaker systems of default inferencing, including (Del87; KLM90; Pea90; Bou92; GP92; Gol92). As discussed in the next section, while these systems may have some very nice properties, none handles all of specificity, inheritance, and relevance adequately. On the other hand, stronger systems, such as (McC80; Rei80; Moo85; Poo88), which handle inheritance and relevance well, do not explicitly deal with specificity. For example, in the naïve representation of Example 1 above in Default Logic, if P is true we still obtain an extension (i.e. a set of default conclusions) in which F is true; one is required to use semi-normal defaults to block this extension.

The approach presented here is based on the notion of *preferential model structures*, explored in depth in (Sho88), but going back at least to (McC80) in Artificial Intelligence, and with roots extending at least to (Sta68; Lew73). The idea is that (some) sentence is valid, not when it is true in all models, but when it is true in some *preferred* subset of models. In the approach presented here though, preference will be expressed not in terms of models, but rather in terms of orderings on worlds or "possible states of affairs" – thus the formalism will be phrased in a modal context. In any case, we begin with a particular weak approach to default reasoning; this supplies us with a notion of preference. If, for example, we have a default $A \rightarrow B$, then this default *prefers* a world in which $A \supset B$ is true over a world in which $A \wedge \neg B$ is true. From a set of defaults then we obtain orderings on worlds. A formula B follows by default from A just when, in each ordering, in the least (in terms of preference) worlds in which A is true, B is true also. Thus, given only the default $A \rightarrow B$, then *all* $A \wedge B$ worlds are preferred to all $A \wedge \neg B$ worlds. In particular, for example, all $A \wedge C \wedge B$ worlds are preferred to all $A \wedge C \wedge \neg B$ worlds and so B follows by default in this structure from $A \wedge C$.

Clearly though there are complicating factors in this notion of preference. For example, given that birds fly and that elephants are grey (i.e. $B \rightarrow F$ and $E \rightarrow G$), we would want to prefer a world in which $BFEG$[1] is

[1] For readability I will indicate conjunction at times by

true over one in which $BFE\neg G$ is true or one in which $B\neg FE\neg G$ is true. However we want no preference to obtain between a $B\neg FEG$ world and a $BFE\neg G$ world. Similarly, for Example 1, since the notion of penguinhood is more specific than that of birdhood, we would want a $BP\neg F$ world to be preferred over a BPF world.

Lastly, the notion of preference given by a default should be *semantic* rather than *syntactic*. That is, in Example 1, we have from $B \rightarrow W$ that (all other things being equal) worlds in which birds have wings are to be preferred to those where they do not. However, what about two worlds, one in which $B\neg PF$ is true and another in which BPF is true? The *set* of defaults is of no help here; nonetheless it *follows* in our logic of defaults that birds are not normally penguins, and so the first world is preferred to the second.

The appropriateness of the approach is argued from a number of directions. First and most importantly, the approach, like any semantic approach, (hopefully) formalises plausible and sound intuitions. Second, a number of examples are presented and are argued to be appropriately handled here. These include common "benchmark" problems as well as particularly nasty (in the author's view) examples that have not (to the author's knowledge) appeared in the literature. Lastly a proof-theoretic analogue, motivated by complementary intuitions is presented and shown to be sound with respect to the semantic formulation.

The next section discusses the background. In after this an extant logic of defaults is briefly presented, followed by the formal details of the approach. This is followed by a discussion and a brief concluding section. Proofs of theorems and further details are to be found in (Del94).

Background

Related Work

Approaches to default reasoning can be broadly characterised as falling into one of two groups: *weak* systems wherein some desirable default inferences are not obtained, and *strong* systems, wherein unwanted inferences may be obtained. Many of the earlier and better-studied systems of default reasoning fall into the "strong" category. Autoepistemic Logic (Moo85), Circumscription (McC80), Default Logic (Rei80), and Theorist (Poo88) are examples of approaches that may be overly *permissive* and that do not explicitly deal with specificity information. Again, in the naïve representation of Example 1 we obtain a set of default conclusions in which, given P, F is also true. Various modifications and restrictions have been proposed to handle such difficulties in each of these systems, but the application of these modifications is necessarily outside of the system. Without a formal theory, it is not clear if such modifications are appropriate or in any sense complete. Moreover, since defaults per se are not part of the formal system, one cannot reason *about* defaults. Thus, as an example, one could not conclude from Example 1 that birds are not normally penguins.

means of juxtaposition.

