1

Partial Models and Non-monotonic Inference

K. Konolige

Artificial Intelligence Center, SRI International and
Center for the Study of Language and Information, Stanford University, USA

Abstract

The non-monotonic character of common-sense reasoning is well recognized, as we often jump to conclusions that are not strictly justified by our partial knowledge of a situation. Most formalizations of this idea are best described as syntactic transformations on theories, with little or no semantic underpinnings. In this paper we develop a method of non-monotonic reasoning from a strictly semantic viewpoint, namely, as conjectures about how the missing information in a partial model should be filled in. The advantages of this approach are a natural and intuitively satisfying formalization of diverse types of non-monotonic reasoning, among them domain closure, the unique names hypothesis, and default reasoning.

1. INTRODUCTION

The importance of non-monotonic reasoning for common-sense domains is widely recognized in the field of Artificial Intelligence (AI). In this paper we will be concerned with such reasoning in its most general form, that is, in inferences that are *defeasible*: given more information, we may retract them.

The purpose of this paper is to introduce a form of non-monotonic inference based on the notion of a *partial model* of the world. We take partial models to reflect our partial knowledge of the true state of affairs. We then define non-monotonic inference as the process of filling in unknown parts of the model with *conjectures*: statements that could turn out to be false, given more complete knowledge. To take a standard example from default reasoning: since most birds can fly, if Tweety is a bird it is reasonable to assume that she can fly, at least in the absence of any information to the contrary. We thus have some justification for filling in our partial picture of the world with this conjecture. If our knowledge includes the fact that Tweety is an ostrich, then no such justification exists, and the conjecture must be retracted.

Of course, there are many different ways to represent partial knowledge of the world; in AI, first-order theories (FOTs) are a widely used method. However, FOTs are in a sense *too* partial for the purpose of non-monotonic inference—it is often difficult to decide just how the 'partial' should be filled. For example, consider the sentence

$$Bird(Tweety) \vee Bird(Opus) \tag{1.1}$$

This sentence gives us partial information about the world, in the sense we know either Tweety or Opus (or both) is a bird; but given just (1.1) it is impossible to conclude that we know Tweety to be a bird, or that we know Opus to be a bird.

Now suppose we are given a default rule stated informally as

In the absence of conflicting information, assume that a bird flies. (1.2)

How can this rule be applied to our bird theory (1.1) to make conjectures about the ability of Tweety and Opus to fly? One approach is to relate the application of the default to a consistency condition on the theory, as in the default theories of Reiter (1980). Roughly speaking, our informal rule translates into the following rule for extending an FOT:

If in a theory x is a bird and it is consistent to assume that x can fly, do so. (1.3)

Unfortunately such a default rule yields no new information when applied to (1.1). The disjunction does not permit us to conclude that any particular individual is a bird, and so it is impossible to instantiate the variable x in the antecedent of the default rule.

But clearly our intuitions are that (1.2) tells us something more about the theory (1.1). Suppose we ask what possible partial states of affairs would make (1.1) true. One of the following two is a minimally necessary condition:

1. Tweety is a bird.
2. Opus is a bird.

Now the application of the default rule is straightforward for each case, so we conjecture that either Tweety or Opus can fly.

One conclusion to be drawn from this example is that default reasoning should be based on an analysis of the models that a theory admits. It is the claim of this paper that partial models are an appropriate and natural level of description for the application of default rules, and other types of non-monotonic reasoning as well. In the next section, we support this claim by discussing general principles for implementing non-monotonic reasoning as conjectures on partial models, and by criticizing another model-based framework for non-monotonic inference, McCarthy's (1980, 1984) circumscription schema, from this point of view. The rest of the paper is devoted to illustrating the general principles using a particular type of partial model based on Hintikka interpretations, defined in Section 3. Because these models use the constants of a theory as their domain, they admit very natural treatment of assumptions involving equality and the naming of individuals, which are illustrated in Section 4, along with other types of default reasoning, including domain closure and the assumption of disjoint domains.

2. A SEMANTICS FOR NON-MONOTONIC INFERENCE

In this section we consider some general principles of a partial-model approach to non-monotonic inference, and introduce notation to be used throughout the paper. An analysis of circumscription based on these principles is also presented.

2.1. Conjectures on partial models

Any consistent set of sentences (or *theory*) T in a first-order language is satisfied by a set of (first-order) models. To continue the example from the Introduction: let *Tweety* refer to the individual TWEETY, and *Opus* to OPUS, and let BIRD and FLY be the properties of being a bird and flying, respectively. Now consider the models of $Bird(Tweety) \vee Bird(Opus)$:

