An Approach to Default Reasoning Based on a First-Order Conditional Logic

James P. Delgrande

School of Computing Science,
Simon Fraser University,
Burnaby, B.C.,
Canada V5A 1S6

Abstract

This paper presents an approach to default reasoning based on an extension to classical first-order logic. In this approach, first-order logic is augmented with a "variable conditional" operator for representing default statements. Truth in the resulting logic is based on a possible worlds semantics: the default statement $\alpha \Rightarrow \beta$ is true just when β is true in the least exceptional worlds in which α is true. This system provides a basis for representing and reasoning about default statements. Inferences of default properties of individuals rely on two assumptions: first that the world being modelled by a set of sentences is as uniform as consistently possible and, second, that sentences that may consistently be assumed to be irrelevant to a default inference are, in fact, irrelevant to the inference. Two formulations of default inferencing are proposed. The first involves extending the set of defaults to include all combinations of irrelevant properties. The second involves assuming that the world being modelled is among the simplest worlds consistent with the defaults and with what is contingently known. In the end, the second approach is argued to be superior to the first.

1. Introduction

Many commonsense assertions about the real world express default or prototypical properties of individuals or classes of individuals, rather than strict conditional relations. Thus for example, "birds fly" seems to be a reasonable enough assertion, even though birds with broken wings generally don't fly, and quite probably no penguin flies. The import of "birds fly" then certainly isn't that *all* birds fly, but rather is more along the lines of "typically birds fly". The issues and problems of such "exception-allowing general statements" have of course been extensively addressed in Artificial Intelligence, most notably with the various default reasoning schemes and approaches based on various theories of uncertainty.

In [Delgrande 86] and [Delgrande 87a] another alternative was introduced. In this approach, "birds fly" is interpreted as "all other things being equal, birds fly", or "ignoring exceptional conditions, birds fly". For this approach, an operator, $\Rightarrow$, is introduced into classical first-order logic (FOL). The statement $\alpha \Rightarrow \beta$ is interpreted as "in the normal course of events, if α then β". In the resulting logic, called N, one can consistently assert, for example, that:

$(x)(Bird(x) \Rightarrow Fly(x))$, $Bird(opus)$, but $\neg Fly(opus)$;

or that:

$(x)(Raven(x) \Rightarrow Black(x))$ and
$(x)((Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x))$,

or that:

$(x)(Penguin(x) \supset Bird(x))$, $(x)(Bird(x) \Rightarrow Fly(x))$ and
$(x)(Penguin(x) \Rightarrow \neg Fly(x))$.

Thus in the first case, all birds normally fly, but *opus* is a bird that does not fly. In the second and third examples, the sentences are satisfiable while having the antecedents of the conditionals being true also.

An advantage of this approach is that one can represent and reason about defaults. Thus it is a theorem of the system that

$$\Diamond\alpha \supset ((\alpha \Rightarrow \beta) \supset \neg(\alpha \Rightarrow \neg\beta)).$$

Hence, if α is possible and $\alpha \Rightarrow \beta$ is true, then it is not the case that $\alpha \Rightarrow \neg\beta$ is true. As a second example, we have the derived rule:

If $\vdash_N (x)(P(x) \Rightarrow Q(x))$ and $\vdash_N (x)(Q(x) \supset R(x))$
then $\vdash_N (x)(P(x) \Rightarrow R(x))$.

From this it follows that we can say say that ravens are normally black; black things are not white; and hence ravens are normally not white.

This approach arguably provides an appropriate basis for representing and reasoning about statements of default properties; in particular, it is meaningful to talk about the consistency of a set of default statements. However the logic N did not — in fact *could not* — allow *modus ponens* as a rule of inference for the variable conditional. For, if it did, then in the first example above we could deduce $Fly(opus)$ and so arrive at an inconsistency. Similarly, in the second example, if we knew $Raven(opus)$ and $Albino(opus)$, then we could conclude both $Black(opus)$ and $\neg Black(opus)$.

The reason that inconsistency does not arise with the above examples is that the truth of $\alpha \Rightarrow \beta$ depends not on the present state of affairs, but on "simpler" or "less exceptional" states of affairs. Thus $Raven(opus) \Rightarrow Black(opus)$ is true if, in the least exceptional states of affairs in which *opus* is a raven, *opus* also is black. Hence in such states of affairs, exceptional conditions such as being an albino, being painted red, being in a strong yellow light, etc. are "filtered out". In this way, it is quite possible that $Raven(opus) \Rightarrow Black(opus)$ is true, even though $Raven(opus) \supset Black(opus)$ is not.

However it nonetheless seems reasonable that if we knew only that $Raven(opus) \Rightarrow Black(opus)$ and $Raven(opus)$ that we should be able to conclude "by default" that $Black(opus)$ is true. One possible way to do so is to translate assertions expressed in N into appropriate statements of some default logic for reasoning deductively about individuals. Thus, the previous formula would have the intuitively acceptable translation $\frac{Raven(x): MBlack(x)}{Black(x)}$ in the formalism of [Reiter 80].

