Consiglio Nazionale delle Ricerche

Istituto di Elaborazione dell'Informazione

A Proof-Theoretic Account of Model-Preference Default Reasoning

Fabrizio Sebastiani

Nota Interna IEI-B4-05-1990

Via S. Maria, 46 - 56126 Pisa (Italy)
Phone : +39.(0)50.553159
Fax : +39.(0)50.554342
Telex : 590305 IEICNR I

A Proof-Theoretic Account of Model-Preference Default Reasoning

Fabrizio Sebastiani[1]
Istituto di Elaborazione dell'Informazione
Consiglio Nazionale delle Ricerche
Via S. Maria, 46 - 56126 Pisa (Italy)

Technical Report IEI-B4-05-1990

Abstract

The "model-preference" account of default reasoning recently proposed by Selman and Kautz overcomes many limitations of previous default formalisms, as it has a strong model-theoretic flavour and provides a formal justification for the limited cognitive load that default reasoning seems to require of human beings. In this paper we describe $\mathcal{L}(\mathcal{D}^+)$, a non-standard proof system for model-preference default reasoning; $\mathcal{L}(\mathcal{D}^+)$ is non-standard in the sense that rules have a global (instead of the usual local) character, and that it deals with proofs of the minimality of object level theories rather than with proofs of the theoremhood of formulae. Nonetheless, $\mathcal{L}(\mathcal{D}^+)$ retains the essential character of a proof system, namely, the independence of provability from the order of application of the rules, and may thus prove a useful tool for the integration of this model-preference default reasoning with other forms of reasoning that are typically dealt with proof-theoretically.

1 Introduction

Default reasoning plays an important role in everyday practical reasoning. Agents, be they natural or artificial, typically face situations in which they have to act and take decisions on the basis of a body of knowledge that is far from being an exhaustive description of the domain of discourse; their lack of such a description is a direct consequence of the limited capacity of their physical repositories of knowledge, of the limited throughput of their channels of communication with the external world (e.g. the visual

[1] Current address: Department of Computer Science, University of Toronto, M5S 1A4 Toronto, Ontario, Canada. E-mail: fabrizio@ai.toronto.edu

apparatus), and, above all, of the fact that the processes involved in the acquisition of knowledge (both from external sources -- e.g. books -- and internal ones -- e.g. speculative reasoning) are computationally demanding and time consuming.

Nevertheless, action and decision-making is often so complex to require more than the knowledge the agents actually possess; this forces them to make up with the limited coverage of their knowledge bases (KBs) by means of "default" assumptions which are brought to bear in the reasoning task. As the name implies, "assumptions" are items of knowledge endowed with an epistemic status that is far from being solid: that is, they can be invalidated by further reasoning or by future acquisition of empirical data. These phenomena are well-known in cognitive science, and their lack of resemblance with deductive patterns of reasoning has sometimes been taken to imply that a great deal of human reasoning does not conform to the canons of "logic" and hence escapes attempts at formalization (Johnson-Laird 1983).

Doubtless, the overall effectiveness of human action in the face of incomplete information testifies to the effectiveness of this modality of reasoning: in fact, humans are much quicker at creating surrogates of missing knowledge than at actually acquiring that knowledge in a more reliable way, either through reasoning or empirical investigation, and have the ability to come up with *plausible* surrogates, surrogates that in most occasions turn out to be accurate predictions of the actual reality. Once these surrogates have been created, humans are much quicker at reasoning on the resulting exhaustive, albeit "epistemically shakier", description of the domain of discourse than they would be had they to rely on the smaller part of this description that they trust as being accurate *tout court*. These observations are at the heart of the recent interest that the KR community has shown in *vivid knowledge bases* (Levesque 1986, 1988; Etherington *et al.* 1989), i.e. exhaustive descriptions of the domain of discourse consisting of collections of atomic statements[2]. Reasoning on these KBs, which may be considered as "analogues" of the domain being represented, is easily shown to be efficient.

It is precisely in the face of the above-mentioned empirical considerations that the bad computational properties of current formalisms that attempt to account for default reasoning (such as the formalisms based on Circumscription (McCarthy 1980; 1986) or on Autoepistemic Logic (Moore 1985; Konolige 1987)) are particularly disturbing: arguably, a formalism for default reasoning not only should characterize the class of conclusions that agents draw in the presence of incomplete information, but should also possess radically better computational properties than formalisms accounting for reasoning tasks at which humans are notoriously inefficient (such as e.g. classical logic in the case of deductive

[2] In formally introducing vivid KBs Levesque (1988) actually situates his discussion in the framework of the first order predicate calculus; hence, for him a vivid KB is "a collection of ground, function-free atomic sentences, inequalities between all different constants (...), universally quantified sentences expressing closed world assumptions (...) over the domain and over each predicate, and the axioms of equality" . As our discussion will be situated in the framework of the statement calculus, we will take this definition of vivid KBs instead.

reasoning).

These considerations have lead researchers to look with special interest at formalizations of default reasoning that emphasize computational tractability. In their recent paper "The complexity of model-preference default theories" (hereafter [MPD]), Selman and Kautz (1988) describe $\mathcal{DH}_a{}^+$, a tractable system for performing inferences on acyclic theories of Horn defaults; in this system a vivid, complete KB may be obtained in polynomial time starting from an incomplete one and from an acyclic theory of Horn defaults. This tractability result accounts for what both intuition and empirical evidence suggest us, namely, that in order to obtain KBs upon which subsequent reasoning can be carried out efficiently, humans use a reasoning method that is itself efficient.