$$
\begin{array}{lll}
M_1 = \text{BIRD}: \{\text{TWEETY}\} & \text{FLY}: \{\} & \\
M_2 = \text{BIRD}: \{\text{TWEETY}\} & \text{FLY}: \{\text{TWEETY}\} & \\
M_3 = \text{BIRD}: \{\text{TWEETY}, e_1\} & \text{FLY}: \{\} & \\
M_4 = \text{BIRD}: \{\text{TWEETY}, e_1\} & \text{FLY}: \{\text{TWEETY}\} & \\
M_5 = \text{BIRD}: \{\text{TWEETY}, e_1\} & \text{FLY}: \{\text{TWEETY}, e_1\} & \\
M_6 = \text{BIRD}: \{\text{TWEETY}, e_1, e_2\} & \text{FLY}: \{\} & \\
\quad \vdots & & \\
M_i = \text{BIRD}: \{\text{OPUS}\} & \text{FLY}: \{\} & (2.1) \\
M_{i+1} = \text{BIRD}: \{\text{OPUS}\} & \text{FLY}: \{\text{OPUS}\} & \\
M_{i+2} = \text{BIRD}: \{\text{OPUS}, e_1\} & \text{FLY}: \{\} & \\
M_{i+3} = \text{BIRD}: \{\text{OPUS}, e_1\} & \text{FLY}: \{\text{OPUS}\} & \\
M_{i+4} = \text{BIRD}: \{\text{OPUS}, e_1\} & \text{FLY}: \{\text{OPUS}, e_1\} & \\
\quad \vdots & & \\
M_j = \text{BIRD}: \{\text{OPUS}, \text{TWEETY}\} & \text{FLY}: \{\} & \\
\quad \vdots & &
\end{array}
$$

These models naturally fall into two groups, corresponding to one of the two disjuncts in the theory: either Tweety is a bird, or Opus is (there are models such as M_j in which both these are true; such models fall into both groups). We can represent these groups by using the notion of a *partial model.* A partial model contains only a part of the information necessary in a (complete) model; by *extending* the partial model, we arrive at a set of models. In this example, we could construct two partial models by specifying just a part of the extension of the BIRD relation:

$$m_1 = \text{TWEETY} \in \text{BIRD}$$
$$m_2 = \text{OPUS} \in \text{BIRD}. \quad (2.2)$$

The extension of m_1 includes $M_1 - M_5$ and M_j; the extension of m_2 includes $M_i - M_{i+4}$ and M_j. We write $E(m)$ for the set of extensions of a partial model m.

We have in m_1 and m_2 a formal model-theoretic counterpart of the informal reasoning we carried out in the Introduction. We can formulate the default rule 1.2 as the following conjecture:

If within a partial model x is a bird and it is consistent to assume that x flies, do so. (2.3)

Note that this is exactly the default rule (1.3), except that 'theory' has been changed to 'partial model'. A proposition P is *consistent* with a partial model m if there is an extension of m satisfying P. In the case of m_1 there are models in which Tweety flies, and so (2.3) picks out just that subset $\{M_2, M_5, \ldots\}$ of $E(m_1)$; similarly, for m_2 we get the subset $\{M_{i+1}, M_{i+4}, \ldots\}$. Since the world could be described by either m_1 or m_2, we take the union $\{M_2, M_5, M_{i+1}, M_{i+4}, \ldots\}$ of these models as the result of default reasoning. Obviously, $Fly(Tweety) \vee Fly(Opus)$ is satisfied by each of these models.

To sum up: let T be a theory and α a conjecture on partial models. A conjecture picks out a non-empty subset of the extensions of a partial model. Non-monotonic inference can be viewed as the following process:

1. Let $\mathbf{M}$ be all models of T. Form a set of all partial models $\mathbf{m}$. Let $\mathbf{M}'$ be $\mathbf{M} - E(\mathbf{m})$, i.e. all models not in the extension of some member of $\mathbf{m}$.
2. Let $C(\alpha, \mathbf{m})$ be the set of extensions of $\mathbf{m}$ chosen by the conjecture.
3. We say that a set of sentences T' is *inferred by α from T* if every member of T' is satisfied by each of $\mathbf{M}' \cup C(\alpha, \mathbf{m})$. We write this as $T \vdash_\alpha T'$.

Remarks. The general nature of non-monotonic inference here is the pruning of the set of models of a theory. For any given language L, we may have in mind certain types of models, the intended interpretations of L. For example, in studying resolution, we restrict our attention to Herbrand interpretations, in which all terms denote themselves.

We consider some general technical points of this definition. First, the inference operator $T\vdash_\alpha$ can be non-monotonic in T, as is easily shown by example. Let α be the conjecture that picks out only those extensions of a partial model in which P is false. We have:

$$\{Q\} \vdash_\alpha \neg P \quad (2.4)$$

but

$$\{Q, P\} \nvdash_{\alpha} \neg P. \tag{2.5}$$

A special case, which is monotonic in T, is the conjecture δ that picks out *all* extensions of a partial model. The operator $\vdash_{\delta}$ is simple logical deduction, that is, for $T \vdash_{\delta} T'$, T' is the set of logical consequences of T, and hence also deductive consequences, by the completeness theorem for first-order logic.

