RC 20125 (07/07/95)
Computer Science

IBM Research Report

Defeasible and Pointwise Prioritization: Preliminary Report

Benjamin N. Grosof

IBM Research Division
T.J. Watson Research Center
P.O. Box 704, Yorktown Heights, NY 10598
(914) 784-7783 direct -7455 fax -7100 main
Internet: grosof@watson.ibm.com (alt.: grosof@cs.stanford.edu)
World Wide Web: http://www.watson.ibm.com

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

This paper was distributed *as a companion to [Grosof, 1993] at the Commonsense '93 Symposium.*

April 20, 1992
revised January 9, 1993
reformatted with minor corrections June 30, 1995

Abstract

We suggest why, and show how, to represent *defeasible* reasoning about prioritization-type precedence. We define Defeasible Axiomatized Policy (DAP) circumscription: it is the first formalism to express defeasible prioritization. DAP circumscription can represent one or more (generally, a finite *reflective tower*) of meta-levels of such reasoning, *without* resorting to a more powerful logical language. We argue for the usefulness, and analyze the expressive significance, of this representational generalization. We show that it can often be achieved with only a modest increase in the mathematical complexity of inference: DAP circumscription often reduces to a series of prioritized predicate circumscriptions, for which inference procedures are currently available.

DAP circumscription also offers an improved approach to *pointwise* prioritization and circumscription, even in the basic, monotonic case of reasoning about prioritization. We observe that unsatisfiability and representational awkwardness trouble the previous approach, due to Lifschitz [1988]. DAP circumscription overcomes these difficulties.

Note to Reader: This is a preliminary report. We define our new DAP circumscription formalism in full detail. However, we only sketch the proofs of several theorems that characterize its behavior. In a future, expanded version of this paper, we will give these proofs in full, and elaborate more in general. With several more illustrating examples, we will support our argument that the DAP formalism behaves well in terms of its intuitive behavior (conclusions entailed) and its representational conciseness.

The LaTeX sans serif style, often with [square brackets], indicates additional material that will be found **in a future, expanded version of this paper**.

Contents

1 Problem Addressed

The representation of beliefs about beliefs has long been a central issue in AI. Generally, one would like to equip agents with the ability to express and reason reflectively and introspectively. Being able to formalize this within a common logical system offers obvious advantages over a scheme with promiscuous proliferation of meta-languages, especially if the cost in formal and/or computational complexity in doing so is not too high.

Here, we address this standard goal of KR for the special case of prioritized default reasoning in circumscription. In particular, our goal here is to represent reasoning about prioritization, and, more generally, about what formulas should be *adopted* as default axioms. In circumscription, the minimization of a predicate (e.g., an abnormality) can be viewed as corresponding to a default axiom, i.e., to the maximization of a default formula. (See Section 5 for more about this view.) In this paper, therefore: in circumscription, we will view selecting (specifying) a predicate to minimize as adopting a default.

There are several different aspects to such reasoning about prioritization and adoption that we would like to represent. The first aspect is *monotonic* reasoning: e.g., to be able to deduce, in ordinary classical-logical fashion, a prioritization on the basis of some other, more directly available information. The second aspect is *"pointwise"* reasoning, of the kind defined and explored by Lifschitz in his idea of pointwise circumscription [1987] [1988] and, essentially, also by Shoham in his idea of chronological minimization [1988]. The third, and most innovative aspect, is *non-monotonic* (*defeasible*) reasoning: e.g., to be able to retract and reverse, as well as infer, prioritization, on the basis of other information; i.e., to make default conclusions about the prioritization itself. The fourth aspect is a particular kind of *reflection*: we would like to be able to represent non-monotonic reasoning at one level *about* the prioritization and adoption used in non-monotonic reasoning at a second level, which is in turn *about* the prioritization and adoption used in non-monotonic reasoning at a third level, etc.. However, we would like to avoid a proliferation of meta-languages: we would like to represent such multi-level reasoning about prioritization and adoption within a single logical language and logical system. [We discuss the obvious, "cheating" alternative: to reason in multiple meta-languages.]