Recently, much attention has been paid to weaker systems of default inferencing. In fact, given the essential similarity among systems such as N (Del87), ϵ-entailment (Pea88) (or 0-entailment or p-entailment (Ada75)), preferential entailment (KLM90), and CT4 (Bou92), among others, it would seem that some consensus has been reached as to what should constitute a "minimal" system of default reasoning. As (Pea89) suggests with respect to 0-entailment, such systems may be taken as specifying a *conservative core* or set of inferences that ought to be common to all nonmonotonic inference systems.[2] These approaches deal satisfactorily with specificity. However, not unexpectedly, they are much too weak. In particular relevance and inheritance of properties are not handled. Hence, even though a penguin may be assumed to not fly by default (i.e. in Example 1 we only derive $\neg F$ but not F), a green bird cannot be assumed to fly by default (since it is *conceivable* that greenness is relevant to flight).

There has also been less agreement on how to strengthen these weak systems. One approach has been to assume things are as simple (or unexceptional) as possible. Again, some convergence is obtained with System **Z** and 1-entailment (Pea90), rational closure (KLM90), and CO^* (Bou92). These approaches handle relevance well. However they fail to allow full inheritance of properties; moreover (see (GP92)) they allow some unwanted specificity relations. This locus of approaches has been extended in various ways, including (GP91; GMP90; BCD$^+$93); all however suffer from one or another of the deficiencies of the original approach.

Of other approaches, (Del88) gives a syntactic strengthening of defaults using meta-theoretic assumptions; consequently, as with the strong approaches, it is difficult to formally characterise the set of default inferences. (GP92) presents another strengthening of the above-mentioned "conservative core", called *conditional entailment*. There are two difficulties with this approach: first that it is quite complex in its formulation and, second, that it does not sanction inheritance of default properties. The present approach however subsumes conditional entailment. This relation is discussed further after the approach is presented.

Examples

The example given at the outset is perhaps overly familiar; however it illustrates the principles of specificity, inheritance, and relevance. The following, involving ravens, albinoism, and blackness, is closely re-

[2]Of course this is not entirely uncontentious: (Gab85) and (LM92) suggest a weaker "core", while the logic of defaults used here is slightly stronger than Pearl's conservative core.

lated:

$$\{R \to Bl, R \wedge Al \to \neg Bl\}. \quad (2)$$

Again, specificity dictates that Bl should be concluded given R, but that $\neg BL$ should be concluded given $R \wedge Al$. Also, given R (arguably) one would want to conclude $\neg Al$.

Similar remarks apply to conflicting defaults. Consider the standard Quaker/republican example:

$$\{Q \to P, R \to \neg P\}. \quad (3)$$

Given only Q, we would want to conclude P. Given $Q \wedge R$ we would want to conclude nothing concerning P. However if we add the default $Q \wedge R \to \neg P$ then clearly, given $Q \wedge R$ we would now want to conclude $\neg P$.