The framework described in [MPD], quite similarly to other recent proposals (Shoham 1987; Brown & Shoham 1989), has the added appeal of possessing a strong model-theoretic flavour[3]. In this paper we attempt to complete the picture by describing $\mathcal{L}(\mathcal{D}^+)$, a proof theory for $\mathcal{D}^+$, the most general system described in [MPD] of which $\mathcal{DH}_a{}^+$ is a tractable subset[4]. Quite surprisingly, $\mathcal{L}(\mathcal{D}^+)$ turns out also to be a proof theory for $\mathcal{DH}_a{}^+$ (and for the other subsets of $\mathcal{D}^+$ that are discussed in [MPD]); this happens essentially because $\mathcal{L}(\mathcal{D}^+)$ has no logical axioms and because the above mentioned subsets are obtained from $\mathcal{D}^+$ simply by restricting the representation language[5]. $\mathcal{L}(\mathcal{D}^+)$ is non-standard in nature, because of two quite different reasons:

- it must account for the *global* character of default reasoning, i.e. for the fact that the complete KB, and not single items of it, has to be brought to bear in order to infer a conclusion. When defaults are involved, the relation of logical consequence is not a relation between two formulae, but a relation between *the whole KB* and a formula. We want to emphasize the fact that globality is inherent in the endeavour of default reasoning, and is not a feature of our specific approach to it;
- unlike in more standard proof theories, the minimality of a theory (see below) has to be proven rather than the theoremhood of a formula. Again, concern with a property of a whole theory (i.e. KB) rather than one of a single formula is another aspect of the above mentioned globality.

Nonetheless, $\mathcal{L}(\mathcal{D}^+)$ retains the essential character of a proof system, that is, the independence of the notion of provability from the order of application of the rules. This is also its most interesting feature,

[3] A semantics for Selman & Kautz's model-preference default systems that fully embraces the model-theoretic credo is described in (Sebastiani 1990).

[4] In this paper we will implicitly rule out from consideration the system $\mathcal{D}$, as its lack of commitment to any specificity ordering between defaults (see below) makes it the least interesting among the systems of [MPD]; arguably, the presence of $\mathcal{D}$ in [MPD] is only instrumental to the establishment of the complexity results. The other systems discussed in [MPD], $\mathcal{DH}^+$ and $\mathcal{DH}_a{}^+$, are restrictions of $\mathcal{D}^+$ to the Horn case and to the Horn Acyclic case, respectively.

[5] Quite similarly, resolution is a proof theory both for classical propositional logic and for Horn propositional logic.

the one which makes it a viable alternative to the graph-theoretic approaches to model-preference reasoning that are proposed in [MPD]. In fact, while a graph-theoretic approach seems the more natural choice for dealing with model-preference reasoning *in isolation*, the unifying framework of proof theory is probably the most sensible choice when its integration with other forms of reasoning (being these mostly dealt with proof-theoretically) is considered.

This paper is organized as follows. In order to make it self-contained, in Section 2 we give a brief overview of $\mathcal{D}^+$; this overview is not completely faithful to the original system described in [MPD] in that it incorporates the modifications that have been suggested in (Sebastiani 1989) in order to make $\mathcal{D}^+$ behave correctly in the presence of both certain information and "defeasible" (default) information. In Section 3 we spell out the proposed proof system in detail and describe in which ways $\mathcal{L}(\mathcal{D}^+)$ departs from more traditional ones. In Section 4 we apply $\mathcal{L}(\mathcal{D}^+)$ to a specific problem in default reasoning, namely one that involves reasoning about inheritance hierarchies. Section 5 concludes.

2 An overview of Selman & Kautz's system $\mathcal{D}^+$

Roughly speaking, the idea around which the systems of [MPD] revolve is that the import of a default $d \equiv \alpha \rightarrow q$ is to make a model (that is, an exhaustive specification of what the domain of discourse is like) where both α and q are true be *preferred* to another model where α is true but q is not. By combining the effects of the preferences due to the individual defaults, a set of defaults identifies a set of "maximally preferred" models; as any model identifies one and only one vivid KB, maximally preferred models are meant to represent possible ways in which the agent may "flesh out" his body of certain knowledge by the addition of defeasible knowledge. For instance, according to a set of defaults such as $\{a \rightarrow b, b \rightarrow c\}$, the model where a, b and c are all true would be a maximally preferred model.

However, the systems in [MPD] also account for the fact that a more specific default (i.e. one with a more specific antecedent) should override a less specific one, and they do so by "inhibiting", where a contradiction would occur, the preference induced by the less specific default; for instance, this prevents a set of defaults such as $\{a \rightarrow b, b \rightarrow c, ab \rightarrow \neg c, a\neg b \rightarrow \neg c\}$ to generate maximally preferred models where a and c are both true.

The first thing we need to do in order to introduce $\mathcal{D}^+$ formally is to describe what the language for representing knowledge in $\mathcal{D}^+$ is. Let $P = \{p_1, p_2, ..., p_n\}$ be a finite set of propositional letters and L be the language of literals built from P (a literal being a propositional letter p or its negation $\neg p$). We define a *default* d to be an expression of the form $\alpha \rightarrow q$, where q is a literal and α is a set of literals[6]. We will also use the standard definition of a *model* for L as a function $M: P \dashrightarrow \{\texttt{True},$

[6] For notational convenience we will omit to draw braces in antecedents of defaults. Hence we will

False}; accordingly, we will say that M satisfies a theory T of L (written as $M \mid= T$) iff M assigns True to each literal in T, negation being evaluated with respect to M in the standard manner[7].

The above-mentioned specificity ordering between defaults is captured by postulating that, given a set of defaults D, a default $d \equiv \alpha \rightarrow q$ in D is *blocked* at a model M iff there exists a default d' in D such that $d' \equiv \alpha \cup \beta \rightarrow \neg q$ and $M \mid= \alpha \cup \beta$. A default $d \equiv \alpha \rightarrow q$ is then said to be *applicable* to a model M iff $M \mid= \alpha$ and d is not blocked at M. If d is applicable at M, the model $d(M)$ is defined as the model which is identical to M with the possible exception of the truth assignment to the propositional letter occurring in q, which is assigned a truth value such that $d(M) \mid= q$.

Naturally enough, a preference ordering induced on models by a set of defaults D may at this point be defined. Given a set of defaults D and a theory T, the relation "$\leq+$" is defined to hold between models M and M' that both satisfy T (written $M \leq+ M'$) iff there exists d in D such that d is applicable to M and such that $d(M) = M'$. The relation "$\leq$" is defined as the transitive closure of "$\leq+$".[8]

Finally, we will say that a model M is *maximally preferred* (or *maximal*) with respect to a set of defaults D and a theory T iff for all models M' either $M' \leq M$ is the case or $M \leq M'$ is not the case.

We will illustrate the way $\mathcal{D}^+$ works by way of an example[9].