2 Previous Work

Lifschitz: Lifschitz [1988] showed how, in a variant of circumscription, to represent one level of monotonic and pointwise reasoning about prioritization and adoption, within a single logical language (and system). I.e., he shows how to do this *without* resorting to a more powerful logical language / system. His formalism is defined as a *conjunction* of circumscriptions, one per minimized predicate. Lifschitz' formal representational mechanism for specifying the non-monotonic activity, i.e., the *"policy"* of these circumscriptions is *variability* (V in his formalism): i.e., which of the other predicates are permitted to vary (as opposed to being fixed) in the context of each particular predicate's circumscriptive minimization. (Section 5 reviews the idea of fixing.) Lifschitz introduces this policy V as a predicate in the same logical language as the predicates being minimized. E.g., in the abnormality-style representation of the Quaker-Republican example of default reasoning, $ab1$, $ab2$, *Quaker*, *Republican*, $Pacifist$, are all first-order predicates in the base logical language, and V is also a first-order predicate in that same language. V is a binary relation whose arguments are the names of predicates, or, more generally, are points. For example,

$V('ab1, 'Pacifist)$

says that, in the conjunct circumscription in which the predicate $ab1$ is minimized, the predicate (symbol) $Pacifist$ is varied.

V being in the same language enables monotonic, first-order-quantified reasoning about the policy: e.g., sentences about V may be deduced by deduction from axioms that relate V to the other predicates. The first-order-quantifiability over V's arguments allows "pointwise control" of the policy. [We give more explanation of Lifschitz' formalism.]

Defaults About Defaults: Previous approaches to representing (a single agent's) default beliefs about default beliefs include non-monotonic (NM) modal logics, such as Autoepistemic Logic (AEL) [Moore, 1985], its predecessor [McDermott and Doyle, 1980], and its offshoots, e.g., [Konolige, 1988b] and [Levesque, 1990]; and NM conditional logics, such as [Delgrande, 1987]. AEL is highly oriented towards the issues of introspection, e.g., nesting of belief, as traditionally formulated in modal logics of belief and knowledge. In addition, some recent work has addressed issues of multiple-agent non-monotonic reasoning (NMR) using AEL [Morgenstern, 1990] [Morgenstern and Guerriro, 1991]. Related is the work on NMR about speech acts: [Appelt and Konolige, 1988] using Hierarchical Autoepistemic Logic (HAEL) [Konolige, 1988b]; and [Perrault, 1987] using Default Logic [Reiter, 1980].

Explicit Prioritization-type Precedence: Interestingly, none of these previous systems include the ability to express prioritization-type precedence explicitly (as opposed to implicitly via specificity dominance). This ability, however, is useful for many applications of non-monotonic reasoning (NMR) (see section 4). HAEL comes the closest of any of these to providing this expressive capability. However, its concept of precedence is quite different, and, we believe, much less useful than prioritization. [1]

3 Overview

Circumscription has the advantage of being equipped with explicit prioritization. But it has no obvious analogue of defeasible belief about defeasible belief. We show how to overcome this limitation.

Here, we define a new formalism: Defeasible Axiomatized Policy (DAP) circumscription. DAP circumscription is a different kind of variant of circumscription than Lifschitz' scheme: it is a *single* general circumscription [Lifschitz, 1984], rather than a conjunction of circumscriptions (see section 9 for more about this difference).

Like Lifschitz' formalism, DAP circumscription is able to express monotonic and pointwise reasoning about prioritization and adoption. However, it has several advantages over his for the monotonic, pointwise case. DAP is better-behaved (e.g., with respect to satisfiability). And it is more adequate and more natural for representing prioritization, especially non-layered (non-stratified) prioritization.

DAP circumscription generalizes predicate circumscription and prioritized predicate circumscription as defined in [McCarthy, 1986] [Lifschitz, 1985] [Grosof, 1991] [Grosof, 1992].

Equipped with DAP circumscription, we show, for the first time, how to express *non-monotonic (defeasible)* reasoning about prioritization and adoption, in one overall circumscriptive theory, also *without* resorting to a more powerful logical language. Moreover, we show how to express two, three, or any finite number of meta-levels of such reasoning, still within the same language. I.e., we show how to express a (finite-height) *reflective* tower of prioritized default circumscriptive reasoning, each level's subject being the prioritization and adoption at the next-lower level. These beliefs and inferences are represented via ordinary first-order sentences; the overall circumscription is, as usual, a (classical) second-order formula.

DAP circumscription can express pre-orders (model-preference criteria) that are fundamentally more general than those employed in prioritized predicate (/ default) circumscription. In section 9, we analyze the expressive power of DAP circumscription, and the problem of inference in it. Via reduction

[1] HAEL cannot represent unresolved conflict between defaults. The Nixon diamond, where no precedence information is available, results in inconsistency of the NM theory.

theorems, we show that the non-monotonic and reflective aspects of reasoning about prioritization and adoption can always be achieved at only a modest increase in the mathematical complexity of inference over the monotonic case; the monotonic case, in turn, can often be achieved with only a modest increase over ordinary prioritized predicate circumscription.