The preceding examples are standard (although no extant theory of nonmonotonic reasoning appears to handle all cases). The next examples however seem to the author to be particularly nasty.

$$\{B \to W, W \to F\} \quad (4)$$

This example seems innocuous: birds have wings and winged things fly. The difficulty is that there are cases (for example, in inheritance) where we would want the default $B \to W$ to take priority over $W \to F$ even though there is no explicit conflict. In terms of preference among worlds, we would want a $(B \supset W) \wedge W \neg F$ world (where the second conditional is falsified) to be preferred to a $B \neg W \wedge (W \supset F)$ world (where the first conditional is falsified). The difficulty is that there is no specific conflict between B and W. In (Del94) we argue that this is the reason that conditional entailment fails to allow full inheritance reasoning.

$$\{A_1 \to B_1, A_2 \to B_2, A_3 \to B_3,\} \text{ but } \neg(B_1 \wedge B_2 \wedge B_3). \quad (5)$$

In this example we have some number of independent defaults (here three) which cannot be jointly applied. In (the appropriate extension to) System **Z** and related systems, all defaults are at the same "level", and so if a default is falsified, no default conclusions can be drawn.[3] In these approaches, n levels (again, here three) are wanted, corresponding to the number of possible default violations. Thus, we would want to conclude B_1 by default from $A_1 \wedge A_2$. However we don't want to conclude B_1 from $A_1 \wedge A_2 \wedge A_3$ (since by symmetry we would have to also conclude B_2 and B_3, which is inconsistent, or else revert to the notion of an extension). Note though that from $A_1 \wedge A_2 \wedge A_3$ we would want to conclude that two of the default conclusions are true, even though we don't know which:

$$(B_1 B_2 \neg B_3) \vee (B_1 \neg B_2 B_3) \vee (\neg B_1 B_2 B_3).$$

If a default is violated, we would still want to carry out default inferences. Hence, given $A_1 A_2 A_3 \neg B_3$ we would want to conclude $B_1 B_2$. Given $A_1 A_2 \neg B_2 A_3 \neg B_3$ we would want to conclude B_1.

$$\{B_1 \to C_1, B_2 \to C_2, A \to (B_1 \neg C_1) \vee (B_2 \neg C_2)\}. \quad (6)$$

The problem here is that we know that the third default is more specific than one of the first two; however we don't know which. (Alternately, in the "context" A, one of the other defaults is violated, and therefore inapplicable, but we don't know which.) This example again generalises to n defaults. In System **Z** (and so in equivalent systems and generalisations) B_1 *and* B_2 are assumed to be less specific than A; however, if *one of* B_1 and B_2 are less specific than A it doesn't seem intuitive to then assume that *both of* B_1 and B_2 are less specific than A. Or if this doesn't seem implausible here, it presumably does if we increase the number of defaults that are falsified by A.

A Logic of Defaults

We begin with a theory of defaults corresponding essentially to an extension of the "conservative core" suggested in (Pea89) for default inferences. While we could have used any of the systems cited in the section on previous work, for uniformity with the approach to be presented we will use a *conditional logic* formulation for default properties. See for example (Del87; Bou92) for a further exposition, details, etc. on the formal system.

The fundamental idea is straightforward: worlds are arranged according to a notion of "exceptionalness"; a default $A \to B$ is true just when there is a world in which $A \wedge B$ is true and, in all worlds that are not more exceptional, $A \supset B$ is true at those worlds. Thus, roughly, "birds fly", $B \to F$, is true if, in the least exceptional worlds in which there are birds, birds fly. Intuitively, we factor out exceptional circumstances such as being a penguin, having a broken wing, etc., and then say that birds fly if they fly in such "unexceptional" circumstances.

More formally, we let $\mathcal{L}$ be the language of propositional logic (PC) augmented with a binary operator $\to$. (We reserve $\supset$ for material implication.) For simplicity we restrict the language so that there are no nested occurrences of the $\to$ operator. Sentences of $\mathcal{L}$ are interpreted in terms of a *model* $M = \langle W, E, P\rangle$ where:

1. W is a set (of worlds),
2. E binary *accessibility* relation on worlds, with the following properties:

   Reflexive: Eww for every $w \in W$.

   Transitive: If Ew_1w_2 and Ew_2w_3 then Ew_1w_3.

   Forward Connected: If Ew_1w_2 and Ew_1w_3 then Ew_2w_3 or Ew_3w_2.

[3]The approach of maximum entropy (GMP90) appears to handle this example well, but is problematic for other reasons.