In this paper, a second alternative for reasoning deductively about default and prototypical properties of individuals is described. Consider where we know only that

$(x)(Raven(x) \Rightarrow Black(x))$, $Raven(opus)$,
and $Has\_wings(opus)$.

Given this information we cannot deduce anything about *opus*'s blackness, simply because it is consistent with what is known that *opus* may not in fact be black. However, if we pragmatically and *a priori* decide that the world at hand is one of the least exceptional worlds consistent with what's known, and we decide also that having wings is irrelevant to blackness, then we could conclude *Black(opus)*. In terms of "states of affairs" or "possible worlds" this means that if we assume that the world being modelled is as "normal" as possible consistent with the above sentences, and that having wings is irrelevant to blackness then *Black(opus)* must be true at the world being modelled.

The next section provides an overview of related work in AI, while the following section provides a brief description of the logic *N*. Section 4 introduces the overall approach to default reasoning. Section 5 expands on this, and describes two approaches to default inferencing. Section 6 discusses some examples of default reasoning in this framework, while the last section examines what we have gained from this approach. Further details and proofs of theorems may be found in [Delgrande 87b].

2. Related Work

Much of the work in AI for dealing with defaults and prototypical properties has centred around systems of default and non-monotonic reasoning. McDermott and Doyle, for example, in their augmentation of first-order logic [McDermott and Doyle 80], represent "birds fly" by the statement:

$$(x)((Bird(x) \wedge MFly(x)) \supset Fly(x)).$$

This can be interpreted as "for every *x*, if it is true that *x* is a bird, and it is consistent that *x* flies, then conclude that *x* flies". On the other hand, in Reiter's system [Reiter 80], "birds fly" would be represented by the rule:

$$\frac{Bird(x): MFly(x)}{Fly(x)}$$

This can be interpreted as "if something can be inferred to be a bird, and if that thing can be consistently assumed to fly, then infer that that thing flies. Circumscription [McCarthy 80] permits similar inferencing; in this case, one typically circumscribes an "abnormality" predicate to minimise the number of abnormal (with respect to flight) birds.

A general limitation with these approaches is that one cannot generally reason *about* defaults. Thus in Reiter's approach, if we knew that every penguin had to be a bird and that birds normally fly but that penguins do not normally fly, there is no means within the system of concluding that birds that aren't penguins normally fly. Similarly, in most systems the assertions "penguins are birds" and "typically penguins aren't birds" can be asserted without difficulty – in Reiter's system the default rule is never applied and in McDermott and Doyle's the truth value of the formula $Fly(x)$ is independent of that of $MFly(x)$. Yet these sentences seem to be inconsistent: if every penguin must necessarily be a bird, then it certainly seems that "typically penguins aren't birds" should be false.

A second, epistemological difficulty with these approaches is that their semantics rests on a notion of consistency with a set of beliefs. Thus, in the above approaches, one would conclude that a bird flies if this does not conflict with other beliefs. However the issue of whether birds fly or not (or normally fly or whatever) is a matter that deals with birds and the property of flight, and not with particular believers. Hence the relation between birds and flight, whatever it may be, should be phrased independently of any set of beliefs. Yet, on the other hand, if all that I know is that birds normally fly and that *opus* is a bird, then it would seem reasonable to assume that, *ceteris paribus*, *opus* flies. Thus perhaps these approaches are best viewed as telling us how to consistently *extend* a belief set, rather than as *representing* the relation between, say, birds and flight.

Another approach in AI for dealing with default properties is prototype theory [Rosch 78], [Brachman 85]. Here membership in the extension of a term is a graded affair and is a matter of similarity to a representative member or prototype. Prototype theory is concerned generally with descriptions of individuals, or predicting properties of individuals. Hence such approaches appear to address a concern that is somewhat different from ours – perhaps recognising an individual as a bird, based on the fact that it flies, or alternatively, predicting whether an individual flies, given other information about it. However in the present approach we want to attribute flight as following in the normal course of events from the conditions of being a bird. In such a case, notions of typicality and resemblance to a prototype appear too weak for our requirements.

Finally, Donald Nute [Nute 86] has investigated default reasoning in a conditional logic for representing subjunctives. However his approach is limited to reasoning with a restricted set of sentences in a propositional logic.

3. A Logic for Representing Defaults

In [Delgrande 87a] a *conditional logic* [Chellas 75],[Nute 80] called *N*, for representing default statements, was presented. The language of this logic is that of FOL, but augmented with a binary connective $\Rightarrow$. The intended interpretation of $\alpha \Rightarrow \beta$ is "if α then normally β" or "all other things being equal, if α then β". In this logic one can represent statements such as "ravens are normally black" or "albino ravens are normally not black". Truth in the logic is based on a possible worlds semantics. Informally, $\alpha \Rightarrow \beta$ is true at a world if, ignoring exceptional conditions, β is true whenever α is. What this amounts to is, if we consider "less exceptional" states of affairs, then $\alpha \Rightarrow \beta$ is true just when the least exceptional worlds in which α is true also have β true.

