A Logic for Representing Default and Prototypical Properties

James P. Delgrande

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

ABSTRACT

The area of default reasoning may be considered as consisting of two separate but closely related subareas. The first deals with representing default information, and addresses issues such as reasoning about default statements or determining the consistency of a set of defaults. The second addresses reasoning about the default properties of an individual, given a set of default statements. Most extant work in default reasoning has concentrated on the second area. This paper addresses the first area only, as an initial investigation into the area of default reasoning. The approach *is* based on adding a "variable conditional" operator to first-order logic. This operator is used to express prototypical, rather than strict, relations between entities and properties. A possible worlds semantics is provided for the logic along with a proof theory; soundness and completeness results are also provided. It is argued that the variable conditional in the logic captures common intuitions concerning defaults and prototypical properties.

1. Introduction

Many general statements concerning the real world are not by themselves true, but rather appear to rely on a collection of tacit background assumptions. Thus for example, "birds fly" seems to be a reasonable assertion, even though it isn't the case that *all* birds fly. Birds with broken wings don't fly; tethered birds don't fly; and even whole species of birds, such as penguins and ostriches, don't fly Rather "birds fly" seems to have the force of "normally birds fly" or perhaps "ignoring exceptional conditions, birds fly".

There are two main approaches in Artificial Intelligence (AI) for dealing with statments such as "birds fly": default reasoning and prototype theory. In the first case, "birds fly" is interpreted, roughly, as saying that if an object is known to be a bird. and it is consistent to believe that it flies, then conclude that it flies. In the second case, the statement is generally interpreted as being descriptive or predictive, and may be taken as meaning. again roughly, "most birds fly".

In this paper, a third alternative is introduced The general idea is that a *variable conditional operator* =► is introduced into standard first-order logic (FOL). where the statement A=>B is interpreted as "in the normal course of events, if *A* then *B".* So. for example. *(Vx)(Bird(x)=> Fly(x))* would have the intended interpretation "for every object x. in the normal course of events if that object is a bird, then it flies" Intuitively, the connection between the antecedent and consequent of these conditionals may be thought of as a relationship in a scientific theory. Birds fly. but only if some set of presumed "additional assumptions" is satisfied: the bird isn't tethered, isn't a penguin, etc. This approach is intended then to be applicable to terms standing for naturally occurring kinds or *natural kinds,* such as "raven", "lemon", etc. More generally, the approach is intended to also be applicable to general statements of typicality, and so we can use this connective to represent statements such as "typically Quakers are pacifists" or "typically adults are employed".

Since the intended interpretation of *A=>B* is "in the normal course of events, if *A* then *B".* the semantics of => will rest on the notion of other courses of events and. in particular, other *more normal* or *less exceptional* courses of events. In extending the semantics of FOL to account for => then I adopt a possible worlds approach, where the truth of *A=>B* at a world w relies not on the world w. but on other "less exceptional" worlds: *A=>B* is true if *B* is true in the "least exceptional" worlds where *A* is true. Roughly then this says that birds fly. if we "factor out" exceptional circumstances such as being featherless, being tethered, being a penguin, etc. A major issue then is the choice of a suitable metric for "less exceptional than" between possible worlds. Clearly the bounds placed on this metric will constrain the semantics of the => operator. The formal system we obtain is a *conditional logic,* of a class of logics that have been developed for representing counterfactual conditionals, conditional obligation, and other related notions.

The next section reviews related work in AI. while the third section introduces conditional logics. Section 4 informally introduces the underlying semantic theory of the logic for representing default information. Section 5 develops a formal semantics for the logic, while section 6 presents a proof theory. Section 7 discusses what we have gained from this approach, while the last section provides a conclusion. Proofs of theorems may be found in [Delgrande 86b); an earlier version of this work was presented in [Delgrande 86a].

2. Related Work

There has been extensive work in AI in extending or augmenting classical first-order logic to deal with default and prototypical properties. Most of these approaches can be termed "consistency-based", in that a default conclusion is typically

warranted in part by being consistent with some given set of beliefs. For example, in Reiter's augmentation of first-order logic [Reiter 80]. "birds fly" would be represented by the default rule:

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

This can be read, roughly, as "if. for some object, you can prove that that object is a bird, and it is consistent that it flies, then infer that it flies' [McCarthy 80]. [McDermott and Doyle 80]. and [Moore 83] describe other consistency-based approaches.