3. P is a mapping of atomic sentences and worlds onto $\{0,1\}$.

Truth at a world w in model M ($\models^M_w$) is as for PC, except that:

$\models^M_w A \rightarrow B$ iff there is a w_1 such that Ew, w_1 and $\models^M_{w_1} A \wedge B$ and for every w_2 where Ew_1, w_2, we have $\models^M_{w_2} A \supset B$, *or* for every w_1 where Ew, w_1 we have $\models^M_{w_1} \neg A$.

We define $\Box A$ as $\neg A \rightarrow A$ (read "necessarily A") and we define $A \Rightarrow B$ as $\Box(A \supset B)$ (read "necessarily A implies B" or "A strictly implies B").

Thus the accessibility relation between worlds is defined so that from a particular world w one "sees" a sequence of successively "less exceptional" sets of worlds. $A \rightarrow B$ is true just when (trivially) A is false at all accessible worlds, or there is a world in which $A \wedge B$ is true, and $A \supset B$ is true at all equally or less accessible worlds.

Space considerations preclude a lengthy discussion of this logic (or, indeed, any of the other "equivalent" weak systems). Suffice to say however that this system supplies us with a weak, but semantically justified, system of default inferencing: Given a set of defaults and strict implications Γ, B follows by default from A just when B is true in the least A worlds in all models of Γ. Hence (as previously discussed) from Example 1 we can conclude that a penguin does not fly, while a bird does; and if something flies then it is not a penguin. However we cannot conclude that green birds fly (since there are models in which green birds do not fly), nor can we conclude that penguins have (or *inherit*) wings. However this approach does provide us with a rich notion of specificity, and we can use this notion of specificity, as described next, to specify a system wherein relevance and inheritance are properly handled, as are the previously-described examples.

The Approach

The general idea of the overall approach is straightforward. If a default $A \rightarrow B$ is true in the original default theory, then this default *prefers* a world in which the material counterpart (viz. $A \supset B$) is true over a world in which it is false. The (weak) logic of defaults also provides us with a notion of specificity between formulas. We say that a world w_1 *is preferred* to a world w_2 just when there is a default that prefers w_1 to w_2, and, if there is a conditional that prefers w_2 to w_1, then there is a conditional that "overrides" this conditional and prefers w_1 to w_2. Thus, essentially, there is some reason to prefer w_1 to w_2, and if there is any reason to prefer w_2 to w_1 then there is a stronger reason to prefer w_1 to w_2.

More formally, a default theory T consists of a set of default and strict (necessary) conditionals. We take T to be closed under logical consequence in the logic of the previous section. Thus we will write $T \models A \rightarrow B$ to mean that $A \rightarrow B$ is true in all models of T; we sometimes also write $A \rightarrow B \in T$. Given a default theory T, we first define a *specificity* ordering on formulas, given by $\prec$:

Definition 1

$$A \prec_T B \quad \textit{iff} \quad T \models A \vee B \rightarrow \neg B \; \textit{and} \; T \models \neg\Box\neg A.$$

Since the default theory T is always understood, for simplicity I will henceforth write just $A \prec B$. The right hand side of the definition says that for every model of T, at some $A \vee B$ world, w, $\neg B$ is true, and $A \vee B \supset \neg B$ is true at all equally- or less exceptional worlds. Since worlds are consistent, this means that at w it must be that A is true, and that there are no equivalently-exceptional or less exceptional worlds in which B is true. (If there were such a world in which B was true then this would also be a least $A \vee B$ world, contradicting $A \vee B \rightarrow \neg B$.) Furthermore, there is an accessible A world.

We have that $\prec$ is irreflexive, asymmetric, and transitive; also, $\prec$ and $\rightarrow$ are interdefinable (Lew73). The following will also be convenient:

$$A \preceq B \stackrel{\text{def}}{=} \neg(B \prec A).$$

Separately, we will also deal with the full set of mappings of the set of atomic sentences $\mathbf{P}$ onto $\{0,1\}$; for simplicity we assume that $\mathbf{P}$ is finite. These mappings we will call "worlds". From the theory T, we will specify partial orders on these worlds; these partial orders will constitute our ultimate preference structure, with respect to which we will define a stronger notion of default reasoning.