The accessibility relation E between worlds in this system then is formulated so that Ew_1w_2 holds between worlds w_1 and w_2 just when w_2 is at least as uniform, or at least as unexceptional, as w_1. In [Delgrande 87a], the following conditions were argued to be required for the accessibility relation E:

Reflexive: Eww for all worlds 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 either Ew_2w_3 or Ew_3w_2.

The propositional modal logic corresponding to this accessibility relation is the standard temporal logic S4.3 [Hughes and Cresswell 68]; it subsumes S4 but does not subsume S5.

The language L for representing defaults has the following primitive symbols: denumerably infinite sets of *individual variables* $x, y, z, \cdots$, *individual constants* $a, b, c, \cdots$, and *predicate symbols*, $P, Q, R, \cdots$ (each with some presumed arity), together with commas, parenthesis, and the symbols $\neg$, $\supset$, $\Rightarrow$, and $\forall$. Variables and constants together make up the set of *terms*. The set of well-formed formulae (wffs) is specified in the usual fashion. Where no confusion arises, lower-case words may be used to stand for constants and capitalised words may be used to stand for predicate symbols. As usual, conjunction, disjunction, biconditionality, and the existential quantifier are introduced by definition. The symbols $\alpha, \beta, \gamma, \ldots$ will stand for arbitrary well-formed formulae of L.

Sentences of L are interpreted in terms of a *model* $M = \langle W, E, DI, V \rangle$ where W is a set, E is a reflexive, transitive and forward connected binary relation on elements of W, DI is a domain of individuals, and V is a function on terms and predicate symbols so that

1. for term t, $V(t) \in DI$,

2. for n-place predicate symbol P, $V(P)$ is a set of $(n+1)$-tuples $<t_1, \ldots, t_n, w>$ where each $t_i \in DI$ and $w \in W$.

Informally W is a set of possible worlds, E is an accessibility relation on possible worlds, and V maps atomic sentences onto worlds where the sentence is true, and predicate symbols onto relations in worlds. For wff α, the symbolism $\|\alpha\|^M$ stands for the set of worlds in M in which α is true. The symbolism $\models_w^M \alpha$ is used to express that α is *true in the model M at world w* (or simply *true*, if some M and w are understood). Validity, denoted $\models \alpha$, and satisfiability have their usual definitions. For convenience, we define a *world selection function f*, in terms of which the truth conditions for $\Rightarrow$ are specified:

Definition: $f(w, \|\alpha\|^M) = \{w_1 \mid Eww_1$ and $\models_{w_1}^M \alpha$, and for all w_2 such that Ew_1w_2 and $\models_{w_2}^M \alpha$, we also have $Ew_2w_1\}$.[1]

This function then, given a world w and proposition $\|\alpha\|^M$, picks out the least exceptional worlds in which α is true. Given a model $M = <W, E, DI, V>$, truth at a world w is given by:

Definition:

(i) For n-place predicate symbol P, terms $t_1, \ldots, t_n$, and $w \in W$, $\models_w^M P(t_1, \ldots, t_n)$ iff $<V(t_1), \ldots, V(t_n), w> \in V(P)$.

(ii) $\models_w^M \neg\alpha$ iff not $\models_w^M \alpha$.

(iii) $\models_w^M \alpha \supset \beta$ iff if $\models_w^M \alpha$ then $\models_w^M \beta$.

(iv) $\models_w^M \alpha \Rightarrow \beta$ iff $f(w, \|\alpha\|^M) \subseteq \|\beta\|^M$.

(v) $\models_w^M (x)\alpha$ iff for every V' which is the same as V except possibly $V(x) \neq V'(x)$, and where $M' = <W, E, DI, V'>$, $\models_w^{M'} \alpha$.

The conditional logic N is the smallest set of sentences of L that contains classical first-order logic and that is closed under the following axiom schemata and rule of inference.

Axiom Schemata[2]

ID	$\alpha \Rightarrow \alpha$
CC	$((\alpha \Rightarrow \beta) \wedge (\alpha \Rightarrow \gamma)) \supset (\alpha \Rightarrow (\beta \wedge \gamma))$
RT	$(\alpha \Rightarrow \beta) \supset (((\alpha \wedge \beta) \Rightarrow \gamma) \supset (\alpha \Rightarrow \gamma))$
CV	$\neg(\alpha \Rightarrow \beta) \supset ((\alpha \Rightarrow \gamma) \supset ((\alpha \wedge \neg\beta) \Rightarrow \gamma))$
CC′	$((\alpha \Rightarrow \gamma) \wedge (\beta \Rightarrow \gamma)) \supset ((\alpha \vee \beta) \Rightarrow \gamma)$
∀N	$(x)(\alpha \Rightarrow \beta) \supset (\alpha \Rightarrow (x)\beta)$ if α contains no free occurrences of x.

Rule of Inference

RCM From $\beta \supset \gamma$ infer $(\alpha \Rightarrow \beta) \supset (\alpha \Rightarrow \gamma)$.