A general limitation with such approaches is that one cannot reason *about* defaults. Thus if we had default rules corresponding to "ravens are normally black" and "albino ravens are normally not black", there is no means within the system of drawing the conclusion "non-albino ravens are normally black . Similarly, in most systems the assertions "ravens are birds" and "typically ravens aren't birds" can co-exist peacefully — in Reiter's system the default rule is simply never applied, and in McDermott and Doyle's no problems arise because there is no semantic connection between a statement *A* and the statement *MA* (i.e. "consistent *A").* Yet it seems that such a pair of sentences should be inconsistent: if every raven must necessarily be a bird, then it seems unreasonable to assert that typically ravens aren't birds.

A second difficulty with these approaches arises from the fact that their semantics rests on a notion of consistency with a set of beliefs. Thus, paraphrasing the first example, a bird may be believed to fly. if this does not conflict with prior beliefs. However, the relationship between birds and flight, whatever it may be. is clearly a relation just between birds and the property of flight; it does not. in particular, rely on any set of beliefs or any particular believer. Thus, such consistency-based approaches seem useful for telling us how to appropriately *extend* a belief set. The goal here, on the other hand, is to attempt to *represent* the relation between, say. birds and flight

A second approach in AI for dealing with default and prototypical properties is prototype theory [Rosch 78]. In this case membership in the extension of a term is a graded affair and is a matter of similarity to a representative member or prototype. Thus one might say that birds fly but only with some given certainty. A difficulty with these approaches is that there is no clear agreement as to what is meant by "certainty", nor how one may combine and work effectively with such certainties [Thompson 85]. However the problems appear to run deeper than this. Prototype theory seems concerned generally with descriptions of individuals, or predicting properties of individuals. Thus generally there is no way of distinguishing the force of a statement such as "birds fly" from "students like pizza". This though is too weak for our purposes. We would want to be able to attribute the property of flight, in some absolute sense, to the class of birds, and leave the connection between students and pizzas as holding contingently in the state of affairs in which we're interested. In sum. notions of typicality and resemblance to a prototype appear too weak to be useful for our purposes.

Insofar as representing defaults is concerned, a related problem in the area of linguistic semantics is the treatment of *generic* statements. Examples of generic statements include "birds fly", "the dodo is extinct", and "John walks to work". In these examples, "birds", "dodo" and "walks" are used generically Perhaps the best-known work with respect to these statements is that of Gregory Carlson [Carlson 80], [Carlson 82], who uses the framework of Montague semantics for their treatment. Carlson argues against any quantificational treatment of generics and instead proposes that kinds be treated as individuals. For example, the term "dogs" in "dogs bark" is treated as a proper name and the statement is true, roughly, if the kind "dogs" has the property of occasionally or normally barking. This approach appears to address a wide class of recalcitrant problems associated with the generic in linguistics. However it appears to not directly address our concerns. There is no indication as to what it means to be a "normal" property of a kind, nor how such properties interrelate, nor how one could reason about such properties. In addition the approach requires an increased ontology of kinds and stages (roughly space/time instantiations of kinds or individuals).

3. Conditional Logics

Consider the following passage, taken from [Lewis 73]:

> "If Otto had come, it would have been a lively party; but if both Otto and Anna had come, it would have been a dreary party: but if Waldo had come as well, it would have been lively; but ".

These statements represent counterfactual conditionals: the antecedent of each statement is false, but each statement could be either true or false. These statements also exhibit a curious pattern: as the antecedent is strengthened, the consequent changes to its negation. It is apparent then that the standard material conditional is inadequate for representing counterfactual statements: since the antecedents of the above statements are false, the statements formed using a material conditional would have to come out true. However, clearly we would want to retain the option of having a counterfactual statement come out false — after all. it is conceivable that if Otto had come to the party, it would have been a boring affair.

David Lewis, in [Lewis 73] and building on [Stalnaker 68], proposes using a *variably strict conditional,* or simply *variable conditional,* to represent counterfactual statements. The formal systems that employ such a connective are called *conditional logics.* If we use $\Rightarrow$ to represent this conditional, then the general idea is that the truth value of $A \Rightarrow B$, relative to a world, depends on a subset of those worlds in which *A* is true. Thus for example in Lewis's approach. $A \Rightarrow B$ is true if the closest set of worlds (or *sphere)* most like our own that have *A* true also. for those worlds, have *B* true, and for no world in the sphere is $A \supset B$ false.