Because conjectures pick out a subset of the possible models of T, the inference operator has the reflexive property

$$T \vdash_{\alpha} T. \tag{2.6}$$

Conjectures are thus appropriate for default reasoning or defeasible reasoning in general, where the initial facts, though sparse, are assumed to be accurate. There are, of course, other types of non-monotonic reasoning that are not naturally expressed as conjectures: for example, events are often treated formally as arbitrary transformations on models, and the revision of belief on the basis of new information requires changing a theory to admit models it did not originally have.

There is no guarantee that partial models exist, or if they do, that their extensions fully cover the set of models **M**. **M**′ is designed to take up the slack in these situations, so that all models of **M** are 'accounted for'. This, and the fact that conjectures are a pruning operation on sets of models, yield the following consistency property for the inference operator: if the initial theory T is consistent, then any set of inferred sentences is also consistent; that is, it is impossible to have

$$T \vdash_{\alpha} p \wedge \neg p. \tag{2.7}$$

The notion of the *coverage* of partial models is an important one, and is in some sense a completeness criterion for this method. If there are *no* partial models for a given theory, then for every conjecture α the operator $\vdash_{\alpha}$ becomes logical deduction, and no non-monotonic inference takes place. If the partial models of a theory fully cover the intended models (that is, every intended model is an extension of some partial model), then a conjecture on the partial models takes into account all of the interpretations of the theory. For example, the two partial models (2.2) cover all the models of $T = Bird(Tweety) \vee Bird(Opus)$, and so the conjecture (2.3) gives us the maximum restriction on the models of T. An important feature of conjectures is that they degrade gracefully when some models are not covered, either because of theoretical or computational limitations. If for some reason only m_1 is used as a partial model of T, then the conjecture (2.3) produces the weaker result

$$T \vdash_{\alpha} Bird(Tweety) \supset Fly(Tweety). \tag{2.8}$$

One of the strengths of the method is that there are many different ways to construct partial models of the world. The type of partiality we choose to represent will influence the nature of the non-monotonic operator $\vdash_\alpha$. For example, we might take partial models to be a subset of each relation's (positive) extension, as we did in (2.2); data bases are often viewed in this way (Gallaire *et al.*, 1978). A partial model of this sort covers a set of models that agree on the common subset, but can otherwise disagree. It invites the conjecture that the subset is the complete extension: there are no other true positive facts about the world (sometimes referred to as the *closed-world assumption*; see Section 4).

An important type of partiality, and one we will exploit for most of the remainder of this paper, is the ability to leave unspecified the equality (or inequality) of terms in a theory. One way to do this is by introducing syntactic elements into the partial models, as we do with Hintikka sets in Section 3. Partial models then become sets of atoms and their negations, including equality predications. For example, the set $\{Bird(Tweety), Bird(Opus)\}$ has extensions in which Tweety and Opus are the same individual, and in which they are different. Assumptions about the uniqueness of named individuals can be framed in terms of conjectures on this partial model.

2.2. Circumscription

Predicate circumscription is a proof-theoretic technique in which an FOT T is augmented by a circumscription formula. We can summarize its current formulation (from Etherington *et al.*, 1984) as follows: let P be a predicate, and P' a finite sequences of predicates of a finite theory T. Then $Circ(T, P, P')$ is a particular second-order formula expressing the circumscription of P, letting the predicates P' vary.

The semantics of circumscription come from the notion of P-minimal models. A model M is P-minimal if there is no other model N, agreeing with M everywhere except for the predicates P and P', such that the extension of P in N is a proper subset of that in M. Circumscription is sound with respect to minimal models, in the sense that $Circ(T, P, P')$ is true in all P-minimal models of T; however it is known to be incomplete (these results are summarized in Minker and Perlis, 1985).

Partial model conjectures have close ties with reasoning about minimal models. In fact, we can express the intended semantics of circumscription as a conjecture in the following way. We take partial models to be the P-minimal models, where the extension of a partial model M is the set of all models N which agree with M, except possibly on P and P'. The conjecture α is to pick only the minimal model itself.

Reasoning about minimal models was first employed in AI by McCarthy (1980) in an attempt to deal with what he called the *qualification problem.* In brief, this is the problem of stating formally what objects and conditions *do not obtain* in a given situation. Using minimal models is a means of applying Occam's razor: only those objects are assumed to exist that are actually required by the statements of a theory.

It is not clear, however, that reasoning in minimal models is the best means of performing defeasible reasoning in general. For example, it can lead to a complicated statement of defaults by means of an abnormality predicate. Compare the compact formulation of Example 4.9 with the corresponding circumscriptive rendering on pp. 300–302 of McCarthy (1984). But the evidence here is not yet in, and awaits a fuller exploration of the application of circumscription.

With regard to assumptions about equality, certain inherent limitations are already known (see Etherington *et al.*, 1984). Because minimal models are defined with respect to a fixed denotation function for the terms of a theory, it is impossible to perform non-monotonic inferences about the equality of terms by reasoning in such models. However, there have been attempts to account for equality by importing names and their denotations as objects of the domain (Lifschitz, 1984; McCarthy, 1984).