The notions of theoremhood in N, and derivability and consistency are defined in the usual manner. The symbolism $\Gamma \vdash_N \alpha$ means that α is derivable from Γ in N. We obtain:

Theorem: $\models \alpha$ iff $\vdash_N \alpha$.

Soundness is proven by a straightforward inductive argument. Completeness is proven by showing that there is a *canonical N*-model, in which every non-theorem of N is invalid. This proof is an adaptation of the method of canonical models in first-order modal logics [Hughes and Cresswell 84], but modified to accommodate the variable conditional operator.

4. Default Reasoning: Initial Considerations

A *default theory T* is an ordered pair $<D, C>$ where D is a set of wffs of N and C is a non-empty consistent set of wffs of FOL. D is intended to represent necessary or conditional sentences constraining how the world *must* be or *could* be, while C is a set of contingent sentences constraining how the world being modelled *is*. Thus in D we would include statements such as "all ravens must be birds" and "all ravens are normally black". Included in C would be statements such as "*opus* is a raven" and "everyone taking CMPT882 this semester is under 6 feet tall". The goal is to define a "default" provability operator which, following [McDermott and Doyle 80], I will write as $T \mid\!\sim p$ to indicate that p follows by default from T.

The first part of this enterprise is startlingly easy. Consider for example where all that is known is $Bird(opus) \Rightarrow Fly(opus)$ and $Bird(opus)$. As argued, we should not be able to conclude from this that $Fly(opus)$, simply because, while the truth of $Bird(opus)$ relies of *this* state of affairs, the truth of $Bird(opus) \Rightarrow Fly(opus)$ relies on *other* less exceptional states of affairs, and there is no necessary connection between this state of affairs and the other states of affairs. Yet nonetheless it does seem reasonable to conclude $Fly(opus)$ "by default". The key point here is that in drawing this default conclusion, one is relying on a tacit assumption: that the world at hand is as unexceptional as possible, consistent with what is known. That is, given the above, it is entirely consistent that *opus* is a penguin, is tethered, or simply (for no known reason) does not fly. The default conclusion relies on assuming that if none of these exceptional factors are known to hold, they are assumed not to hold. This assumption can be stated as follows:

The Assumption of Normality: The world being modelled is among the least exceptional worlds according to D in which the sentences of C are true.

Thus it seems that we would want to say that $T \sim p$ just when, in the presence of "background information" D, p is true in all least exceptional worlds in which C is true[3]. Hence:

Temporary Definition: $T \sim p$ iff (so it seems) $D \vdash_N C \Rightarrow p$.

This does in fact work in a large class of simple cases. For example if we have that

$D = \{Raven(x) \Rightarrow Black(x),$
$(Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x)\}$,[4]
$C = \{Raven(a)\}$,

then we can make the default conclusion that $Black(a)$. If we have that $C = \{Raven(a), Albino(a)\}$, then we can derive by default that $\neg Black(a)$; we cannot now derive $Black(a)$, because we cannot prove

$$\{Raven(x) \Rightarrow Black(x), (Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x)\} \vdash_N (Raven(a) \wedge Albino(a)) \Rightarrow Black(a).$$

Similarly, if we have

$D = \{Quaker(x) \Rightarrow Pacifist(x), Republican(x) \Rightarrow \neg Pacifist(x)\}$

and learn that $Quaker(a)$ then we can conclude $Pacifist(a)$. If we learn also that $Republican(a)$, then we can conclude nothing concerning whether a, by default, satisfies *Pacifist*.

However the approach to this point also fails to work for a large class of simple cases. If we have that:

$D = \{Raven(x) \Rightarrow Black(x)\}$,
$C = \{Raven(a), Has\_wings(a)\}$

then the relation

$$\{Raven(x) \Rightarrow Black(x)\} \vdash_N (Raven(a) \wedge Has\_wings(a)) \Rightarrow Black(a)$$

cannot be shown. The difficulty, from the semantic end of things, is that in all models of these sentences, in the simplest

[1] Note that the apparent circularity in this and the next definition is benign.

[2] I am following the conventions of [Chellas 75] and [Nute 80] for naming axioms and rules of inference.

[3] By this I mean of course the conjunction of elements in C is true; in following formulae I will use C to stand also for the conjunction of the elements of C.

[4] Henceforth I will, in the interests of readability, generally drop the universal quantifier. Universally quantified versions of the formulae follow trivially by universal generalisation.

worlds where $Raven(a)$ is true, $Black(a)$ is also true. However there are models where, in the simplest worlds where $Raven(a)$ and $Has\_wings(a)$ are true, $Black(a)$ may not be true. Hence $(Raven(a) \wedge Has\_wings(a)) \Rightarrow Black(a)$ is not entailed by D.

It seems however that, based on what is known, there is no good reason for supposing that having wings has any effect on blackness. In a word, having wings seems *irrelevant* to whether a raven is black. This is the second assumption that I make in order to be able to draw default inferences in a default theory. It may be stated as:

Assumption of Relevance: Only those sentences known to bear on the truth value of a conditional relation will be assumed to, in fact, have a bearing on that relation's truth value.