The counterfactual conditional also differs from the material conditional, in that it does not necessarily support tran-

sitivity. Consider the following example, taken from [Stalnaker 68]:

> "If Carter had been born in Russia, he would be a communist. If Carter were a communist, he would be sending American defense secrets to the Kremlin. Therefore, if Carter had been born in Russia, he would be sending American defense secrets to the Kremlin."

The conclusion is clearly dubious, and so a logic for a counterfactual conditional must also allow for failure of the transitivity of the conditional.

Related approaches for reasoning with counterfactuals and subjunctives, and reasoning about conditional obligation are described in [Nute 75] and [van Fraassen 72], while [Chellas 75] and [Nute 80] provide general discussions. The underlying semantic theory for such approaches is typically expressed using a possible worlds formulation, generally along the lines of Lewis'.

Consider now the case of default and prototypical properties. It is fairly easy to come up with patterns of reasoning that are quite similar to those cited for counterfactual reasoning. So. for example, it seems reasonable to say that ravens are black, but albino ravens are not black. In a similar fashion, perhaps there is some disease X that turns albino ravens black again, and so we would also want to be able to assert that albino ravens with disease X are black. Hence we should allow that a strengthening of an antecedent may reverse the truth value of a consequent. Failure of transitivity is also easy to show. Thus, every penguin is necessarily a bird: birds normally fly: yet penguins normally don't fly. In addition perhaps. Quakers are normally pacifists; pacifists are normally vegetarians; yet Quakers are not normally vegetarians.

There are major differences though between representing counterfactual knowledge and representing default knowledge. Most significantly, counterfactual reasoning treats conditionals where the antecedent is false but the conditional as a whole may be either true or false. For default knowledge, we are interested in the case where the antecedent of the conditional is true. Thus, for example, when we assert that ravens are black and albino ravens are not black, we expect that there are in fact ravens and albinos. So it appears that what we require is a conditional logic, but one where the properties of the logic are "tailored" to the problem at hand. Such a logic then would have the same relation to conditional logics in general as particular epistemic logics of belief have to modal logics in general: each would constitute a particular system formulated in response to a particular need.

4. Initial Considerations

A statement such as "ravens are black" is to be interpreted as "normally ravens are black". The approach taken is to say. roughly, that a raven is black, or would be black, if we "factor out" exceptional circumstances such as being an albino, being painted, being in a strong red light, or whatever. Similarly we would want to say that an albino raven is not black if we again "factor out" exceptional circumstances such as having some disease that turns individuals black, being painted, etc. Thus, roughly, if everything else was equal, a raven would be black, and an albino raven would be non-black. A key point is that in general there will be an arbitrarily large set of exceptional circumstances and so we will not be able to specify all possible exceptional conditions.

Since the truth of "ravens are black" involves ignoring exceptional circumstances, this means that we are considering, in some sense, less exceptional states of affairs in order to determine the truth of this statement. So we can say that "ravens are black" is true if in the least exceptional states of affairs in which there are ravens, ravens are black. Similarly, in the least exceptional states of affairs in which there are albino ravens, albino ravens are black. (From this it follows that the least exceptional states of affairs in which there are albino ravens are more exceptional that the least states in which there are ravens.) This then is the way statements such as "ravens are black", "albino ravens are non-black", and so on are informally interpreted: we consider not the state of affairs being modelled, but other "less exceptional" states of affairs.

The semantics for the formal system to be presented then is based on a possible worlds formulation (where a "possible world" corresponds to a "possible state of affairs"). The accessibility relation *E* between worlds is interpreted 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. From this. A=> *B* is true at a world just when the least exceptional worlds in which *A* is true also have *B* true. The notion of "at least as exceptional as" between possible worlds is. to be sure, a rather imprecise one. Yet. arguably, one uses just such a metric when asserting that "birds fly" or "ravens are black", or any other commonsense statement. Moreover, there are some conditions that can be placed on such a metric, and. arguably, these conditions yield a system that conforms to common intuitions concerning default assertions.

In [Delgrande 86b], 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.