Example Let $P = \{a, b, c, d\}$, $D = \{a \rightarrow b, b \rightarrow c, ab \rightarrow \neg c, a\neg b \rightarrow \neg c, a\neg c \rightarrow \neg d\}$, $T = \{d\}$. **¬abcd**, **¬a¬bcd**, **ab¬cd** and **¬a¬b¬cd** are all and the only maximal models and all and the only intended models[10]. Note that if $b \rightarrow c$ had not been blocked at **ab¬cd**, then **abcd** would have been maximal too, contrary to intuitions. The example is represented graphically in Figure 1.

write e.g. $ab \rightarrow \neg c$ instead of $\{a, b\} \rightarrow \neg c$.

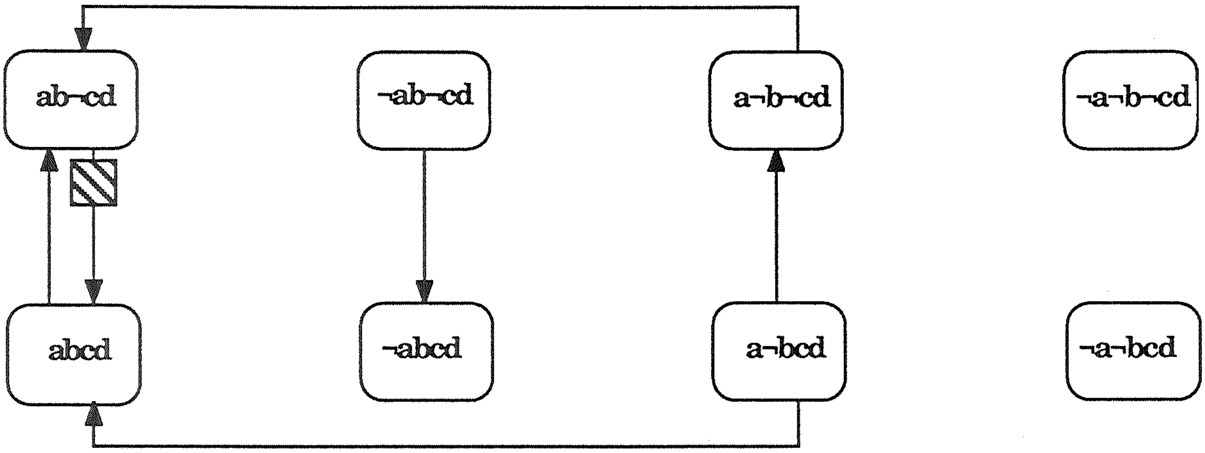
[7] In this paper, unlike in [MPD], we will not deal with theories including connectives other than negation; hence, a theory T will just be a set of literals which do not contain both a propositional letter p and its negation. Besides being less in keeping with the philosophy of vivid KBs, the commitment of [MPD] to arbitrary theories brings about some unintuitive behaviour, as first noted in (Boddy et al. 1989).

[8] [MPD] defines "$\leq$" to be the *reflexive* transitive closure of "$\leq+$"; that this is redundant may be seen by inspecting the way "$\leq$" is used in the definition of maximal model. Also, the requirement that both models satisfy T is the key modification introduced in (Sebastiani 1989) and allowing a correct interaction between certain and defeasible knowledge.

[9] In the drawings of the following examples, rectangles will denote models represented the obvious way (e.g. **a¬bcd** will represent the function that assigns True to a, c and d and False to b). Arrows will represent "$\leq+$" relationships. Slashed arrows will represent what would have been "$\leq+$" relationships unless a blocking had occurred. Also, we will omit drawing arrows corresponding to simple loops (i.e. arrows starting and ending in the same model) as they do not contribute in supporting the maximality or non- of a model. Models that do not satisfy T will also be omitted for similar reasons.

[10] We will call an "intended model" a model which our intuition suggests should be a maximal model. A model-preference default system will thus be empirically satisfactory iff for every set of defaults D every intended model is also a maximal model and viceversa.

Figure 1



3 A proof system for model-preference default reasoning

Before introducing $\mathcal{L}(\mathcal{D}^+)$ in formal detail we will sketch the basic ideas that underlie it in order to give the reader a feeling for what we are after.

As mentioned above, $\mathcal{L}(\mathcal{D}^+)$ will not be concerned with proving the theoremhood of a formula. Let us recall that in more traditional proof systems the notion of theoremhood of a formula may be seen as the proof-theoretic counterpart of the model-theoretic notion of validity, i.e. truth in all models. As we have seen above, there is no notion of validity in $\mathcal{D}^+$ (apart from the "trivial" one imported from classical logic, i.e. the one that does not take the import of defaults into account). The key semantic notion for us is that of a theory identifying a model which is maximal with respect to the given input data, i.e. to a theory T and a set of defaults D; hence, $\mathcal{L}(\mathcal{D}^+)$ will be concerned with proving the proof-theoretic counterpart of this notion, that, as such a theory will have a "minimal" number of models (namely, a single model)[11], we will call "minimality of a theory" wrt $<T, D>$.

According to our framework, the construction of a proof will roughly consist of the synthesis of a sequence of progressively larger theories $T \equiv T_1, T_2, ..., T_{n-k}$, where n is the cardinality of the alphabet, k is the cardinality of T, and each T_{i+1} is obtained through the application of an inference rule that adds to T_i a literal that is guaranteed not to contain propositional letters that already appear in T_i. The proof ends when the cardinality of the most recently obtained theory equals n; this theory will contain n literals made out of the n propositional letters of the alphabet, and will hence be a vivid KB[12]. We will call a theory obtained in this way a *minimal theory*.

[11] Strictly speaking, a theory can also have zero models (iff it is inconsistent); however, we will disregard this case here.

[12] To be consistent with the definition given in (Levesque 1988), it is the set of positive literals

The literal that is being added to T_i to yield a theory T_{i+1} will usually be the consequent of a default $d \equiv \alpha \rightarrow q$ in D that is relevant to the most recently obtained theory T_i; here "relevant" means that its antecedent α is a subset of T_i and is not a subset of the antecedent of any other default in D. Alternatively, this literal may be the result of a nondeterministic expansion (a "fleshing out" operation) of T_i, which must take place whenever no more defaults are relevant to T_i.

After having reached a general feel for what $\mathcal{L}(\mathcal{D}^+)$ is like, we are ready to delve into its formal specification. We will start by defining what a "step" in a proof is.