This is of course rather vague, and part of the task in the next section is to make this notion more precise.

5. Two Approaches for Default Reasoning

The general idea in this paper is to use the logic N for representing defaults, and use metatheoretic considerations to sanction contingent default inferences. To this end, two assumptions were identified in the previous section as being essential for default inferences. In the previous section also, the formal system N was used to suggest an initial approach for default inferencing. As mentioned though, this approach fails for a wide class of simple cases. Two approaches are presented in this section to rectify these difficulties. The general idea in both approaches is to consider only a subset of the models of a default theory T for default inferences. Interestingly, the approaches derive from somewhat complementary intuitions, yet there is a high degree of symmetry between them.

The First Approach: Consider the statement $\alpha \Rightarrow \gamma$. This statement is true at a world w in model M iff $f(w, \|\alpha\|^M) \subseteq \|\gamma\|^M$. Intuitively, β is irrelevant to the truth of this statement if knowing β doesn't alter our judgement of the truth of the consequent of the conditional. Hence, according to our truth conditions for the conditional, β is irrelevant to $\alpha \Rightarrow \gamma$ iff $f(w, \|\alpha \wedge \beta\|^M) \subseteq \|\gamma\|^M$ and $f(w, \|\alpha \wedge \neg\beta\|^M) \subseteq \|\gamma\|^M$. So one approach is to assume, whenever possible, that a proposition β has no effect on the truth value of $\alpha \Rightarrow \gamma$. Hence, informally, we begin with a set of assertions D and extend this set by iteratively considering each conditional $\alpha \Rightarrow \gamma$ in D and each wff β of FOL, and if $\alpha \wedge \beta \Rightarrow \gamma$ is consistent, adding it to D. Thus if D is

$$\{Raven(x) \Rightarrow Black(x),\ (Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x)\}$$

we will add statements including

$$(Raven(x) \wedge Has\_wings(x)) \Rightarrow Black(x) \text{ and}$$
$$(Raven(x) \wedge Albino(x) \wedge \neg Has\_wings(x)) \Rightarrow \neg Black(x).$$

However, this isn't quite right. If D is $\{Q(x) \Rightarrow P(x), R(x) \Rightarrow \neg P(x)\}$ we could consistently add either $(Q(x) \wedge R(x)) \Rightarrow P(x)$ or $(Q(x) \wedge R(x)) \Rightarrow \neg P(x)$ (but not both) by this recipe. The solution is to add $\alpha \wedge \beta \Rightarrow \gamma$ only if there is no other "relevant" conditional that denies γ. This can be accomplished as follows:

Definition: $\alpha \Rightarrow \gamma$ is *supported* in Γ if there is β such that:

1. $\vdash_{FOL} \alpha \supset \beta$.
2. $\Gamma \vdash_N \beta \Rightarrow \gamma$.
3. If there is β' such that $\vdash_{FOL} \alpha \supset \beta'$ and $\Gamma \vdash_N \neg(\beta' \Rightarrow \gamma)$ then $\vdash_{FOL} \beta \supset \beta'$.

Using this we can define the procedure for forming an extension. If $\beta_0, \beta_1, \cdots$ is some ordering of wffs of FOL, we obtain:

Definition: An *extension* $E(D)$ of D is defined by:

1. $E_0 = D$.
2. $E_{i+1} = \delta$, where δ is defined by:
   Initially $\delta = \emptyset$.
   For each $D \vdash_N \alpha \Rightarrow \gamma$, $\delta = \delta \cup \{\alpha \wedge \beta_i \Rightarrow \gamma\}$ if $\alpha \wedge \beta_i \Rightarrow \gamma$ is supported in D;
   $\delta = \delta \cup \{\alpha \wedge \neg\beta_i \Rightarrow \gamma\}$ otherwise.
3. $E = \bigcup_{i=0}^{i=\infty} E_i$.

The procedure may be thought of as adding an inordinate number of default frame axioms to a set of defaults, in order to say that apparently irrelevant sentences are in fact irrelevant. Clearly only a single extension is produced. We obtain that $D \subseteq E(D)$ and $E(E(D)) = E(D)$ for an extension. Hence, under the process of forming an extension, an extension is a *fixed point* of the set of defaults. However, if $D_1 \subseteq D_2$, it may not be the case that $E(D_1) \subseteq E(D_2)$. An example is $D_1 = \{\alpha \Rightarrow \gamma\}$ and $D_2 = D_1 \cup \{\alpha \wedge \beta \Rightarrow \neg\gamma\}$, wherein any $E(D_1)$ contains $\alpha \wedge \beta \Rightarrow \gamma$ but no $E(D_2)$ does. We also obtain:

Theorem: $E(D)$ is consistent if D is.