By contrast, we can choose partial models in such a way that non-monotonic inferences about equality are possible. As we show in the next section, partial model conjectures enjoy a natural treatment of assumptions about equality, including domain closure and the unique names hypothesis.

Finally, non-monotonic inference using partial model conjectures has been defined to always yield a consistent extension for a theory. For circumscription this is not the case, unless every model of the theory is an extension of a minimal model. In those instances where this is not the case, it has been shown that the circumscription formula can be inconsistent with an originally consistent theory (Etherington *et al.*, 1984).

3. HINTIKKA SETS AS PARTIAL MODELS

We now introduce a particular type of partial model, based on the method of analytic tableaux. [We do not give more than a cursory presentation of this method. See Smullyan (1968) for a general introduction; the method used here is based on work by Hintikka (1955).] Consider a theory T, which may be infinite. A tableau for T is a tree whose nodes are sentences, constructed in the following manner. The root of the tree is an arbitrary element of T. The tree is grown in a systematic manner from its leaves by adding new nodes, either elements of T or sentences derived from previous nodes by a small set of rules. Some of these rules, those dealing with disjunction, cause splits in the tree. The end result is a (perhaps infinite, but finitely branching) tree with an important property. All of the branches which are not *closed* (no sentence and its negation both appear on the branch) are partial models of T of a certain sort. To take a very simple example, here is a tableau for the theory $T = Bird(Tweety) \vee Bird(Opus)$:

$$Bird(Tweety) \vee Bird(Opus)$$

$$Bird(Opus) \qquad\qquad Bird(Tweety)$$

There are two open branches, each comprising one partial model. The partial models are just sets of atoms and their negations. Our intended models, then, are *Herbrand interpretations* of the theory, in which all terms of the language denote themselves.

The method of analytic tableaux has two pleasing properties, which we state here without proof. The first is that a tableau for T forms a complete survey of all of the Herbrand interpretations of T: any such interpretation is isomorphic to the extension of some open branch of the tableau. The second is that the partial models make the minimal assumptions about the equality of terms. This is a consequence of the particular effect of the rule that operates on existential statements. If a node in the tree has the form $\exists x . A[x]$, this rule allows the introduction of a node $A[a]$, where the constant a is *new*, that is, has not been previously used in the tableau. So, in effect, this rule makes no assumption about the identity of the individual satisfying $A[x]$, since a could be equal or not equal to any term already in the tableau.

3.1. Hintikka sets

We now give a formal definition of partial models based on tableaux. Let L be a first-order language with equality and constant symbols, but no other function symbols. $A[\bar{x}]$ is a formula in which the free variables $\bar{x} = x_1, x_2, \ldots, x_n$ occur. $A[\bar{a}]$ is the result of substituting a_i for each corresponding x_i. If S is a set of formulas, the *universe* of S is the set of constants of S.

The set of sentences on an open branch of a tableau is called a *Hintikka set.* A Hintikka set S with universe U satisfies the following conditions.

1. If $\neg A \in S$, then $A \notin S$.
2. If $A \wedge B \in S$, then both $A \in S$ and $B \in S$.
3. If $A \vee B \in S$, then either $A \in S$ or $B \in S$.
4. If $\exists x . A[x] \in S$, then for some constant $a \in U$, $A[a] \in S$.
5. If $\forall x . A[x] \in S$, then for all constants $a \in U$, $A[a] \in S$
6. $a \neq a \notin S$ for all $a \in U$.
7. If $a = b \in S$, then $A[a] \in S$ if and only if $A[b] \in S$.

In a Hintikka set, the equality predicate obeys the standard constraints of symmetry, reflexivity, transitivity, and substitution. Here are some examples of Hintikka sets that are open branches of tableaux, rooted in the first element of the set.

Example 3.1. $\{Pa \vee Pb, Pa\}$ with universe $\{a, b\}$.

Example 3.2. $\{Pa \wedge Pb \wedge \forall x.x = a, \quad \forall x.x = a, \quad Pa, \quad Pb, \quad a = a, \quad b = a\}$ with universe $\{a, b\}$.

Example 3.3. $\{\exists x.\forall y.Rxy, \forall y.Ray, Raa\}$ with universe $\{a\}$.

Example 3.4. $\{\forall x.\exists y.Rxy, \exists y.Ra_1y, Ra_1a_2, \exists y.Ra_2y, Ra_2a_3, \exists y.Ra_3y, Ra_3a_4, \ldots\}$ with universe $\{a_1, a_2, \ldots\}$.

Example 3.4 has an infinite domain.

Example 3.5. Let H be the conjunction of

1. $\forall x.\exists y.Rxy$
2. $\exists y.\forall x.\neg Rxy$
3. $\forall xyz.Rxy \wedge Rxz \supset y = z$
4. $\forall xyz.Ryz \wedge Rzx \supset y = z$

There is one Hintikka set S of T, for which

$$v(S) = \bigcup_{i \geq 0} \{Ra_ia_{i+1}\} \cup \bigcup_{j \neq i+1} \{\neg Ra_ia_j\}$$

If we interpret the constant a_i as the natural number i, this model is the natural numbers with $a_0 = 0$ being the least element.