Let us recall that in more traditional proof systems a step is, roughly, either an axiom or a formula that follows from the application of an inference rule to one or more of the preceding steps. Instead, given the global character of default reasoning, $\mathcal{L}(\mathcal{D}^+)$ steps will have to be *global* descriptions of the state of advancement of the proving process. Therefore, at the very least, a step will have to contain a description of what the most recently obtained theory is; in particular, we will see that the last step of a proof will consist exactly of a (minimal) theory. However, as we have previously anticipated, the set of actions that can possibly be taken at a given stage in the proof depends on the existence of defaults that are relevant to the most recently obtained theory. Hence, for better convenience, an $\mathcal{L}(\mathcal{D}^+)$ step will also typically contain a description of what the defaults that might possibly be relevant to the most recently obtained theory are, and of what the ones that might become relevant in the future are.[13]

Definition A *step* of an $\mathcal{L}(\mathcal{D}^+)$ proof is either a theory T, or a pair $<T, D>$ where T is a theory and D is a set of defaults, or a triple $<T, D, X>$ where T is a theory and D, X are sets of defaults. ■

In the definition above, D is the set of "active" defaults, i.e. the ones that are potentially relevant to the most recently obtained theory, while X is the set of "idle" defaults, i.e. the ones that, although not relevant to the most recently obtained theory, might turn out to be relevant at future stages of the proving process. Note that, unlike in classical proof theories where steps are always formulae of the language, here steps do not all belong to the same syntactic type. Also, notice that although n-k theories are constructed during a proof, this does not mean that the number of steps in a proof is n-k because, as we will see below, only two inference rules out of six build a new theory; in fact, the number of steps in a proof is always larger than n-k.

belonging to $T_{n\text{-}k}$ that can be called a vivid KB; for the purposes of this paper we will be able to overlook the distinction.

[13] The recording of relevant defaults might not prove strictly necessary; consequently, its elimination might allow to cut down the number of rules of the proof system from the current six to four; however, the conditions on the applicability of the remaining rules would then be much more complicated, which is the reason why for the moment being we will stick to the formulation described in this paper.

We are now ready to define what a proof and what a minimal theory are.

Definition A *proof* in $\mathcal{L}(\mathcal{D}^{+})$ is a sequence of m steps ($m \geq 3$) such that: 1) the first step is a pair $<T, D>$ where T is a theory and D is a set of defaults ; 2) the m-th step is a theory T_{n-k}; 3) for all $i=2, \ldots, m$, the i-th step is the result of the application of one of the inference rules to the j-th step, for some $j=1, \ldots, i-1$; 4) n is the cardinality of the alphabet and k is the cardinality of T. ■

Definition A *minimal theory* wrt $<T, D>$, where T is a theory and D is a set of defaults, is the last step of a proof in $\mathcal{L}(\mathcal{D}^{+})$ whose first step is $<T, D>$. ■

We may now proceed to describe, one by one, the rules of inference that will be the constituents of $\mathcal{L}(\mathcal{D}^{+})$. In this section we will use the following abbreviations. We will feel free to write $\neg q_i$ and actually mean p_i in case $q_i \equiv \neg p_i$, and $\neg p_i$ in case $q_i \equiv p_i$. Also, we will say that two consistent theories T_j and T_k *are incompatible* when $q_i \in T_j$ and $\neg q_i \in T_k$ for some $i=1, \ldots, n$ (i.e. when their union is inconsistent). We will also use the abbreviations T^m, D^m and X^m to indicate the T, D and X components of step m, respectively.

The first rule we encounter is **S**, which is actually the first rule to be applied in a proof.

S

$$\frac{<T, D>}{<T, D, \{\}>}$$

S ("Start") creates an empty set X of "idle rules"; at each step m the set X^m of idle rules will consist of rules that are not relevant to T^m (because their antecedent is not a subset of T^m) but that are also not "ruled out" by it (i.e. the negation of their consequent is not in T^m), so that further additions of literals to T^m might render them applicable at a future stage of the proof. Note that, since by definition the first step of a proof must be a pair $<T, D>$ and since, as it will soon be clear, **S** is actually the only rule that can be applied to such a pair, **S** is always the first rule to be applied in a proof. Note also that neither **S** nor any other rule return a pair $<T, D>$: this means that **S** is never applied other than to the first step.

QED

$$\frac{<T, D, \{\}>}{T}$$

where the cardinality of T is n

QED ("*Quod Erat Demonstrandum*") is applied to a step m such that the cardinality of T^m is equal to the cardinality of the alphabet: semantically speaking, this means that the extension of T^m is a set consisting of a single model, and that the proof is virtually completed: we only need to get rid of D^m and X^m in order to comply with requirement 2 of the definition of "proof". Note that, since the last step of a proof must be a theory, and since **QED** is the only rule that returns solely a theory T, the last rule to be applied in a proof is always **QED**. Conversely, since there will be no rules with solely a theory T as antecedent, **QED** is only applied as the last rule of a proof.

The next and last four rules we encounter are applied to a triple $<T^m, D^m, X^m>$ to yield a triple $<T^{m+1}, D^{m+1}, X^{m+1}>$. Given the properties of rules **S** and **QED**, all rule applications apart from the first and the last in a proof will involve one of these four rules.

R

$$\frac{<T, \{d_1, ..., d_i, ..., d_p\}, X>}{<T, \{d_1, ..., d_{i-1}, d_{i+1}, ..., d_p\}, X>}$$

where $d_i \equiv \alpha \rightarrow q$
and either α and T are incompatible
or $\{q\}$ and T are incompatible
or $q \in T$

R ("Remove") removes permanently from D^m a default that could never be used in the proof, either because its antecedent could never be a subset of T^{m+l} for any $l \geq 0$ (this is the case when α and T^m are incompatible), or because its consequent could never be added to T^{m+l} for any $l \geq 0$ (this is the case when $\{q\}$ and T^m are incompatible), or because its consequent will always be an element of T^{m+l} for all $l \geq 0$(this is the case when $q \in T^m$). Notice that application of this rule is necessary to empty D^m and consequently let the **F** rule (see below) be triggered when appropriate.