Definition 2 $\mathcal{W} = \{f \mid f : \mathbf{P} \rightarrow \{0,1\}\}$.

Elements of $\mathcal{W}$ will be denoted $w, w_1, w_2, \ldots$. I will also write $w \models A$ if $A \in \mathcal{L}$ is true under the standard (PC) valuation in the mapping w.[4]

Defaults in T provide a basic preference notion on worlds, as follows:

Definition 3

For default theory T and $w_1, w_2 \in \mathcal{W}$, a conditional $A \rightarrow B$ prefers *w_1 to w_2 iff*

1. $T \models A \rightarrow B$,
2. $w_1 \models A \supset B$,
3. $w_2 \models A \wedge \neg B$.

Definition 4

For default theory T and $w_1, w_2 \in \mathcal{W}$, we have:

$$Pref(w_1, w_2) = \{A \rightarrow B \in T \mid A \rightarrow B \textit{ prefers } w_1 \textit{ to } w_2\}.$$

[4]This means that $\models$ is used ambiguously: for a logical consequence of a default theory T, and for a true sentence at a world – compare condition 1. with conditions 2. and 3. in Definition 3. Since these are distinct relations, hopefully no confusion results.

There is one difficulty with specificity orderings, and that is that they may be incomplete, in the sense that we may have $B_1 \vee B_2 \prec A$ but neither $B_1 \prec A$ nor $B_2 \prec A$ (recall Example 6). We define the set of complete (in the above sense) orderings as follows:

Definition 5
Given a default theory T, a specificity ordering is extended to a full specificity ordering *by:*

if $B_1 \vee B_2 \prec A$ and it is not the case that $B_1 \prec A$ then $B_2 \prec A$.

We are now in a position to define orderings on worlds:

Definition 6
Given a full specificity ordering, a preference ordering *$\mathcal{P} = \langle \mathcal{W}, < \rangle$ is defined as follows:*

For $w_1, w_2 \in \mathcal{W}$, we have $w_1 < w_2$ iff

1. *$Pref(w_1, w_2) \neq \emptyset$ and*
2. *for every $C \rightarrow D \in Pref(w_2, w_1)$ there is some $A \rightarrow B \in Pref(w_1, w_2)$ such that $C \preceq A$ and it is not the case that $A \preceq C$.*

$\Pi_T = \{\mathcal{P} \mid \mathcal{P}$ *is a preference ordering with respect to default theory T*$\}$.

That is, $w_1 < w_2$ iff

1. there is some conditional that prefers w_1 to w_2, and
2. for a conditional that prefers w_2 to w_1 there is a conditional that is no less specific than it and that prefers w_1 to w_2, but the converse does not hold.

Consider again Example 1:

$$\{B \rightarrow F, B \rightarrow W, P \Rightarrow B, P \rightarrow \neg F\}.$$

First, we have $B \prec P$, since we can prove in the logic that $B \vee P \rightarrow \neg P$ is a logical consequence of this theory. If we have worlds

$$w_1: \; B, P, \neg F, W \qquad w_2: \; B, P, F, W$$

then:

$Pref(w_1, w_2) = \{P \rightarrow \neg F\}$ and
$Pref(w_2, w_1) = \{B \rightarrow F\}$.

Hence $w_1 < w_2$. Things remain unchanged if instead $\neg W$ is true at both worlds.

Consider next Example 4:

$$\{W \rightarrow F, B \rightarrow W\}.$$

We have $W \preceq B$, but in the full specificity order we do not have $B \preceq W$. For

$$w_1: \; B, W, \neg F \qquad w_2: \; B, \neg W, F$$

we have:

$Pref(w_1, w_2) = \{B \rightarrow W\}$ and
$Pref(w_2, w_1) = \{W \rightarrow F\}$.