In a Hintikka set S with universe U, the atoms (atomic sentences) and their negations comprise a partial interpretation over the domain U, by specifying some of the positive and negative instances of relations. This interpretation is Herbrand in the sense that each constant in U refers to itself. We call the set of atoms and their negations $v(S)$. $v(S)$ is *complete* if for every atom p it contains either p or $\neg p$.

A complete set $v(S)$ is called an *interpretation set with domain U*. It generates an interpretation by assigning the atom Ra the value *true* if it is in $v(S)$, and *false* if its negation is. The truth value of compound statements is determined by the normal rules for quantifiers and Boolean operators. We write $v(S) \models T$ if every member of T is assigned the value *true* (*i.e.* is satisfied) by $v(S)$, and call $v(S)$ a *model set for T*.

A *partial interpretation* is an atom set $v(S)$ that is not complete. It is an important property of a Hintikka set S that $v(S)$ can always be extended, by the addition of atoms and their negations, to a complete Hintikka set; further, every such extension satisfies every sentence in S. This result is Hintikka's Lemma. If every extension of a partial interpretation satisfies a sentence T, we write $v(S) \models T$, and say that $v(S)$ is a *partial model* for T. Using this notation, we can write Hintikka's lemma as

$$v(S) \models S. \tag{3.1}$$

Example 3.2 has a complete atom set except for equality statements of symmetry and reflexivity (which we will generally ignore from now on, assuming they are always present); it is easy to verify that $v(S) \models S$. Example 3.3 is also complete. Example 3.1 has three extensions, one with $\{a = b, Pb\}$, one with $\{a \neq b, Pb\}$, and one with $\{a \neq b, \neg Pb\}$. Example 3.4 has an infinite number of extensions. Example 3.5 has only one extension, formed by including $\bigcup_{i \neq j} a_i \neq a_j$.

Finally, we want to make sure that the partial models contain the least information compatible with satisfying a theory. Consider the theory $T = Pa \vee (Pa \wedge Pb)$, which is equivalent to Pa. A tableau for T generates two partial models, $\{Pa\}$ and $\{Pa, Pb\}$. However, the second partial model is an extension of the first. So we consider only those partial models of a theory that are *minimal*: there are no partial models which are proper subsets. We have not yet been able to prove that such minimal partial models always exist. However, the situation appears hopeful, because we are restricting our attention to Herbrand models generated by the tableau method. A case in point is the theory T of Example 3.5, which has no minimal model in general, but has a single partial model in its tableau. This corresponds well with our intuitions, since all models of T are isomorphic to the natural numbers.

4. NON-MONOTONIC INFERENCE ON PARTIAL MODELS

We consider four types of non-monotonic inference: the 'unique names' assumption, domain closure, disjoint domains, and default reasoning. All can be defined as conjectures about the *extension* of partial models, that is, we derive a theory T' from T by considering only a subset of the possible extensions of partial models for T.

By way of exposition, we first present a simple conjecture called *negative extension* (or *NE*) which is closely related to circumscription and the closed world assumption. Given a partial model, *NE* picks out that subset of its extensions which are maximal in negative atoms. Put another way, *NE* 'fills in' the missing information in a partial model by always adding negated atoms. There is an exception for equality: no equality atoms are added.

Example 4.1. Let $T = Pa \vee (Pb \wedge Pc)$; there are two partial models, $\{Pa\}$ and $\{Pb, Pc\}$. These give the following conjectured extensions under *NE*:

$$\begin{array}{l} \{Pa\} \\ \{Pa, \neg Pb\} \\ \{Pa, \neg Pb, \neg Pc\} \\ \{Pa, \neg Pb, \neg Pc, \neg Pe_1\} \\ \vdots \end{array} \quad (4.1)$$

and

$$\{Pb, Pc\}$$
$$\{Pb, Pc, \neg Pa\} \qquad (4.2)$$
$$\{Pb, Pc, \neg Pa, \neg Pe_1\}$$
$$\vdots$$

In the first of these, the sentence $\forall x \,.\, Px \supset x = a$ is satisfied; in the second, $\forall x \,.\, Px \supset (x = b \vee x = c)$. We have

$$T \vdash_{NE} T \wedge [(\forall x \,.\, Px \supset x = a) \vee (\forall x \,.\, Px \supset (x = b \vee x = c))] \qquad (4.3)$$

or, equivalently,

$$(Pa \wedge \forall x \,.\, Px \supset x = a) \vee (Pb \wedge Pc \wedge \forall x \,.\, Px \supset (x = b \vee x = c)). \qquad (4.4)$$

This result can be compared with the circumscription of T (with respect to P), which yields:

$$(Pa \wedge \forall x \,.\, Px \supset x = a) \vee$$
$$(Pb \wedge Pc \wedge \forall x \,.\, Px \supset (x = b \vee x = c) \wedge a \neq b,c). \qquad (4.5)$$

The difference lies in the treatment of equality. In general, the conjecture NE will make fewer assumptions about the equality of terms than the corresponding circumscription.