I

$$\frac{<T, \{d_1, ..., d_i, ..., d_n\}, \{x_1, ..., x_n\}>}{<T, \{d_1, ..., d_{i-1}, d_{i+1}, ..., d_n\}, \{x_1, ..., x_n, d_i\}>}$$

where $d_i \equiv \alpha \rightarrow q$
and α is not a subset of T
and α and T are not incompatible
and $\{q\}$ and T are not incompatible
and it is not the case that $q \in T$

I ("Idle") moves a default which is not relevant to T^m (i.e. α is not a subset of T^m), but might turn out to be relevant at some future stage of the proof, from the set D^m of active defaults into the set X^m of idle defaults.

A

$$\frac{<T, D \equiv \{d_1, ..., d_i, ..., d_n\}, X>}{<T \cup \{q\}, \{d_1, ..., d_{i-1}, d_{i+1}, ..., d_n\} \cup X, \{\}>}$$

where $d_i \equiv \alpha \rightarrow q$
and α is a subset of T
and $\{q\}$ and T are not incompatible
and there is no proper superset β of α such that $\beta \rightarrow \neg q \in D$

A ("Apply") applies a default d_i which is most specific for T^m, i.e. it adds its consequent to T^m, removes d_i from D^m (because it would no more be useful) and moves all defaults from the set of idle defaults X^m into the set of active defaults D^m (because their antecedents might be subsets of $T^m \cup \{q\}$).

F

$$\frac{<T, \{\}, X>}{<T \cup \{q_i\}, \bigcup \{q_{\pi(1)}, ..., q_{\pi(k-1)}, q_i, q_{\pi(k+1)}, ..., q_{\pi(m_j)} \rightarrow \neg q_{\pi(k)}\}, \{\}>}$$

for all defaults $d_j \equiv q_{\pi(1)}, ..., q_{\pi(m_j)} \rightarrow \neg q_i \in X$ and for all $k = 1, ..., m_j$
where neither p_i nor $\neg p_i$ belong to T

F ("Flesh out") nondeterministically adds either p_i or $\neg p_i$ to T^m if T^m does not already contain either p_i or $\neg p_i$, and moves all defaults from the set of idle defaults X^m into the set of active defaults D^m. However, if q_i is added to T^m, for any default d_j of form $q_{\pi(1)}, ..., q_{\pi(m_j)} \rightarrow \neg q_i$ (where π is any permutation over the alphabet) which originally belonged to X^m, **F** adds to D^m all *contrapositives* of d_j, that is, all defaults d_j of form $q_{\pi(1)}, ..., q_{\pi(k-1)}, q_i, q_{\pi(k+1)}, ..., q_{\pi(m_j)} \rightarrow \neg q_{\pi(k)}$ for all $k = 1, ..., m_j$; this is done in order to make sure that further **F** applications will be consistent with the defaults that have not yet been discarded.

Having terminated the description of the inference rules, our job is completed by the following definition.

Definition We define the logic $\mathcal{L}(\mathcal{D}^+)$ of Model Preference Defaults as the proof system which is composed of the six rules **S**, **QED**, **R**, **I**, **A**, **F**. ■

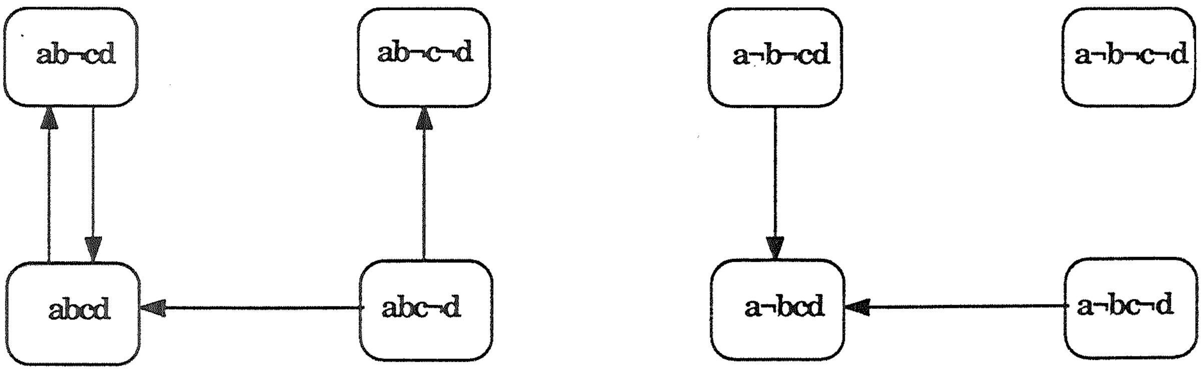
Notice that $\mathcal{L}(\mathcal{D}^+)$ does not have anything corresponding to what, in more traditional proof systems, are logical axioms. In fact, if we followed the parallel between the theoremhood of a formula and minimality of a theory wrt $<D, T>$, the notion corresponding to valid formulae would be that of a theory which is always minimal, irrespective of what D and T are. Quite obviously, there is not such a theory: as in model-preference default reasoning we are interested in finding *one* of the possibly many theories that are minimal wrt $<D, T>$, if there existed one theory which were always minimal this kind of reasoning would be a non-problem.

4 An example

In this section we will illustrate the way $\mathcal{L}(\mathcal{D}^+)$ works on a specific example, against which the reader will be able to check the actual behaviour of the rules described above. We will start by describing an instance problem $<T, D>$ and by finding out, by means of the usual graph-based method originally proposed in [MPD], all its maximal models. We will subsequently go on to apply $\mathcal{L}(\mathcal{D}^+)$ to the same instance problem $<T, D>$ in order to check the equivalence of the two problem-solving methods. In order to gain a better understanding of the behaviour of $\mathcal{L}(\mathcal{D}^+)$, this time we will not be content with finding a single minimal theory, but will go on to find *all* minimal theories of $<T, D>$. It will in fact turn out that these theories are all and the only theories denoting the maximal models found with the graph theoretic method.