Hence $w_1 < w_2$.

We can now define the notion of a default inference based on preference orderings:

Definition 7
B follows as a preferential default inference *from A in theory T, written $A \mid\!\sim_T B$, iff*

for every $\mathcal{P} \in \Pi_T$, for every w_2 where $w_2 \models A \wedge \neg B$ there is a w_1 where $w_1 \models A \wedge B$, and $w_1 < w_2$.

First of all, preference orderings are indeed orderings:

Theorem 1
For a preference ordering $\mathcal{P}$ and $w_1, w_2, w_3 \in \mathcal{W}$:

1. *$w_1 \not< w_1$.*
2. *If $w_1 < w_2$ then $w_2 \not< w_1$.*
3. *If $w_1 < w_2$ and $w_2 < w_3$ then $w_1 < w_3$.*

Second, these orderings preserve truth in the original default theory:

Theorem 2
If $T \models A \rightarrow B$ then $A \mid\!\sim_T B$.

This then concludes the development of the semantical aspects of the approach. However, before discussing properties of this approach, we first give a proof-theoretic characterisation. For this characterisation, the central idea is that beginning with a theory T, we "appropriately" strengthen the elements of T. For example, in Example 1 we have the default $B \rightarrow F$; it would seem safe to allow also that $B \wedge Gr \rightarrow F$, or "green birds fly", since there is nothing in the theory that would make us believe otherwise. On the other hand we would not want to allow that $B \wedge P \rightarrow F$, since here there is a reason to believe that this conditional may not hold, namely that we have $P \rightarrow \neg F$. We can informally state this principle of irrelevance as:

Unless there is reason to believe that a property is relevant to the truth of a conditional, assume that it is irrelevant.

We will call a default *supported* if there is a reason to hold it, based on this notion of relevance, even though it may not be a logical consequence of T. A set of defaults Γ is supported iff every default in Γ is supported. The formal definition is straightforward, if a bit long-winded:

Definition 8
$A \rightarrow B$ is supported in a default theory T iff

1. *$T \models A \rightarrow B$, or*
2. *If Γ is supported and $\Gamma \models A \rightarrow B$ then $A \rightarrow B$ is supported, or*

3(a) *$T \models A \rightarrow A'$ but $T \not\models A' \rightarrow A$ and $A' \rightarrow B$ is supported, and*

(b) *if $T \models A \rightarrow A''$ but $A'' \rightarrow \neg B$ is supported then $T \models A' \vee A'' \rightarrow \neg A'$.*

Thus, in the first two parts, a conditional is supported if it is a consequence of T or of a set of supported conditionals. For the third part, (a) states that there is a reason to accept the conditional: for some strictly less specific formula A', we have the supported conditional $A' \rightarrow B$. For part (b), if there is also a less specific formula A'' that denies B, then this formula is strictly less specific than A'. So in this last case we could say that A' "overrides" A''.

At present, we have a "soundness" result, in that the consequent of a supported conditional follows as a preferential default inference from the antecedent:

Theorem 3
If $A \rightarrow B$ is supported in T then $A \mid\!\sim_T B$.

I believe that the converse also holds, but have yet to show a rigorous proof. Nonetheless the partial result provides a second, intuitive, indication of what default inferences may be obtained.

Discussion

Conditional entailment (GP92) was formulated in part as an attempt to reconcile what has been called here "strong" and "weak" approaches to nonmonotonic reasoning. For the approach at hand, the methodology was to formulate from first principles a notion of preference between worlds, based on an extant logic of defaults. However, as noted earlier, there are strong similarities between the systems; it proves to be the case that the default inferences sanctioned by the present approach subsume those of conditional entailment

Theorem 4 *If a proposition q is conditionally entailed by a default theory $\langle K, E\rangle$ then $E \mid\!\sim_K q$.*