4.1. The unique names hypothesis

This is the assumption that distinct names refer to distinct objects. The term was introduced by Reiter (1980) in formalizing a common naming convention in data bases. It is often useful to make this assumption in some form in common-sense reasoning. There are three additional criteria in this case:

1. It should be possible to state in the theory T that some distinct terms are equal (e.g. *Morningstar* = *Eveningstar*).
2. It should also be possible to exclude terms from the unique names hypothesis by saying for a term that it may be equal to another term, without saying what that other term is. Skolem constants, for example, have this property.
3. The assumption of unique names should be defeasible, because it may turn out later that two distinct names actually do refer to the same individual.

It is very easy to implement the unique names assumption for partial models, because they are defined to minimize equality assumptions among names: it is the conjecture that adds as many equalities as possible to a partial model. More formally, we define $UN\colon\{a_1, a_2, \ldots\}$ as the conjecture that picks out those extensions of a partial model which are maximal in inequalities of the form $a_i \neq a_j$, $i \neq j$. This forces the names a_1, $a_2, \ldots$ to be maximally unique, i.e. two names a_i and a_j are considered to refer to different individuals unless the partial model forces them to be the same.

Example 4.2. Continuing Example 4.1, let $T = Pa \vee (Pb \wedge Pc)$. $T \vdash_{UN:\{a,b,c\}} a \neq b \wedge b \neq c \wedge a \neq c$.

Example 4.3. Let T be as defined in Example 3.5. $T \vdash_{UN:\{a_1,\ldots\}} T$.

Example 4.4. Let T be the sentence $ms = es \wedge x_0 \neq c$. In this example, es and ms are constants that refer to the same individual, while x_0 is a skolem constant, i.e. it may refer to some individual already named; however it cannot refer to c. We form the UN conjecture for the constants ms, es, and c:

$$T \vdash_{UN:\{ms,es,c\}} T \wedge c \neq ms. \tag{4.6}$$

That is, c is different from ms (and es), and it is not known whether x_0 is the same as ms or not. The conjecture that $c \neq ms$ is defeasible, because if it is later learned that $c = ms$, this can be added to T, and

$$T \wedge c = ms \vdash_{UN:\{ms,es,c\}} T \wedge c = ms. \tag{4.7}$$

This last example shows that the three criteria above for the unique names hypothesis are satisfied by the conjecture UN.

4.2. Domain closure

This term was originated by Reiter (1980) in the logical reconstruction of data base theory, as the assumption that only the individuals mentioned by the data base exist. If $a_1, a_2, \ldots$ are the constants of the data base, then the domain closure axiom is

$$\forall x\,.\,(x = a_1 \vee x = a_2 \vee \cdots). \tag{4.8}$$

This axiom is finite if the number of constants is finite. In data base theory, domain closure is usually invoked along with the unique names hypothesis, so that there are as many individuals as there are constants.

We will pursue a more semantically oriented form of domain closure, so that we can work with theories (such as $T = \forall x.\exists y.Rxy$) that have models containing individuals not explicitly named in the theory. Call DC the assumption that *only those individuals exist that are minimally required to satisfy a theory*. This idea was first articulated by McCarthy (1980). Note that it is a stronger assumption than domain closure, because it picks out models that are minimal in their domains. For $T = Pa \vee Pb$, domain closure only implies $\forall x\,.\,(x = a \vee x = b)$, while the minimality requirement forces $(\forall x\,.\,x = a) \vee (\forall x\,.\,x = b)$, i.e. either everything is a, or it is b.

In terms of conjectures on partial models, DC picks out the extensions of a partial model with the fewest individuals. For Hintikka sets, DC is implemented by picking extensions V of a partial model $v(S)$ with the following properties:

1. The universe of V is the universe of S.
2. V is maximal in positive equality atoms, i.e. there is no extension of $v(S)$ that contains more occurrences of positive equality atoms.

The first condition conjectures that all actual individuals have already been named in S (this is Reiter's domain closure), and the second attempts to give as many names as possible a common interpretation.

Example 4.5. Let S be the Hintikka set of Example 3.4. All equalities of the form $a_i = a_j$ can be consistently added, so DC picks out the single extension

$$\left\{v(S) \cup \bigcup \{a_i = a_j\}\right\}. \tag{4.9}$$

This is the Hintikka set with the smallest number of individuals satisfying $\forall x.\exists y.Rxy$. We thus have

$$\forall x.\exists y.Rxy \vdash_{DC} \forall xy\,.\,x = y \tag{4.10}$$

which means that DC conjectures a one-element domain.

As noted, in general DC is a stronger assumption than domain closure. As this next example shows, DC tries to identify different names with the same individual, thereby reducing the size of the domain; often this is not the desired result.