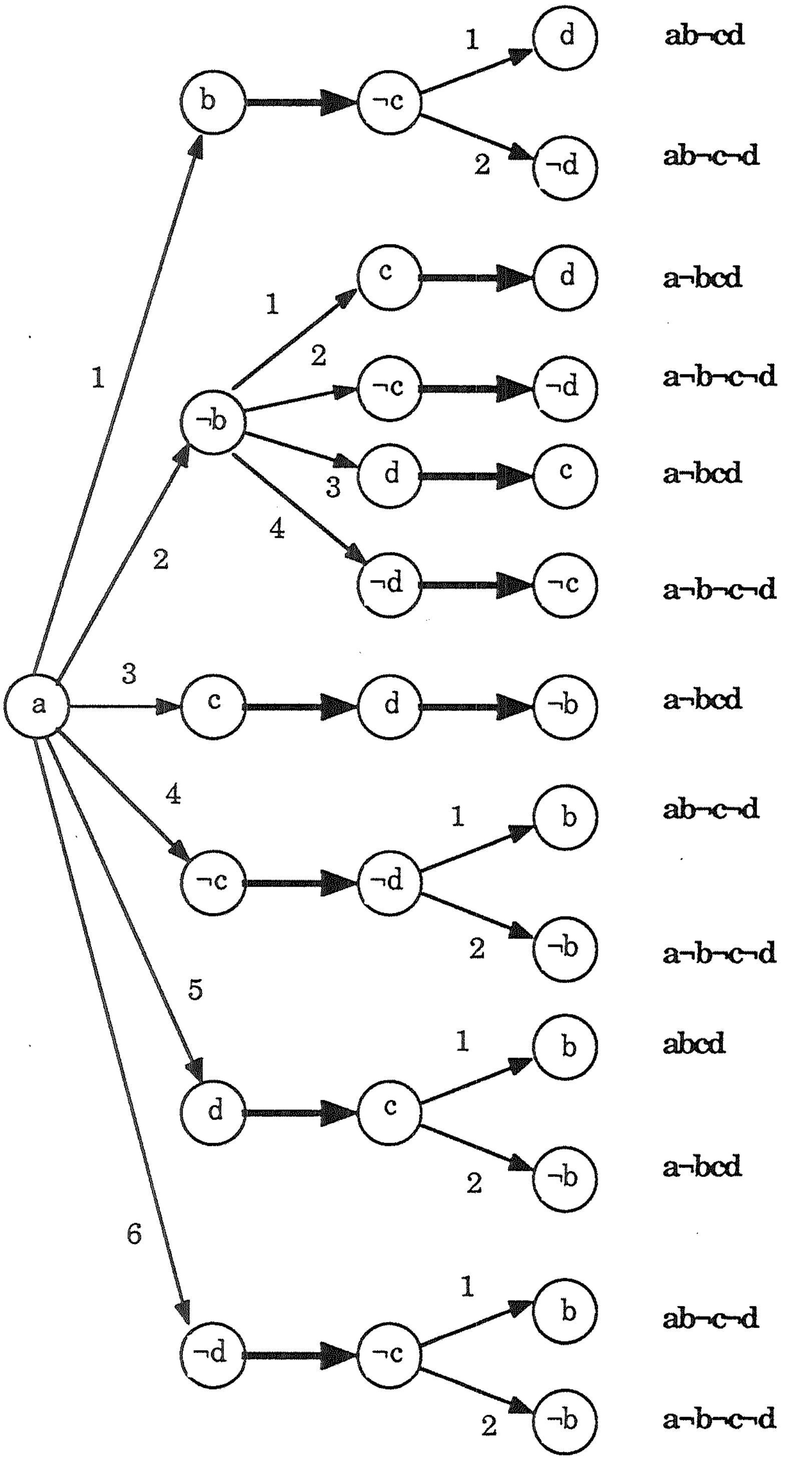
Example Let $P = \{a, b, c, d\}$, $D = \{d \to c, b \to \neg c, d \to \neg a, c \to d, \neg a \to d\}$, $T = \{a\}$. This is a typical example of an *inheritance hierarchy*, i.e. a set of defaults whose preconditions are sets with a single element; in recent times these hierarchies have been widely investigated also from a formal point of view (see e.g. Touretzky *et al.* 1987; Selman & Levesque 1989). The example is represented graphically in Figure 2.

Figure 2



Here **ab¬cd, ab¬c¬d, a¬bcd, a¬b¬c¬d,** and **abcd** are all and the only maximal models (and are also all and the only intended models). We will be interested in showing that all and only the corresponding theories are minimal according to $\mathcal{L}(\mathcal{D}^+)$. The need to find all minimal theories will compel us to explore the proof tree exhaustively. For convenience of presentation, in Figure 3 we detail the proof tree.

Figure3



In the tree-based representation of all possible proofs with respect to <T, D> each path represents a proof, and the nodes it traverses represent literals added to the theory T during the process. When a multiplicity of simple arrows depart from a node, they indicate that at that stage in the proof an **F** ("Flesh-out") rule had to be used; the multiple paths departing from the node represent the various nondeterministic expansions that could be performed at that stage. Solid arrows indicate that an **A** ("Apply") rule had to be (deterministically) applied.

The proofs so summarized in the proof tree are detailed in the following table. Each row represents a step in a proof, and is identified by a number (x/y) in the first column; this is to be interpreted as saying that this is the x-th step of all proofs corresponding to paths in the proof tree whose identifier starts with y. For example, step 12/4 is the 12-th step of the proofs identified by paths 41 and 42. In the i-th step the following columns represent the T^i, D^i and X^i components, while the last column gives information as to which rule is applied to which piece of information.

#	*T*	*D*	*X*	Rule applied
1/-	{a}	{d→c, b→¬c, d→¬a, c→d, ¬a→d}		
2/-	{a}	{d→c, b→¬c, d→¬a, c→d, ¬a→d}	--	S
3/-	{a}	{b→¬c, d→¬a, c→d, ¬a→d}	{d→c}	I (d→c)
4/-	{a}	{d→¬a, c→d, ¬a→d}	{d→c, b→¬c}	I (b→¬c)
5/-	{a}	{c→d, ¬a→d}	{d→c, b→¬c}	R (d→¬a)
6/-	{a}	{¬a→d}	{d→c, b→¬c, c→d}	I (c→d)
7/-	{a}	{}	{d→c, b→¬c, c→d}	R (¬a→d)
8/1	{a, b}	{d→c, b→¬c, c→d}	{}	F (b)
9/1	{a, b, ¬c}	{d→c, c→d}	{}	A (b→¬c)
10/1	{a, b, ¬c}	{c→d}	{}	R (d→c)
11/1	{a, b, ¬c}	{}	{}	R (c→d)
12/11	{a, b, ¬c, d}	{}	{}	F (d)
13/11	{a, b, ¬c, d}	--	--	QED
12/12	{a, b, ¬c, ¬d}	{}	{}	F (¬d)
13/12	{a, b, ¬c, ¬d}	--	--	QED
8/2	{a, ¬b}	{b→¬c, d→c, c→d}	{}	F (¬b)
9/2	{a, ¬b}	{d→c, c→d}	{}	R (b→¬c)
10/2	{a, ¬b}	{c→d}	{d→c}	I (d→c)
11/2	{a, ¬b}	{}	{d→c, c→d}	I (c→d)