In conditional entailment, defaults are arranged in partial orders. A priority order over the set of defaults $\Delta_{\mathcal{L}}$ is *admissable* relative to a default theory iff every set Δ of assumptions in conflict with a default r contains a default r' that is less than that default in the ordering. Rankings on worlds are derived from priority relations over default rules: If $\Delta(w)$ and $\Delta(w')$ are the defaults falsified by worlds w and w' respectively, then w is preferred to w' iff $\Delta(w) \neq \Delta(w')$, and for every rule in $\Delta(w) - \Delta(w')$ there is a rule in $\Delta(w') - \Delta(w)$ which has higher priority. This then is very close to Definition 6. The primary difference is that, in conditional entailment, in order to prefer w to w' there must be defaults of strictly higher priority "favouring" w. In the approach at hand, (informally) we require that there be a reason to prefer w over w', and that such a reason not obtain for the converse. That is, for Example 4 we can't show (in conditional entailment or here) that W is less specific that B; hence conditional entailment doesn't distinguish the conditionals. In the approach at hand, we can show that W is no more specific than B (i.e. $W \preceq B$), but that the converse fails to be demonstrable. Consequently (all other things being equal) defaults with antecedent B "override" those with antecedent W.

Technically this difference appears to amount to the following: In Definition 6 where we have:

> for every $C \rightarrow D \in Pref(w_2, w_1)$ there is $A \rightarrow B \in Pref(w_1, w_2)$ such that $C \preceq A$ and it is not the case that $A \preceq C$,

conditional entailment (effectively) uses:

> for every $C \rightarrow D \in Pref(w_2, w_1)$ there is $A \rightarrow B \in Pref(w_1, w_2)$ such that $C \prec A$.

As a second minor difference, there may be fewer preferential orderings in the present approach than the admissible structures of conditional entailment.

Of the examples presented earlier, all of the desired default inferences go through: green birds fly; birds that are penguins do not fly; penguins have wings; ravens are black and (by default) non-albino. Quakers are pacifists, and normally non-republican.

Example 4 was discussed earlier. For Example 5 we had three defaults that could not be simultaneously applied. Here we conclude B_1 by default from $A_1 \wedge A_2$, but not from $A_1 \wedge A_2 \wedge A_3$. In terms of (the proof-theoretic notion of) support, we have that $A_1 \wedge A_2 \rightarrow B_1$ is supported, based on $A_1 \rightarrow B_1$. However, $A_1 \wedge A_2 \wedge A_3 \rightarrow B_1$ is not supported: even though there is a reason to accept this conditional (viz. $A_1 \rightarrow B_1$) there is a reason not to accept it (since $A_2 \wedge A_3 \rightarrow \neg B_1$ is supported) that is not overridden by this conditional.

From $A_1 \wedge A_2 \wedge A_3$ we obtain, as desired, that $(B_1 B_2 \neg B_3) \vee (B_1 \neg B_2 B_3) \vee (\neg B_1 B_2 B_3)$. We also obtain default inferences in the face of denied defaults; given $A_1 A_2 A_3 \neg B_3$, for example, we conclude $B_1 \wedge B_2$.

For Example 6, we had the formulas: $B_1 \rightarrow C_1, B_2 \rightarrow C_2, A \rightarrow (B_1 \neg C_1) \vee (B_2 \neg C_2)$. Hence, essentially, in the presence of A, at most one of B_1, B_2 can be "unexceptional". From A we obtain the default conclusion $(B_1 \neg C_1) \equiv (B_2 \supset C_2)$ and so one of B_1, B_2 is guaranteed to be "unexceptional".

A final point concerns the applicability of this approach. Implicitly, defaults are "applied" wherever possible. Consequently, given a chain of defaults $T = \{A_1 \rightarrow A_2, A_2 \rightarrow A_3, \ldots, A_{n-1} \rightarrow A_n\}$ we would obtain that $A_1 \mid\!\sim_T A_n$. This may be fine for default reasoning, but it leads to unintuitive results for temporal reasoning, as has been noted elsewhere for *chronological ignorance* (Sho88). Hence this approach would appear to produce results too strong for such reasoning.