Example 4.6. Let T be $Pb \wedge Pc$. Then

$$T \vdash_{DC} \forall x\,.\,x = c \wedge \forall x\,.\,x = b \wedge b = c. \tag{4.11}$$

In practice, it would be useful to couple DC with a conjecture about the uniqueness of names, as is done in data base theory (see Reiter, 1980). We define the conjecture $UN{:}\bar{a}$; DC as first taking the subset of complete extensions based on $UN{:}\bar{a}$, then further pruning these extensions by the DC conjecture.

Example 4.7. Let T be as in Example 4.6.

$$T \vdash_{UN:\{b,c\};DC} b \neq c \wedge \forall x\,.\,(x = b \vee x = c). \tag{4.12}$$

4.3. Disjoint domains

Often we wish to assume that individuals cannot belong to two different groups, e.g. one is normally either a Democrat or a Republican (or neither), but not both. However, we would like this assumption to be defeasible, since it could turn out that a person is registered for both parties, but in different states.

Let $P_1, \ldots, P_n$ be a set of monadic predicates that we assume to be disjoint. The disjoint domain axioms are expressed as:

$$\bigwedge_{i \neq j} \forall x\,.\,P_i x \supset \neg P_j x. \tag{4.13}$$

We cannot just add these axioms to a theory, however, because of the defeasibility condition. However, the disjoint domain assumption can be stated as the following conjecture: $DD:\{P_1 \cdots P_n\}$ picks out those complete extensions of a partial model with exactly one positive atom from the set $\{P_1a, \ldots, P_na\}$, for every constant a.

Example 4.8. Let *Dem x* and *Rep x* be predicates we wish to be disjoint, and let $\{Dem\, a\}$ be a partial model. $DD:\{Dem, Rep\}$ picks out the following extensions:

$$\begin{aligned}&\{Dem\, a, \neg Rep\, a\}\\&\{Dem\, a, \neg Rep\, a, Dem\, e_1, \neg Rep\, e_1\}\\&\{Dem\, a, \neg Rep\, a, \neg Dem\, e_1, Rep\, e_1\}\\&\vdots\end{aligned} \tag{4.14}$$

and thus

$$\begin{aligned}Dem\, a \vdash_{DD:\{Dem,Rep\}} Dem\, a \wedge \neg Rep\, a &\\ \wedge\, \forall x\,.\, Dem\, x \supset \neg Rep\, x &\\ \wedge\, \forall x\,.\, Rep\, x \supset \neg Dem\, x. &\end{aligned} \tag{4.15}$$

The following theorem shows that DD implies the disjoint domain axioms when there is no evidence to the contrary.

Theorem 4.1. Let T be any tautology.

$$T \vdash_{DD:\{P_1 \cdots P_n\}} \bigwedge_{i \neq j} \forall x\,.\, P_i x \supset \neg P_j x \tag{4.16}$$

Proof. The single partial model of a tautology is the empty set. $DD:\{P_1 \cdots P_n\}$ will pick out all models in which $P_i a$ and $P_j a$ do not coexist, for all a and $i \neq j$.

4.4. Default reasoning

Default reasoning is the assumption of propositions that have a reasonable chance of being true, given the available information. Formally, we implement defaults as conjectures on partial models, where the conjecture mentions specified and unspecified parts of the model. Consider, for example, the default rule that birds normally fly. Let *Bird x* and *Fly x* be the relevant predicates. We express the application of the default to a partial model $v(S)$ by the rule

> If *Bird x* is specified by $v(S)$ and $\neg Fly\, x$ is not specified by $v(S)$, then only consider those extensions of $v(S)$ that specify *Fly x*. (4.17)

We need a language for expressing default conjectures of this sort. A simple one can be formed by considering quantifier-free formulas, perhaps with free variables, together with the monadic operators $\Box$ and $\Diamond$. The default rule 4.17 would be expressed as

$$\alpha = \Box Bird\, x \wedge \Diamond Fly\, x \Rightarrow Fly\, x \tag{4.18}$$

$\Box p$ means that p is true in *all* extensions of the partial model, and $\Diamond p$ that p is true in *some* extension ($\Diamond = \neg\Box\neg$). If the formula to the left of the $\Rightarrow$ sign is satisfied by a partial model $v(S)$ for some instantiation $x = a$, then α picks out only those extensions of $v(S)$ containing $Fly\ a$, the instantiated expression to the right of $\Rightarrow$. Again, it is useful to think of α as 'filling in' a partial model that contains $Bird\, x$, by adding $Fly\, x$ if possible.