Clearly the notion of "at least as unexceptional as" between worlds should be reflexive and transitive; the third condition states that any worlds accessible from some world are themselves comparable. A given world w then "sees" a succession of disjoint sets of worlds arranged in a strict ordering, wherein all the worlds in a set are equivalent with respect to "unexceptionalness The 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.

Most conditional logics, including all that have been cited here, however do not base their semantics on an accessibility relation *E.* but rather employ a "world selection" function /. This function, given a proposition and a world, has as value the sei of least exceptional worlds in which the proposition is true. This means that $A \Rightarrow B$ is true at world w just when the set of worlds picked out by /. applied to w and the proposition expressed by *A.* is contained in the set of worlds in which *B* is true. Clearly then, given an accessibility relation *E.* it is possible to obtain a world selection function /: in [Delgrande 86b], the converse is also shown to hold for the system at hand.

In the interests of bringing the present work into line with this previous work in conditional logic, the formal semantics is in fact formulated in terms of such a function. In addition, this reformulation has the advantage of being, in the end. simpler. For the remainder of this paper then I consider a semantic treatment in terms of a function /; conditions for this function are given in the next section.

As a final point in our informal interpretation of "normally*, note that there is no statistical connotation in "ravens are normally black". In fact, it is quite possible that *no* raven is black in some state of affairs, yet "ravens are normally black" is still true. One can imagine, for example, some new disease that turns ravens white: in this case it is quite reasonable to say something like "ravens are normally black (but will probably remain white until an antidote can be found)". This seems to be a slightly more general reading that that usually adopted in other systems of default reasoning, where "ravens are black" is generally interpreted as "ravens are typically black". This last reading seems a little unusual in the above example, which involves no raven (contingently) being black.

5. A Formal Semantics

The language *L* for the variable conditional consists of the language of FOL. augmented with a connective for the variable conditional. The language has the following primitive symbols: a denumerably infinite set of *individual variables x.y.z.* • • ; a denumerably infinite set of *individual constants a.b.c.* • • • ; for each $n \geq 1$ a denumerably infinite set of n-place *predicate symbols. P.Q.R.* ; together with the symbols $\neg$, $\supset$, $\Rightarrow$, and V. and parentheses and commas for punctuation. The symbols *A.B.C.* will stand for arbitrary well-formed formulae of *L.* The set of individual variables and constants will make up the set of *terms.* 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.

Definition: The *well-formed formulae (wffs)* of *L* constitute the least set such that:

(i) If *P* is a n-place predicate symbol and $t_1, \ldots, t_n$ are terms, then $P(t_1, \ldots, t_n)$ is a wff.

(ii) If *A. B* are wffs and *x* is an individual variable, then $\neg A$, $(A \supset B)$, $(A \Rightarrow B)$, and $(\forall x)A$ are wffs.

As usual, conjunction (A), disjunction (V), biconditionality ($\equiv$), and the existential quantifier ($\exists$) are introduced by definition. Parenthesis may be omitted where no confusion arises. Also the standard notions of scope and of free and bound variables are assumed. *A(x)* will be used to indicate that *A* may have *x* as a free variable: *A(t)* is the result of uniformly substituting term *t* for all free occurrences of x

Sentences of *L* are interpreted in terms of a *model* $M = \langle W.f.D.V \rangle$ where *W* is a set. / is a function from *Wx (W)* to (HO. *D* is a domain of individuals, and V is a function on terms and predicate symbols so that

1. for term t. $V(t) \in D$.
2. for any n-place predicate symbol *P. V(P)* is a set of (n+1)-tuples $\langle t_1, \ldots, t_n, w \rangle$ where each $t_i \in D$ and $w \in W$.

For wff *A.* the symbolism $\|A\|^M$ will be used to stand for the set of worlds in *M* in which *A* is true. If we identify propositions with possible worlds, then $\|A\|^M$ stands for the proposition *expressed* by *A.* Informally *W* is a set of possible worlds and / picks out a set of possible worlds $f(w, \|A\|^M)$ for each possible world *w* and proposition $\|A\|^M$. V maps atomic sentences onto those worlds where the sentence is true, and predicate symbols onto relations in worlds. The symbolism $\models^M_w A$ is used to express that *A* is *true in the model hi at world w* (or simply *true,* if some *M* and w are understood). We write $\models A$ in the case that *A* is true at every world in every model, and say that *A* is *valid. A* is *satisfiable* if and only if $\neg A$ is not valid. Given a model $M = \langle W.f. D. V \rangle$. truth at a world *w* is given by:

Definition:

(i) For n-place predicate symbol P, terms $t_1, \ldots, t_n$, and $w \in W$, $\models^M_w P(t_1, \ldots, t_n)$ iff $\langle V(t_1), \ldots, V(t_n), w \rangle \in V(P)$.