Notice that a sequence of interleaved applications of **R** and **I** (e.g. steps 3/- to 7/-) produces a result that is independent of the particular order in which these rules are applied, because the result of each application does not affect the applicability of the others. Without loss of generality we will then be able to consider a proof as a representative of a whole class of different proofs obtained by shuffling the order of applications of **R** and **I**.

12/21	{a, ¬b, c}	{d→c, c→d}	{}	F (c)
13/21	{a, ¬b, c}	{c→d}	{}	R (d→c)
14/21	{a, ¬b, c, d}	{}	{}	A (c→d)
15/21	{a, ¬b, c, d}	--	--	QED

12/22	{a, ¬b, ¬c}	{d→c, c→d, ¬c→¬d}	{}	F (¬c)
13/22	{a, ¬b, ¬c}	{c→d, ¬c→¬d}	{}	R (d→c)
14/22	{a, ¬b, ¬c}	{¬c→¬d}	{}	R (c→d)
15/22	{a, ¬b, ¬c, ¬d}	{}	{}	A (¬c→¬d)
16/22	{a, ¬b, ¬c, ¬d}	--	--	QED

Step 12/22 features an interesting case of full-fledged application of **F**, with a default (namely, $\neg c \rightarrow \neg d$) which did not originally belong to D being added to it in order to ensure the coherency with D of further applications of **F**.

12/23	{a, ¬b, d}	{d→c, c→d}	{}	F (d)
13/23	{a, ¬b, c, d}	{c→d}	{}	A (d→c)
14/23	{a, ¬b, c, d}	--	--	QED

Notice that the minimality of $\{a, \neg b, c, d\}$ had been already proven before along path 21 (and will be proven again by means of paths 3 and 52).

12/24	{a, ¬b, ¬d}	{d→c, c→d, ¬d→¬c}	{}	F (¬d)
13/24	{a, ¬b, ¬d}	{c→d, ¬d→¬c}	{}	R (d→c)
14/24	{a, ¬b, ¬d}	{¬d→¬c}	{}	R (c→d)
15/24	{a, ¬b, ¬c, ¬d}	{}	{}	A (¬d→¬c)
15/24	{a, ¬b, ¬c, ¬d}	--	--	QED

8/3	{a, c}	{b→¬c, d→c, c→d, c→¬b}	{}	F (c)
9/3	{a, c}	{d→c, c→d, c→¬b}	{}	R (b→¬c)
10/3	{a, c}	{c→d, c→¬b}	{}	R (d→c)
11/3	{a, c, d}	{c→¬b}	{}	A (c→d)
12/3	{a, ¬b, c, d}	{}	{}	A (c→¬b)
13/3	{a, ¬b, c, d}	--	--	QED

8/4	{a, ¬c}	{b→¬c, d→c, c→d, ¬c → ¬d}	{}	F (¬c)
9/4	{a, ¬c}	{d→c, c→d, ¬c → ¬d}	{}	R (b→¬c)
10/4	{a, ¬c}	{c→d, ¬c → ¬d}	{}	R (d→c)
11/4	{a, ¬c}	{¬c → ¬d}	{}	R (c→d)
12/4	{a, ¬c, ¬d}	{}	{}	A (¬c → ¬d)
13/41	{a, b, ¬c, ¬d}	{}	{}	F (b)
14/41	{a, b, ¬c, ¬d}	--	--	QED

13/42	{a, ¬b, ¬c, ¬d}	{}	{}	F (¬b)
14/42	{a, ¬b, ¬c, ¬d}	--	--	QED

8/5	{a, d}	{b→¬c, d→c, c→d}	{}	F (d)
9/5	{a, d}	{d→c, c→d}	{b→¬c}	I (b→¬c)
10/5	{a, c, d}	{c→d, b→¬c}	{}	A (d→c)
11/5	{a, c, d}	{b→¬c}	{}	R (c→d)
12/5	{a, c, d}	{}	{}	R (b→¬c)
13/51	{a, b, c, d}	{}	{}	F (b)
14/51	{a, b, c, d}	--	--	QED

Notice how the order in which subsequent applications of **F** are carried out influences the theories which can be shown minimal in a given subtree: $\{a, b, c, d\}$ has not turned out to be minimal on paths where the first application of **F** adds b or c to T (although b or c do belong to $\{a, b, c, d\}$) while it has been shown to be minimal by using d as the first addition.

13/52	{a, ¬b, c, d}	{}	{}	F (¬b)
14/52	{a, ¬b, c, d}	--	--	QED
8/6	{a, ¬d}	{b→¬c, d→c, c→d, ¬d→¬c}	{}	F (¬d)
9/6	{a, ¬d}	{d→c, c→d, ¬d→¬c}	{b→¬c}	P (b→¬c)
10/6	{a, ¬d}	{c→d, ¬d→¬c}	{b→¬c}	R (d→c)
11/6	{a, ¬d}	{¬d→¬c}	{b→¬c}	R (c→d)
12/6	{a, ¬c, ¬d}	{b→¬c}	{}	A (¬d→¬c)
13/6	{a, ¬c, ¬d}	{}	{}	R (b→¬c)
14/61	{a, b, ¬c, ¬d}	{}	{}	F (b)
15/61	{a, b, ¬c, ¬d}	--	--	QED
14/62	{a, ¬b, ¬c, ¬d}	{}	{}	F (¬b)
15/62	{a, ¬b, ¬c, ¬d}	--	--	QED

Notice how all the theories that we have shown to be minimal correspond to models that had been shown maximal with the graph-based method, and how all such models correspond to theories that we have shown minimal.