Our Main Contributions in this paper are (1) to suggest why, and (2) to show how, to represent reflective, defeasible reasoning about prioritization-type precedence, not just about adoption. DAP circumscription is the first formalism to express defeasible prioritization. (HAEL, by contrast, does not permit reasoning about the precedence at all, much less defeasible reasoning about it.) In addition, (3) we identify some problems (satisfiability and representational adequacy) with the previous formalization of pointwise (prioritization and adoption in) circumscription due to Lifschitz, and show that DAP circumscription overcomes them.

Limitations: However, our approach does not (at least yet) capture some other issues of introspection that have provided important motivation for some previous approaches. E.g., DAP circumscription does not permit the direct syntactic representation of nested belief. In addition, we only permit a finite number of default axioms[2] to be the formulas about which reflective reasoning is conducted. For now, we also consider only the case of a single believing agent (rather than multiple agents).

Next, we discuss the usefulness of this expressive generalization for applications.

4 Motivating Applications

We are interested especially in developing a formal theory to support the design of what we call *learning agents*: automatic programs that store, and revise, a body of conclusions while continually acquiring new, possibly defeasible, premise beliefs.

Application areas for learning agents include inductively forming theories in: rule and concept learning [Russell and Grosof, 1987] [Grosof and Russell, 1990] [Russell and Grosof, 1990a] [Helft, 1989]; natural language understanding and communication [Perrault, 1987] [Appelt and Konolige, 1988] [McRoy, 1989] [Morgenstern, 1990] [Morgenstern and Guerriro, 1991] [Zadrozny, 1988]; diagnosis [Ginsberg, 1986] [Reiter, 1987]; and perception and sensory interpretation (vision, motion, speech understanding). Other applications include: reasoning about time and action; law; Bayesian and evidential reason-

[2] though, these may have an infinite number of instantiations

ing [Grosof, 1988] [Bacchus, 1990] [Bacchus, 1991]; and retractable decision-making (planning, design, heuristics).

In many applications of learning agents, updating with new defaults is frequent, and information about which defaults take precedence over others (have greater prioritization) is important and available. Bases for prioritization information [Grosof, 1991] include: specificity dominance as in inheritance, reliability of sources, authority in the legal and organization senses, temporal sequence, and decision-theoretic superiority (greater probabilistically expected utility) of actions based on defaults.

Reasoning about the relative precedence of one's beliefs and information occurs, for example, when updating perceptual models of the physical world (e.g., on the basis of freshness of information), or when updating scientific knowledge (e.g., on basis of quality of methodology and statistical significance). In particular, much of the information that humans acquire via communication (e.g., reading, listening) and that they need in order to deal with what is acquired, concerns credibility or reliability. (E.g., consider negotiating a bureaucracy.) Accordingly, we would like to represent reasoning about precedence, including reflective / introspective reasoning about precedence, as part of a NM formalism. And, to capture the defeasibility of almost all knowledge, we would like to represent the retractability of the premises themselves (not just of the conclusions they entail) as much as possible.

Prioritizing Multiple, Contradictory Sources of Advice in Common-Sense Learning by Being Told: An important common-sense capability is to learn by "taking advice", or "being told", from a variety of sources of information: messages from various other agents, reading various texts, etc.. Such learned or "assimilated" knowledge provides an important basis for one's actions. Yet a basic feature of life is that one cannot believe everything that one is told. Not only may advice be incorrect; e.g., it may be contradicted by direct experience. Worse, different sources of advice information may contradict each other, or even themselves. How is an advice-taking agent to maintain a consistent, yet usefully actionable, set of beliefs, then? In [Grosof, 1993], we address this problem, which has not received much previous attention in the AI literature. We illustrate, with an actual example, how it arises in the domain of *bureaucracy*. A common human strategy is to treat advice as working belief, which can be overridden by other advice and direct experience. This suggests the first element of our approach: to represent advice as default-status knowledge.

A crucial issue in logically formalizing common-sense reasoning is then how to represent *precedence* among multiple, potentially contradictory sources of information. We observe that bases for such precedence include reliability and authority. We observe furthermore that precedence relationships may themselves be inferrable and even defeasible. Using DAP circumscription, we show how, for the first time, to represent advice-taking, with such precedence among sources, as reasoning in a non-monotonic logical formalism. Precedence among sources is represented explicitly via the mechanism of prioritization in DAP circumscription.

Examples of Defeasible Prioritization: Next, we list some examples where it is useful to represent *defeasible* prioritization.