Conclusion

An approach to nonmonotonic inference, based on preference orderings between worlds, has been presented. The semantics takes as a starting point an extant theory of defaults; from this, given a default theory, we specify orderings on worlds. The original theory of defaults provides a satisfactory notion of specificity; in the orderings based on this theory, irrelevant properties are correctly handled as is reasoning within exceptional circumstances, including inheritance reasoning. Arguably the notion of a preferential default inference satisfactorily formalises intuitions concerning preferences induced by default rules. As well, the approach is shown to handle standard and non-standard examples of default reasoning. Finally, a (sound) proof theory is presented.

There are two shortcomings to the approach as presented. First a completeness result is obviously desirable. Second, computational concerns have not been addressed. Two points ameliorate this second concern: first, the goal here is to present a characterisation of default inference, and then address computational issues; second, presumably a complete proof theory will in fact indicate how an implementation may be effected.

Acknowledgements This research was funded by the Natural Science and Engineering Research Council of Canada grant A0884 and the Institute for Robotics and Intelligent Systems (IRIS) in the Canadian Networks of Centres of Excellence Program. The author was a visitor at York University and the University of Toronto while this work was being carried out. The author also gratefully acknowledges comments from the University of Toronto Knowledge Representation group and the anonymous reviewers.

References

E.W. Adams. *The Logic of Conditionals.* D. Reidel Publishing Co., Dordrecht, Holland, 1975.

Salem Benferhat, Claudette Cayrol, Didier Dubois, Jerome Lang, and Henri Prade. Inconsistency management and prioritized syntax-based entailment. In *Proc. IJCAI-93*, pages 640–645, Chambéry, Fr., 1993.

Craig Boutilier. *Conditional Logics for Default Reasoning and Belief Revision.* PhD thesis, Department of Computer Science, University of Toronto, 1992.

J.P. Delgrande. A first-order conditional logic for prototypical properties. *Artificial Intelligence*, 33(1):105–130, 1987.

J.P. Delgrande. An approach to default reasoning based on a first-order conditional logic: Revised report. *Artificial Intelligence*, 36(1):63–90, 1988.

J.P. Delgrande. A preference-based approach to default reasoning. Technical report, School of Computing Science, Simon Fraser University, 1994. in preparation.

D. M. Gabbay. Theoretical foundations for nonmonotonic reasoning in expert systems. In K. R. Apt, editor, *Proceedings NATO Advanced Study Institute on Logics and Models of Concurrent Systems*, pages 439–457. Springer-Verlag, Berlin, 1985.

Moisés Goldszmidt, Paul Morris, and Judea Pearl. A maxixmum entropy approach to nonmonotonic reasoning. In *Proc. AAAI-90*, Boston, MA, 1990.

Moisés Goldszmidt. *Qualitative Probabilities: A Normative Framework for Commonsense Reasoning.* PhD thesis, Department of Computer Science, University of California, Los Angeles, 1992.

Moisés Goldszmidt and Judea Pearl. System-Z+: A formalism for reasoning with variable-strength defaults. In *Proc. AAAI-91*, pages 399–404, Anaheim, CA, 1991.

Hector Geffner and Judea Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, 53(2-3):209–244, 1992.

G.E. Hughes and M.J. Cresswell. *An Introduction to Modal Logic.* Methuen and Co. Ltd., 1968.

S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207, 1990.

D. Lewis. *Counterfactuals.* Harvard University Press, 1973.

D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.

J. McCarthy. Circumscription – a form of nonmonotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

R.C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Mateo, CA, 1988.

J. Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proc. KR-89*, pages 505–516, Toronto, May 1989. Morgan Kaufman.

J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proc. of the Third Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 121–135, Pacific Grove, Ca., 1990.

D.L. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36(1):27–48, 1988.

R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence.* The MIT Press, Cambridge, Mass., 1988.

R.F. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell, Oxford, 1968.