5 Conclusion

In this paper we have described a formalism for reasoning with default information that attempts to provide a proof-theoretic alternative to the graph-theoretic reasoning style that was originally proposed for model-preference reasoning. Although we do not claim that such proof theory should replace *tout court* the original graph-based algorithms, we think it brings about some substantial insights into how model-preference default reasoning can be accomplished by means of proof-theoretic, and hence more orthodox, tools; it is precisely because of its greater orthodoxy that this approach is especially promising with respect to the prospective integration of default reasoning and other reasoning patterns.

In order for $\mathcal{L}(\mathcal{D}^+)$ to lay claim of being semantically motivated, we should show its soundness and completeness with respect to the semantics sketched in Section 2. Soundness would consist of all minimal theories wrt $<T, D>$ identifying models that are maximal wrt $<T, D>$, while completeness would consist of all models that are maximal wrt $<T, D>$ being denoted by theories that are minimal wrt $<T, D>$. While we are still investigating the issues of soundness and completeness of $\mathcal{L}(\mathcal{D}^+)$ wrt $\mathcal{D}^+$, because of the behaviour that we have observed up to date in testing $\mathcal{L}(\mathcal{D}^+)$ on several examples we conjecture that such properties indeed obtain.

Acknowledgements

This research was carried out in the framework of the TELOS project, dealing with the design and construction of a knowledge representation language supporting the phase of requirements modelling in information systems development. I would like to thank John Mylopoulos for encouraging the work described here and for providing me the environment where this research was carried out. I would also like to thank David Lauzon and Bart Selman for stimulating discussions about the topic of this paper.

Bibliography

Boddy, Mark; Goldman, Robert P.; Kanazawa, Keiji & Stein, Lynn A. (1989). Investigations of model-preference defaults. Technical Report CS-89-13, Department of Computer Science, Brown University, Providence, RI.

Borgida, Alex & Etherington, David W. (1989). Hierarchical knowledge bases and efficient disjunctive reasoning. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Ontario, pp. 33-43.

Brown, Allen L. & Shoham, Yoav (1989). New results on semantical nonmonotonic reasoning. In Reinfrank, Michael; De Kleer, Johan; Ginsberg, Matthew L. & Sandewall, Erik (eds.) (1989), *Nonmonotonic reasoning*, Heidelberg, BRD: Springer, pp. 19-26.

Etherington, David W.; Borgida, Alex; Brachman, Ronald J. & Kautz, Henry A. (1989). Vivid knowledge and tractable reasoning: preliminary report. In *Proceedings of IJCAI-89*, Detroit, MI, pp. 1146, 1152.

Johnson-Laird, Philip N. (1983). *Mental models*. Cambridge, MA: Harvard University Press, 1983.

Konolige, Kurt (1987). On the relation between default theories and autoepistemic logic. In *Proceedings of IJCAI-87*, Milan, Italy, pp. 394-401. [1] An extended version appears as "On the relation between default logic and autoepistemic theories" in *Artificial Intelligence 35*, pp. 343-382, 1987.

Levesque, Hector J. (1986). Making believers out of computers. *Artificial Intelligence 30*, pp. 81-108.

Levesque, Hector J. (1988). Logic and the complexity of reasoning. *Journal of Philosophical Logic 17*, pp. 355-389.

McCarthy, John (1980). Circumscription - A form of nonmonotonic reasoning. *Artificial Intelligence 13*, pp. 81-108. [1] Appears also in Ginsberg, Matthew L. (ed.) (1987), *Readings in nonmonotonic reasoning*, Los Altos, CA: Morgan Kaufmann, pp. 145-152.

McCarthy, John (1980). Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence 28*, pp. 89-116. [1] Appears also in Ginsberg, Matthew L. (ed.) (1987), *Readings in nonmonotonic reasoning*, Los Altos, CA: Morgan Kaufmann, pp. 153-166.

Moore, Robert C. (1985). Semantical considerations on nonmonotonic logic. *Artificial Intelligence 25*, pp. 75-94. [1] Appears also in Ginsberg, Matthew L. (ed.) (1987), *Readings in nonmonotonic reasoning*, Los Altos, CA: Morgan Kaufmann, pp. 127-136.

Sebastiani, Fabrizio (1989). On heterogeneous model-preference default theories. Technical Report IEI-B4-72-1989, Istituto di Elaborazione dell'Informazione - CNR, Pisa, Italy.

Sebastiani, Fabrizio (1990). A fully denotational semantics for model-preference default systems. Technical Report IEI-B4-06-1990, Istituto di Elaborazione dell'Informazione - CNR, Pisa, Italy.

Selman, Bart & Kautz, Henry A. (1988). The complexity of model-preference default theories. In *Proceedings of the 1988 Conference of the Canadian Society for Computational Studies of Intelligence*, Edmonton, Alberta, pp. 102-109. [1] Appears also in Reinfrank, Michael; De Kleer, Johan; Ginsberg, Matthew L. & Sandewall, Erik (eds.) (1989), *Nonmonotonic reasoning*, Heidelberg, BRD: Springer, pp. 115-130. [2] An extended version is forthcoming as "Model-preference default theories" in *Artificial Intelligence.*

Selman, Bart & Levesque, Hector J. (1989). The tractability of path-based inheritance. In *Proceedings of IJCAI-89*, Detroit, MI, pp. 1140-1145. [1] A preliminary version is forthcoming in *Formal aspects of semantic networks*, Los Altos, CA: Morgan Kaufmann, CA, 1990. [2] An extended version is forthcoming in Lenzerini, Maurizio; Nardi, Daniele & Simi, Maria (eds.), *Inheritance hierarchies in knowledge representation*, New York, NY: Wiley, 1990.

Shoham, Yoav (1987). A semantical approach to nonmonotonic logics. In *Proceedings of IJCAI-87*, Milan, Italy, pp. 388-392. [1] Appears also in Ginsberg, Matthew L. (ed.) (1987), *Readings in nonmonotonic reasoning*, Los Altos, CA: Morgan Kaufmann, pp. 227-250.

Touretzky, David S.; Horty, John F. & Thomason, Richmond H. (1987). A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings of IJCAI-87*, Milan, Italy, pp. 476-482.