Theorem: For any $\beta \in$ FOL and $\alpha \Rightarrow \gamma \in D$, $\alpha \wedge \beta \Rightarrow \gamma \in E(D)$ or $\alpha \wedge \neg\beta \Rightarrow \gamma \in E(D)$.

We can define default provability as we did in the last section, but now incorporating assumptions of relevance via the extension. That is:

Definition: $T \vdash \rho$ iff $E(D) \vdash_N C \Rightarrow \rho$.

Thus ρ follows by default from T if, considering all assumptions of irrelevance, ρ follows conditionally from the known facts C.

This approach yields reasonable default inferences, with one exception. Consider where we have $D = \{\alpha \Rightarrow \beta,\ \beta \Rightarrow \gamma,\ \alpha \Rightarrow \neg\gamma\}$ and $C = \{\alpha\}$. If would seem that in this case the best strategy is to conclude neither γ nor $\neg\gamma$. However, since $D \vdash_N C \Rightarrow \neg\gamma$, then in the extension of D we will also conclude $\neg\gamma$. There seems to be no obvious remedy for this difficulty in this approach; fortunately it does not occur in the next approach.

The Second Approach: This approach is perhaps the complement of the first. Whereas before we added assumptions to D to constrain the models that we wanted to consider for a default inference, here we assume that the world at hand is among the simplest worlds, consistent with what is known contingently. Thus for example if we know only that

$$D = \{Raven(x) \Rightarrow Black(x)\}.$$
$$C = \{Raven(opus),\ Has\_wings(opus)\}$$

then if the state of affairs modelled by C were among the simplest worlds according to D then, by the definition of $\Rightarrow$, $Black(opus)$ must be true in that state of affairs. So the idea is to first make whatever conclusions we can about C under the assumption of normality. Given such an extension (or extensions) to C we can specify that ρ follows as default inference from T iff ρ follows in FOL from all extensions of C.

There is a minor difficulty with this approach however arising again from the relative strength of defaults. Consider where we have:

$$D = \{Raven(x) \Rightarrow Black(x),\ (Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x)\}.$$

Thus in the least exceptional states of affairs in which there are ravens, ravens are black, and in the least exceptional states of affairs in which there are albino ravens, ravens are not black. From this it follows that the states of affairs in which there are ravens are less exceptional than the states of affairs in which there are albino ravens. This means that if we have that $C = \{Raven(opus), Albino(opus)\}$, then in extending C we should only consider the second default.

It is interesting also to see how this approach handles possible transitive relations in the defaults. Consider where we have that

$D=\{Quaker(x) \Rightarrow Pacifist(x), Pacifist(x) \Rightarrow Vegetarian(x)\}$ and $C=\{Quaker(a)\}$.

If we assume that the world at hand is among the least exceptional consistent with C, then we can conclude $Pacifist(a)$. However, given this new information, it now also becomes reasonable to conclude that $Vegetarian(a)$, barring evidence to the contrary. So effectively we need to "iterate" over default transitivities, while allowing for the fact that particular transitivities may not be warranted. Hence in the above example, if we were to add $Quaker(x) \Rightarrow \neg Vegetarian(x)$ to D, we would still want to conclude $Pacifist(a)$ but not be able to conclude $Vegetarian(a)$. This is accomplished as follows:

Definition: A *maximal contingent extension* $E(C)$ of C is defined by:

1. $C_0 = C$.
2. If $D \vdash_N \alpha \Rightarrow \gamma$ and $\vdash_{FOL} C_i \supset \alpha$ and if there is α' so that $\vdash_{FOL} C_i \supset \alpha'$ and $D \vdash_N \neg(\alpha' \Rightarrow \gamma)$ then $\vdash_{FOL} \alpha \supset \alpha'$ then $C_{i+1} = C_i \cup \{\alpha \supset \gamma\}$.
3. $E(C) = \bigcup_{i=0}^{i=\infty} C_i$.

This means that $\alpha \supset \gamma$ is added to C_i, if C_i implies α and for any α' implied by C_i, which conflicts with the default conclusion of γ, α' is implied by α. If we use $\vdash'$ for default derivability in this approach, we obtain:

Definition: If $T = <D, C>$ then $T \vdash' \rho$ iff $\cap E(C) \vdash_{FOL} \rho$.

Note that the number of extensions will typically be finite. Two extensions are distinct if and only if there are transitivities in the defaults that conflict. That is, we get more than one extension only if we have defaults of the form $\alpha \Rightarrow \gamma$ along with $\alpha \Rightarrow \beta$, $\beta \Rightarrow \neg\gamma$. We obtain also that, with respect to default derivability, the inferences of the first approach subsume those of the second:

Theorem: If $T \vdash' \rho$ then $T \vdash \rho$.