In an organization, change of authorities (e.g., promotion and demotion of managers) leads to change of prioritization by subordinates in following potentially-conflicting demands from superiors. Each command from a superior may be represented as a default premise belief: e.g., Manager Fred tells Joe to "plan to finish the report today" and Manager Sue tells Joe to "plan to fly to Chicago today". But these commands may conflict: e.g., it is physically impossible for the hapless Joe to both finish the report today and fly to Chicago today. Relative authority within the organization provides a basis for precedence between such commands. E.g., Joe has as a premise belief that Manager Sue is Manager Fred's boss. Joe thus concludes by default that Manager Sue's command takes precedence over Manager Fred. But if Joe receives a new bulletin from the chairman of the organization that Manager Fred has been promoted over Manager Sue's head, then Joe concludes that this precedence should be reversed. This reasoning about precedence is thus itself defeasible.

In law, rulings concern precedence of principles or of jurisdictions, and are themselves reversible by subsequent rulings.

In inductive learning, adoption and prioritization-type credibility change on the basis of changing history and statistics about past predictive accuracy.

[We describe these examples in more detail.]

Our approach can represent these reasoning patterns, in a fashion similar to the Fiscal Agenda example that we detail in section 7.

5 Prioritized Circumscription: Review

General Circumscription: [Lifschitz, 1984] defined general circumscription as minimization with respect to a pre-order *preference relation* (i.e., a preference criterion):

$$C(B;G;Z) \stackrel{\text{def}}{\equiv} B[Z] \wedge \neg\exists Z'.\, B[Z'] \wedge Z' \prec_G Z$$

Here Z is the tuple of all predicate symbols in a classical first-order *base lan-*

guage $\mathcal{L}$. [3] $B[Z]$ is the *base sentence*, representing a conjunction of one or more for-sure (non-defeasible) axioms. Except where explicitly mentioned, in this paper the axioms constituting B will be expressible in first-order. [4]

G is a binary relation that is reflexive and transitive, i.e., is a *pre-order*.

$\forall Z.\, G(Z, Z)$

$\forall Z^1, Z^2, Z^3.\quad G(Z^1, Z^2) \wedge G(Z^2, Z^3) \quad \supset \quad G(Z^1, Z^3)$

Each of G's arguments is taken from the same domain: candidate tuples of predicate symbols similar to Z. By "similar", we mean that each such tuple has the same arity as Z, and each predicate symbol in that tuple has the same arity as the corresponding predicate in Z.

We can view G as a preference criterion. For our purposes in this paper, G can also be viewed equivalently as defined over models.[5]

Fixing: In the circumscription literature, a symbol is said to be *fixed* when it is not varied and second-order quantified over in the right-hand-side part of the circumscription formula. For the sake of simplicity of presentation, in this paper, we *fix* (do not vary and second-order quantify over) all *function* symbols in the circumscriptions. This assumption can easily be relaxed. In addition, in defining prioritized predicate circumscription (as well as DAP circumscription), we neglect the ability to fix any *predicate* symbols, at least explicitly. This assumption, too, can easily be relaxed. The effect of fixing any particular sub-set of the predicates can be accomplished indirectly, but straightforwardly, via the method in [Grosof, 1992] (section 3.5), which generalizes that of [de Kleer and Konolige, 1989]. In short, such fixing is expressively inessential. See [Grosof, 1992] (especially, sections 2.3 and 3.5) for more discussion of fixing. [We give more explanation of fixing versus varying.]

Notation: For any pre-order G, we also write G as $\preceq_G$. We then write $\prec_G$ to stand for its irreflexive version $\preceq_G \wedge \neg\succeq_G$. We also write $\approx_G$ to stand for $\preceq_G \wedge \succeq_G$.

Prioritized Circumscription: [Grosof, 1991][6] defined prioritized circumscription as the special case of general circumscription where G is a *prioritized pre-order* $(H^N; R)$. Here H is a tuple, indexed by N, of *starting* pre-orders, all defined over the same domain as we discussed above. The *prioritization partial order* R is a well-founded (in the direction of ascending

[3] This can easily be generalized to higher-order classical languages.

[4] In this paper, we assume that there are only a finite number of predicates in $\mathcal{L}$. This will be the case if B is finitely axiomatizable and Z is defined as the set of symbols appearing explicitly in B.

[5] See [Etherington, 1988] (section 6.2) and [Grosof, 1992] (section 2.3) for discussion of model theory for general circumscription.