(ii) $\models^M_w \neg A$ iff not $\models^M_w A$.

(iii) $\models^M_w A \supset B$ iff if $\models^M_w A$ then $\models^M_w B$.

(iv) $\models^M_w A \Rightarrow B$ iff $f(w, \|A\|^M) \subseteq \|B\|^M$.

(v) $\models^M_w (\forall x)A$ iff for every V' which is the same as V except possibly $V(x) \neq V'(x)$, and where $M' = \langle W, f, D, V' \rangle$, $\models^M_w A$.

Definition: By a *N-model,* we will mean a model $M = \langle W, f, D, V \rangle$ where the following conditions on f are met:

(i) $f(w, \|A\|^M) \subseteq \|A\|^M$.

(ii) If $f(w, \|A\|^M) \subseteq \|B\|^M$ then $f(w, \|A\|^M) \subseteq f(w, \|A \wedge B\|^M)$.

(iii) If $f(w, \|A\|^M) \not\subseteq \|B\|^M$ then $f(w, \|A \wedge \neg B\|^M) \subseteq f(w, \|A\|^M)$.

(iv) $f(w, \|A \vee B\|^M) \subseteq f(w, \|A\|^M) \cup f(w, \|B\|^M)$.

These conditions are shown in [Delgrande 86b] to yield a semantics equivalent to a reflexive, transitive, forward connected accessibility relation *E* with respect to some given world.

Finally, the conditional operator can be tied to the more familiar modal notions of necessity, or truth in all alternative

states of affairs, and possibility, in the following manner [Lewis 73]:

Definition: $\Box A$ is defined as $\neg A \Rightarrow A$.

$\Diamond A$ is defined as $\neg(A \Rightarrow \neg A)$.

6. A Proof Theory

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

Axiom Schemata

ID	$A \Rightarrow A$
CC	$((A \Rightarrow B) \wedge (A \Rightarrow C)) \supset (A \Rightarrow (B \wedge C))$
RT	$(A \Rightarrow B) \supset (((A \wedge B) \Rightarrow C) \supset (A \Rightarrow C))$
CV	$\neg(A \Rightarrow B) \supset ((A \Rightarrow C) \supset ((A \wedge \neg B) \Rightarrow C))$
CC′	$((A \Rightarrow C) \wedge (B \Rightarrow C)) \supset ((A \vee B) \Rightarrow C)$
$\forall N$	$(\forall x)(A \Rightarrow B) \supset (A \Rightarrow (\forall x)B)$ if A contains no free occurrences of x.

Rule of Inference

RCM From $B \supset C$ infer $(A \Rightarrow B) \supset (A \Rightarrow C)$.

For naming the axioms and rules of inference in the quantifier-free fragment of the logic, I have followed, and will follow, the conventions of [Chellas 75] and [Nute 80]. A wff A of L is a *theorem* of N, just when A is in N. We write $\vdash_N A$ or simply $\vdash A$. A is *derivable* in N from a set of wffs Γ iff there are $A_1, \ldots, A_n \in \Gamma$ such that $\vdash_N (A_1 \wedge \cdots \wedge A_n) \supset A$. A set of wffs is *consistent* in N iff not every wff is derivable from it.

The axiom schemata $\forall N$ has an obvious syntactic similarity to the theorem of FOL:

$(\forall x)(A \supset B) \supset (A \supset (\forall x)B)$ if A contains no free occurrences of x.

Essentially $\forall N$ states that for some A (with proviso), if for every individual x, $A \Rightarrow B(x)$, then $A \Rightarrow (\forall x)B$. This in turn effectively restricts the set of individuals in a world accessible (i.e. in $f(w, |A|^M)$) from some given world to be non-increasing. It is not surprising then that the Barcan formula [Hughes and Cresswell 68] is a theorem of N. That is, we have:

Theorem: $(\forall x)(\neg A \Rightarrow A) \supset (\neg(\forall x)A \Rightarrow (\forall x)A)$.