The two approaches exhibit a high degree of symmetry. The first approach involves extending D. The basic issue in this approach concerns satisfying the assumption of relevance; the assumption of normality is trivially satisfied. The second approach on the other hand involves extending C. The basic issue in this approach concerns satisfying the assumption of normality; the assumption of relevance is trivially satisfied. Of the approaches, the first is similar, from a technical standpoint, to other procedures for forming maximal sets of formulae. However, it does not appear to lend itself to any straightforward implementation. In addition it sometimes leads to over-strong inferences. The second appears to have somewhat more promise for providing a basis for an implementation. In addition, the quantifier-free fragment of the logic N is decidable and so the second approach applied to specific individuals is easily seen to yield a decidable system. The next section describes a set of example default inferences under the second approach.

6. Some Examples

The second approach to default reasoning arguably leads to reasonable and intuitive default inferences. As a first example, assume that we have the default portion of a theory:

$D_1 = \{Adult(x) \Rightarrow Employed(x), Univ\_st(x) \Rightarrow \neg Employed(x)\}$,

say, adults are typically employed, while university students normally are not. If we knew that someone was an adult then we could conclude by default that that individual was employed. If we knew that someone was an adult and a university student, then we could draw no conclusion. If, on the other hand, we knew that someone was an adult and was Dutch, then we would still conclude that they were employed. Of course, we also know that university students are typically adults, and so the defaults could be augmented to:

$D_2 = D_1 \cup \{Univ\_st(x) \Rightarrow Adult(x)\}$.

Now if we were told that someone was an adult and a university student, we would conclude by default that that person was not employed. The reason that we can now draw a conclusion is that in any model of D_2, in the simplest worlds in which someone is a university student, that person is not employed (but is an adult). From the logic N, we have the relation:

$D_2 \vdash_N Adult(x) \Rightarrow \neg Univ\_st(x)$,

and so from N we can *derive* the default that, given D_2, adults are normally not university students.

Consider next the defaults:

$D_3 = \{Raven(x) \Rightarrow Black(x), Raven(x) \Rightarrow Fly(x), (Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x)\}$.

Not unexpectedly, we can conclude by default that ravens with wings are black, and that ravens that fly (or don't fly) are black. Moreover albino ravens are concluded by default to fly but to not be black. Consider further where we augment the defaults so that we have:

$D_4 = D_3 \cup \{Bear(x) \Rightarrow Black(x), (Bear(x) \wedge Has\_illness\_X(x)) \Rightarrow \neg Black(x)\}$.

The default conclusions in D_3 go through as before. However, now if we learn that a particular raven has illness X then we would *not* conclude by default that the raven was not black; rather we would still conclude that the individual was black. The reason for this is that, by our notion of relevance, illness X has no apparent connection with the colouring of ravens, even though it clearly does for bears.

Transitive relations among the defaults appear to be handled correctly. If we have:

$D_5 = \{Quaker(x) \Rightarrow Pacifist(x), Pacifist(x) \Rightarrow Vegetarian(x)\}$,

and we know contingently that $Quaker(a)$ and $Republican(a)$, then we could conclude by default than $Vegetarian(a)$. If we were to augment D_5 with either $Republican(x) \Rightarrow \neg Pacifist(x)$ or $Republican(x) \Rightarrow \neg Vegetarian(x)$ then in neither case could we form the default conclusion $Vegetarian(a)$. Nor could we if $\neg(Republican(x) \Rightarrow Pacifist(x))$ were added. If, on the other hand, we have:

$D_6 = \{Quaker(x) \Rightarrow Pacifist(x), Pacifist(x) \Rightarrow Vegetarian(x), (Pacifist(x) \wedge Republican(x)) \Rightarrow \neg Vegetarian(x)\}$,

and we know that $Quaker(a)$, then we could conclude $Vegetarian(a)$. If we knew that $Quaker(a)$ and $Republican(a)$, then we could conclude that $Pacifist(a)$. However, since we have that pacifists are normally vegetarian, but that republican pacifists are normally not vegetarian, we would conclude $\neg Vegetarian(a)$.

7. Discussion

The logic N together with the approaches described in this paper provide a basis for representing, and reasoning about, default statements, and for performing default inferencing. Arguably the properties of the logic conform to commonsense intuitions concerning default statements. Arguably also, the logic is more appropriate for representing information about defaults than default logics or non-monotonic logics, in that its semantics does not rest on the notion of consistency with a given set of assertions. Thus the relation between ravens and blackness is phrased independently of any particular believer or believers. In addition, one can reason about default-like statements expressed by variable conditionals. So for example, the statement "all ravens must necessarily be birds" can be determined to be inconsistent with "ravens are normally not birds". The logic allows also for the representation of denials of defaults and so also for default reasoning with such denials.

Two approaches to default inferencing in this system were presented. Both approaches were based on the intuition that proposition ρ followed by default from a default theory T if, in the simplest or most uniform worlds consistent with what was known, ρ was true in all such worlds. Both approaches relied on the assumptions:

Normality: The world at hand is one of the least exceptional worlds consistent with the default theory T.

Relevance: Only those sentences known to bear on the truth value of a conditional relation are assumed to have a bearing on that relation's truth value.

Both assumptions seem reasonable and are, arguably, required for any other approach to default reasoning; it is perhaps an advantage of this approach then that such assumptions must be explicitly made before default inferences can be justified.