Example 4.9. This is from McCarthy (1984). There are ostriches, penguins, and canaries. Unless a bird is known to be an ostrich or penguin, we assume that it can fly. Let

$$T = \forall x\,.\ Ostrich\, x \supset \neg Fly\, x \wedge \forall x\,.\ Penguin\, x \supset \neg Fly\, x \tag{4.19}$$

and let α be the conjecture (4.18) above. We have

$$T \wedge Bird\ a \vdash_{\alpha} Fly\ a. \tag{4.20}$$

It is interesting to combine the default rule with assumptions about the disjointness of canaries, ostriches, and penguins:

$$T \wedge Canary\ a \vdash_{\alpha;DD:\{Ostrich,Penguin,Canary\}} Fly\ a \wedge \neg Ostrich\ a \wedge \neg Penguin\ a \tag{4.21}$$

and

$$T \wedge Ostrich\ a \vdash_{\alpha;DD:\{Ostrich,Penguin,Canary\}} \neg Fly\ a \wedge \neg Canary\ a \wedge \neg Penguin\ a \tag{4.22}$$

5. CONCLUSION: SOME ISSUES

5.1. Multiple conjectures

One of the pleasing aspects of a partial model approach is that conjectures of different sorts can be intermixed, as in the domain closure and unique names hypothesis of Example 4.7, and disjoint domain and default rules in Example 4.9. In both these examples there is an obvious order of application of the conjectures. Ordering is important because we first prune possible extensions with one conjecture, and then apply another conjecture to the result; doing this in a different order can lead to different sets of extensions. Ordering is a useful property when conjectures have readily defined priorities; however, especially in the case of default rules, conjectures may have roughly equal weights. Reiter (1980) gives the example the Republicans are normally non-pacifists and Quakers pacifists, so what about Richard Nixon, who is both a Quaker and Republican? As conjectures, these are

1. $\Box Rep\ x \wedge \Diamond\neg Pacifist\ x \Rightarrow \neg Pacifist\ x$.
2. $\Box Quaker\ x \wedge \Diamond Pacifist\ x \Rightarrow Pacifist\ x$.

If we apply (1) first, Nixon will be a non-pacifist; if (2), a pacifist. If both defaults are equal in their plausibility, it would be better not to conclude anything. We could try applying them in parallel, that is, taking the *intersection* of the extensions allowed by (1) and (2). However, this intersection would be empty in the present case, an undesirable result (and, by definition, not a conjecture). Instead, we might use the following rules:

1. $\Box Rep\ x \wedge \neg\Box Quaker\ x \wedge \Diamond\neg Pacifist\ x \Rightarrow \neg Pacifist\ x$.
2. $\Box Quaker\ x \wedge \neg\Box Rep\ x \wedge \Diamond Pacifist\ x \Rightarrow Pacifist\ x$.

But this is not entirely satisfactory either, because the modularity of the rules is compromised.

5.2. Proof theory

This is still unexplored. However there are some directions that appear promising.

1. Theories with finite tableaux. If a theory has a finite tableau, then it has a finite number of partial models, and it is possible to work directly with these. The chief syntactic class with this property are the $\exists\forall$-theories: those whose existential quantifiers all precede universals in prenex form.

2. Approximations based on one or a few partial models of a theory. All of the atoms (positive and negative) of the theory are kept as a partial model, while the more complicated axioms of the theory are treated procedurally as a means of deriving more atoms. Important disjunctions may be split into cases, producing more than one partial model. Note that this is the strategy of typical first-order AI knowledge bases (Nilsson, 1980). For syntactic classes that have a unique partial model, this is a complete technique. An interesting example here is provided by Horn-clause theories, for which the unique partial model is the intersection of all their Herbrand models.

Acknowledgements

I thank Benjamin Grosof and especially David Israel for enlightening discussions on this topic. This research was made possible in part by a gift from the System Development Foundation. It was also supported by Grant N00014-80-C-0296 from the Office of Naval Research.

REFERENCES

Etherington, D. W., Mercer, R. E., and Reiter, R. (1984) On the adequacy of predicate circumscription for closed-world reasoning. *AAAI Workshop on Non-Monotonic Reasoning*. American Association for Artificial Intelligence, Menlo Park, Calif.

Gallaire, H., Minker, J., and Nicholas, J. M. (1978) An overview and introduction to logic and data bases. In *Logic and data bases* (eds H. Gallaire and J. Minker). Plenum Press, New York.
Hintikka, K. J. J. (1955) Form and content in quantification theory. *Acta Philosophica Fennica* **8**, 7–55.
Lifschitz, V. (1984) Some results on circumscription. *AAAI Workshop on Non-Monotonic Reasoning*. American Association for Artificial Intelligence, Menlo Park, Calif.
McCarthy, J. (1980) Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence* **13**, 1–2.
McCarthy, J. (1984) Applications of circumscription to formalizing common sense knowledge. *AAAI Workshop on Non-Monotonic Reasoning*. American Association for Artificial Intelligence, Menlo Park, Calif.
Minker, J. and Perlis, D. (1985) Completeness results for circumscription. University of Maryland, College Park, Md. Unpublished manuscript.
Nilsson, N. (1980) *Principles of artificial intelligence*. Tioga, Palo Alto, Calif.
Reiter, R. (1980) A logic for default reasoning. *Artificial Intelligence* **13**, 1,2.
Reiter, R. (1980) Equality and domain closure in first-order data bases. *J. Association for Computing Machinery* **27**(2), 235–49.
Smullyan, R. M. (1968) *First-order logic*. Springer-Verlag, New York.