From our definition of $\Box$, this is equivalent to $((\forall x)\Box A) \supset (\Box(\forall x)A)$. The converse of the Barcan formula, which essentially restricts sets of individuals in accessible worlds to be non-decreasing, is easily shown to also be a theorem of N.

The logic N is, in the terminology of [Chellas 75], *normal* – that is, it is closed under the rules:

RCEA From $A \equiv B$ infer $(A \Rightarrow C) \equiv (B \Rightarrow C)$.

RCK From $(B_1 \wedge \cdots \wedge B_n) \supset B$ infer $((A \Rightarrow B_1) \wedge \cdots \wedge (A \Rightarrow B_n)) \supset (A \Rightarrow B)$, $n \geq 0$.

Hence it follows that equivalent propositions can be substituted into the antecedent or consequent of a variable conditional, and so we have full substitution of equivalents.

However, the key result here is that the proof theory is sound and complete with respect to the semantics given in the previous section. We obtain:

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

Soundness is proven by a straightforward inductive argument. Completeness is proven by showing that there is a N-model, called the *canonical 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.

Finally, by a standard result concerning a type of derived model called a *filtration* [Chellas 75] we obtain:

Theorem: The quantifier-free fragment of N is decidable.

7. Discussion

The previous sections gave a formal specification of the logic N. It remains to be shown that the logic, and in particular the variable conditional, captures common intuitions regarding prototypical and default properties. To begin with, the various difficulties encountered with the material conditional for representing default information do not arise with the variable conditional. So, for example, the following set of sentences is satisfiable:

$$\{Raven(x) \Rightarrow Black(x),\ (Raven(x) \wedge Albino(x)) \Rightarrow \neg Black(x)\}^1.$$

Moreover the sentences are satisfiable while having *Raven(x)* and *Albino(x)* true at a world. Hence, as required, we can strengthen the antecedent of a variable conditional, and reverse the truth of the consequent, without falling into inconsistency. This pattern of strengthening the antecedent to reverse the truth value of the consequent may be extended arbitrarily.

In addition,

$$\{Penguin(x) \supset Bird(x),\ Bird(x) \Rightarrow Fly(x),\ \neg(Penguin(x) \Rightarrow Fly(x))\}$$

is satisfiable, and satisfiable with true antecedents, in each of the conditionals. Also the following set of statements is similarly satisfiable:

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

So, as required, we lose transitivity of the variable conditional.

Thirdly, we can consistently assert universal conditional statements, together with statements about "exceptional" individuals. Thus for example the following set of sentences is satisfiable:

$$\{Bird(x) \Rightarrow Fly(x),\ Bird(opus),\ \neg Fly(opus)\}.$$

[1] In the interests of readability I have omitted universal quantifiers. Clearly universally quantified versions follow easily and trivially by universal generalisation.

However, we do obtain weakened forms of strengthening the antecedent and transitivity for the variable conditional. Axiom CV provides such a weakened form for strengthening the antecedent:

$$((Raven(x) \Rightarrow Black(x)) \wedge \neg(Raven(x) \Rightarrow Albino(x))) \supset ((Raven(x) \wedge \neg Albino(x)) \Rightarrow Black(x)).$$

Thus if ravens are normally black, but it isn't the case that ravens are normally albino, then we can conclude that ravens that aren't albino are normally black.

A consequence of axiom CC" provides another restricted form of strengthening the antecedent:

$$A \Rightarrow C \supset (((A \wedge B) \Rightarrow C) \vee ((A \wedge \neg B) \Rightarrow C)).$$

Hence if we have some conditional relation then, for any property, the consequent follows conditionally from the antecedent conjoined either with the property or its negation. So. if ravens are normally black, then it is either the case that albino ravens are normally black, or that non-albino ravens are normally black.

A rule of inference that follows from RCM provides a restricted form of transitivity:

$$\text{If} \vdash Raven(x) \Rightarrow Black(x) \text{ and} \vdash Black(x) \supset NonWhite(x) \text{ then} \vdash Raven(x) \Rightarrow NonWhite(x).$$

That is. if ravens are normally black and black things are not white then ravens are normally not white. We also have restricted transitivity from axiom RT:

$$Raven(x) \Rightarrow Black(x) \supset (((Raven(x) \wedge Black(x)) \Rightarrow BlackEyed(x)) \supset Raven(x) \Rightarrow BlackEyed(x)).$$

Thus if ravens are normally black and ravens that are black normally have dark eye pigment, then ravens normally have dark eye pigment. If the second occurrence of *Raven(x)* were dropped from the above, then we would of course have full transitivity.