For the default inference of a proposition ρ, the first approach involves extending the set of default sentences D, and then saying that ρ follows from the original default theory if ρ follows contingently from C in the extension of D. The second approach was to assume that the world being modelled by C was among the simplest consistent with D, and then using this assumption to enable further information concerning that world to be obtained. Arguably both approaches are grounded in reasonable intuitions — the first with expanding the set of default assumptions and the second with inferring further contingent information about the world being modelled. The first approach however was shown to lead to overly strong default inferences in the case of transitive defaults.

Unlike the approach of [Reiter 80], which addresses ways in which a set of beliefs can be extended consistently, the approaches in the paper deal with situations in which default inferences may reasonably be carried out. Hence the approach produces only a single set of plausible default "theorems" rather than some set of default theory extensions. Unlike the approaches of [Reiter 80] and [McCarthy 80], there is no need to rely on, or explicitly represent, combinations of exceptional conditions in order to obtain reasonable results. Thus we can say that ravens are normally black and albino ravens are normally not black, without the need of explicit "abnormality" predicates, or the addition of other explicit default statements to constrain possible "interactions".

The approach of McDermott and Doyle has been criticised in [Davis 80] because there is no way of associating the truth of β with $M\beta$ (i.e. "consistent β"). The same point can be made about the system at hand, although arguably here it is not a disadvantage. There is no connection in N between the truth of α and β, and of $\alpha \Rightarrow \beta$. However this seems reasonable: α and β tell us something about the world being modelled; $\alpha \Rightarrow \beta$ tells us something about other, more uniform, states of affairs, and there is no reason to expect that there be a connection. It is only when we connect the truth values, by declaring that the world at hand is one of the least exceptional worlds consistent with what's known, that we can form default inferences.

An interesting extension to this work would be to consider its applicability to other conditional logics. There is nothing about the approaches described above that limit them specifically to N. In particular, there has been much work in philosophy in developing conditional logics for representing counterfactual and subjunctive relations [Lewis 73], [Nute 80], and for representing obligation [van Fraassen 72]. The techniques presented herein then should, with simple modification, be applicable to default inferences concerning counterfactuals, subjunctives, and notions of obligation. Thus, as an example, if we had the statements "if John comes, it will be a good party" and "if John and Sue come, it will be a dull party" represented in one of the logics of [Lewis 73], and if we also knew that only John would be going to the party, then using techniques similar to those of this paper it should be possible to formalise the reasoning that would let us conclude that (likely) it will be a good party.

Acknowledgements

I would like to thank Robert Hadley and Sharon Hamilton for their very helpful comments on an earlier draft of this paper. This research was supported by the Natural Science and Engineering Research Council of Canada grant A0884.

References

[1] R.J. Brachman, "'I Lied About the Trees' or Defaults and Definitions in Knowledge Representation", *The AI Magazine*, 6,3, 1985, pp 80-93.

[2] B.F. Chellas, "Basic Conditional Logic", *Journal of Philosophical Logic 4*, 1975, pp 133-153.

[3] M. Davis, "The Mathematics of Non-Monotonic Reasoning", *Artificial Intelligence* 13, 1980, pp 73-80

[4] J.P. Delgrande, "A Propositional Logic for Natural Kinds", *AI-86 Canadian Society for Computational Studies of Artificial Intelligence Conference*, May 1986

[5] J.P. Delgrande, "A First-Order Logic for Prototypical Properties", *Artificial Intelligence* (to appear), 1987

[6] J.P. Delgrande, "A Formal Approach to Default Reasoning Based on a Conditional Logic (Extended Report)", LCCR TR 87-9, School of Computing Science, Simon Fraser University, 1987.

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

[8] G.E. Hughes and M.J. Cresswell, *A Companion to Modal Logic*, Methuen and Col. Ltd., 1984.

[9] D. Lewis, *Counterfactuals*, Harvard University Press, 1973.

[10] J. McCarthy, "Circumscription — A Form of Non-Monotonic Reasoning", *Artificial Intelligence 13*, pp 27-39, 1980.

[11] D. McDermott and J. Doyle, "Non-Monotonic Logic I", *Artificial Intelligence 13*, 1980, pp 41-72.

[12] D. Nute, *Topics in Conditional Logic*, Philosophical Studies Series in Philosophy, Volume 20, D. Reidel Pub. Co., 1980.

[13] D. Nute, "A Non-Monotonic Logic Based on Conditional Logic", *ACMC Research Report 01-0007*, Advanced Computational Methods Center, The University of Georgia, 1986

[14] R. Reiter, "A Logic for Default Reasoning", *Artificial Intelligence 13*, 1980, pp 81-132.

[15] E. Rosch, "Principles of Categorisation" in *Cognition and Categorisation*, E. Rosch and B.B. Lloyds eds., Lawrence Erlbaum Associates, 1978.

[16] B.C. van Fraassen, "The Logic of Conditional Obligation", *Journal of Philosophical Logic 1*, 1972, pp 417-438.