In addition, we do not have a law of the excluded middle consequents of the variable conditional. Hence if a kind does not normally have a particular attribute, we are not bound to attribute to it the negation, unlike the material conditional. This is a consequence of the fact that the following sentence is satisfiable:

$$\neg((A \Rightarrow B) \vee (A \Rightarrow \neg B)).$$

Thus perhaps residing in North America is irrelevant with respect to ravenhood. and we would not want to say either that ravens normally live in North America or that they normally do not live there.

Also, we cannot have conflicting variable conditionals. A theorem of *N* is:

$$\Diamond A \supset (\neg(A \Rightarrow B) \vee \neg(A \Rightarrow \neg B)).$$

So if $A \Rightarrow B$ is true then $A \Rightarrow \neg B$ cannot be true (unless *A* is necessarily false).

The logic *N* differs from conditional logics for counterfactual reasoning primarily in that logics for representing counterfactual assertions generally allow some sort of connection between contingent truths and variable conditionals. For example, many systems contain the following theorem:

MP $\quad (A \Rightarrow B) \supset (A \supset B)$.

Thus, given a "counterfactual" where the antecedent is true, the consequent is also taken as true. An example perhaps is

"If Otto had come, a would have been a lively party."
"But Ouo *did* come."
"Hence, it must have been a lively party."

However, in *N* we obtain an unsurprising relation between implication and the variable conditional:

RCE $\quad \text{If} \vdash A \supset B \text{ then} \vdash A \Rightarrow B$

and so. if we were to add MP to *N.* we would obtain:

$$\vdash A \supset B \text{ iff} \vdash A \Rightarrow B$$

and the variable conditional would collapse into entailment.

Another formula that appears occasionally is the following:

CS $\quad (A \wedge B) \supset (A \Rightarrow B)$.[2]

CS requires that if the antecedent and consequent happen to be true at a world. then that world is a member of the least exceptional worlds with respect to the antecedent. This would have the undesirable consequence of tying our notion of "normally", which heretofore had rested on the notion of alternative stales of affairs, to contingent truths. Neither MP nor CS then is appropriate for our concerns, wherein contingent truths at a world should play no part in the truth of a variable conditional.

Conditional logics of obligation [van Fraassen 72] similarly differ in detail from *N.* most notably in rejecting the axiom ID.

ft. Conclusion

This paper has presented a logical system *N* for representing statements of default and prototypical properties. The language of the system consists of that of first-order logic, augmented with a variable conditional $\Rightarrow$. The intended interpretation of $A \Rightarrow C$ is "all other things being equal, if *A* then C" or "if *A* then normally C". The semantics of the system is based on a possible worlds formulation, wherein $A \Rightarrow C$ is true if in the least exceptional worlds in which *A* is true. C is also true. This allows the statements $A \Rightarrow C$ and $(A \wedge B) \Rightarrow \neg C$ to be jointly satisfiable. and with contingently true antecedents, simply by having the least exceptional worlds in which $A \wedge B$ is true differ from the least exceptional worlds in which *A* is true. In a similar fashion, transitivity of the variable conditional is blocked, and so the statements $A \Rightarrow B$, $B \Rightarrow C$, and $\neg(A \Rightarrow C)$ are jointly satisfiable with true antecedents. Similarly, it is possible to assert a variable conditional along with statements that would conflict with the corresponding material conditional. Hence $A \Rightarrow C$, *A.* and $\neg C$ are jointly satisfiable. We can then, via universal generalisation, assert general statements concerning "normal" properties without inconsistency. Hence, for example, we can assert that every bird normally flies, every penguin is

[3] See {Lewis 73] and (Nute 80] for taxonomies of such systems.

necessarily a bird, yet every penguin normally does not fly. In addition we can assert that *Tweety* is a bird that does not fly.

On the other hand, the variable conditional is strong enough to permit reasonable and intuitive relations between sentences. In particular we obtain restricted versions of strengthening the antecedent and of transitivity of the variable conditional. Thus, as an example of restricted transitivity, we have a derived rule that lets us state that if it is true that every bird normally flies, and that if everything that flies has the ability to become airborne, then it follows that every bird normally has the ability to become airborne. In addition we have restrictions such that A=>C and *A*=>-*C* are jointly satisfiable only if *A* is necessarily false.

The key point here then is that the logic allows one to *represent* statements of default and prototypical properties, and to *reason* about such statements. Thus it makes sense in the system to talk of one default statement being derivable in the logic from a set of others. In addition, the logic seems more appropriate for representing information about prototypical properties than extant default 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 birds and flight is phrased independently of any particular believer or believers. Finally, the quantifier-free fragment of the logic is decidable. and so there exist complete mechanical procedures for reasoning about the default properties of, for example, a single individual. One such procedure, based on the method of semantic tableaux, is described in [Groeneboer 87].

However, the work presented here contains one rather obvious omission, the logic contains no mechanism for reasoning "in the absence of other information". Thus for example if we know that *A* and that A=>C. we have no mechanism for concluding something like "in the normal course of events. C". This, from the semantic end of things is quite reasonable, in that the truth conditions for *A* and C are independent of the conditions for *A*=>*C*. Yet. on the other hand, it also seems reasonable that if we knew that birds normally fly and that *Tweety* is a bird, then we should be able to conclude "by default" that *Tweety* flies. The difficulty of course is that if we allowed *modus ponens* as a rule of inference, we would run into problems with strengthening the antecedent.

One solution is to use the model theory of the logic to suggest conclusions about individuals. Consider, for example, where we know only that ravens are normally black, and that albino ravens are normally not black, and that individual *opus* is a raven. This gives us some information concerning the world at hand, but not enough to conclude that *opus* is (or is not) 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 then, we would have also that *opus* is black. That is. since in the simplest worlds in which there are ravens, ravens are black, if the world at hand is among these worlds then ravens clearly are black at that world. If we were to subsequently add the information that *opus* was an albino, then we could no longer obtain such a result. The reason for this is that in the simplest worlds in which there are albino ravens, such ravens are not black, and so we would draw the alternative conclusion that *opus* was not black. This approach is developed and put on a formal footing in [Delgrande 87].

Acknowledgements

This research was supported in part by the Natural Science and Engineering Research Council of Canada grant A0884 „

References

[I] G.N. Carlson. *Reference to Kinds in English,* Garland Publishing. 1980.

[2] G.N. Carlson, "Generic Terms and Generic Sentences". *Journal of Philosophical Logic 11.* 1982. pp 145-181.

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

[4] J.P. Delgrande. "A Propositional Logic for Natural Kinds". *A1-86 Canadian Society for Computational Studies of Artificial Conference*. May 1986

[5] J.P. Delgrande. "A First-Order Logic for Prototypical Properties: Extended Report". Technical Report 86-15. Laboratory for Computer and Communications Research. School of Computing Science. Simon Fraser University. 1986.

[6] J.P. Delgrande. "An Approach to Default Reasoning Based on a First-Order Conditional Logic", *American Association for Artificial Intelligence Conference.* Seattle. July 1987.

[7] C. Groeneboer. "Theorem Proving in a Propositional Conditional Logic". M.Sc. thesis (in preparation). School of Computer Science. Simon Fraser University. 1987

[8] G.E. Hughes and M.J. Cresswell. *An Introduction to Modal Logic.* Methuen and Col. Ltd.. 1968.

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

[10] D. Lewis. *Counterfactuals*. Harvard University Press. 1973.

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

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

[13] R.C. Moore. "Semantical Considerations on Nonmonotonic Logic". *Proc. IJCAI-83,* Karlsruhe. 1983. pp 272-279.

[14] D. Nute. "Counterfactuals". *Notre Dame Journal of Formal Logic, Vol 16,* 1975. pp 476-482.

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

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

[17] R. Reiter and G. Criscuolo. "On Interacting Defaults". *Proc. IJCAI-81.* Vancouver. 1981. pp 270-276.

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

[19] R.F. Stalnaker. "A Theory of Conditionals", in *Studies in Logical Theory,* N. Rescher (ed.), Basil Blackwell. Oxford. 1968. pp 98-112.

[20] T.R. Thompson. "Parallel Formulation of Evidential-Reasoning Theories" *Proc. IJCAI-85,* Los Angeles. 1985. pp 321-327

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