[6] See also [Grosof, 1992] for elaboration.

priority) strict partial order, defined over N. (In general, N may be infinite.) $R(j,i)$ means that the preference criterion Hj is to be minimized with higher precedence than the preference criterion Hi. The prioritization operation, then, takes as input a tuple of starting preference criteria and results in a single, aggregated preference criterion:

$$Z^1 \preceq_{(H^N;R)} Z^2 \stackrel{\text{def}}{\equiv}$$
$$\forall i \in N.\ [\forall j \in N.\ R(j,i) \supset (Z^1 \approx_{Hj} Z^2)] \ \supset \ (Z^1 \preceq_{Hi} Z^2)$$

In this paper, we will focus technically on minimizing predicates. We will define DAP circumscription as a generalization of prioritized predicate circumscription (PPC).

Prioritized Predicate Circumscription: [Grosof, 1991] defined prioritized *predicate* circumscription[7] as, in turn, the special case in which the minimized pre-order is a prioritized predicate pre-order, i.e., when each Hi is an individual *predicate* pre-order. For any predicate P in Z, its corresponding predicate pre-order, which we will denote in our notation simply with P, is defined as:

$$Z^1 \preceq_P Z^2 \stackrel{\text{def}}{\equiv} \ P^1 \leq P^2$$

Here we let $D \leq E$ stand for $\forall x.\ D(x) \supset E(x)$, where D and E are open formulas with the same arity, and x stands for a tuple of individual object variables. Let $D{=}E$ and $D < E$ then be defined analogously to $\approx_G$ and $\prec_G$ above. [8]

Thus prioritized predicate circumscription (PPC) is the minimization of a *preference relation* (corresponding to a model-preference criterion) that is a prioritized predicate pre-order.

Notation for PPC: Notationally, in PPC, we can use N to index the sub-set of predicates in Z that are minimized. Henceforth, we will write Z^N to stand for the tuple of minimized predicates, among the overall tuple Z of all predicates. Thus, we use N as a kind of indexing selector.

We thus write a prioritized predicate predicate circumscription as:

$$PPC(B; Z^N; R; Z)$$

or as: $PPC(B; N; R; Z)$

where the minimized pre-order is defined as:

$$Z^1 \preceq_{(Z^N;R)} Z^2 \stackrel{\text{def}}{\equiv}$$
$$\forall a \in N.\ [\forall b \in N.\ R(b,a) \supset (Zb^1 = Zb^2)] \ \supset \ (Za^1 \leq Za^2)$$

[7] This generalized [Lifschitz, 1985], who made precise [McCarthy, 1986]'s conceptual proposal of prioritization. See also [Grosof, 1992] for further elaboration.

[8] For simplicity, in this version of this paper, we notationally often ignore what arity a particular predicate has.

Here, Za stands for the predicate (member of the tuple Z) whose index is a; likewise, Zb stands for the predicate whose index is b.

We view the minimization of a predicate (e.g., $ab3$) as, essentially, a *default* premise belief (schema) (e.g., $\neg ab3(x)$), analogous to an open normal default without pre-requisite in Default Logic [Reiter, 1980]. [Grosof, 1992] (sections 3.3 and 8.5) shows the formal equivalence up to conservative extension, in prioritized circumscription, of maximizing (first-order, open) default formulas to minimizing predicates: essentially, via the well-known "abnormalities" method ([McCarthy, 1986] [Grosof, 1984] [Lifschitz, 1984]). The example in section 7 will use abnormalities in this spirit. [9] Thus, in the above notation for PPC, N specifies which default formulas are to be *adopted* as default premise beliefs, i.e., as *default axioms*.

In the above PPC notation, $R(b, a)$ means that the predicate (/ default) indexed (i.e., named) by b is to be minimized with higher precedence than the predicate (/ default) indexed by by a. Each predicate's minimization is a kind of primitive preference relation (criterion): the preference for the predicate's extension (in the model-theoretic sense) to be smaller. The prioritization operation, then, takes as input a tuple of starting preference relations (criteria) and results in a single, aggregated preference relation (criterion).

Concept of Policy: We define the *policy* of a circumscription as the part of its specification other than its base sentence: i.e., the specification of the overall minimized preference relation in the sense of general circumscription. In a prioritized predicate circumscription, the policy is the list of minimized predicates, and their relative prioritization.

In our definition above of PPC:
Together N (adoption) and R (prioritization) constitute the *policy* of the PPC; i.e., they specify the aggregated preference relation.

For more details about the definition of prioritized circumscription, see [Grosof, 1992] (chapter 2).

6 Defeasible Prioritization and Adoption

Next, we define our new formalism: DAP circumscription. For the sake of simplicity of presentation, we begin with the case without pointwise control. I.e., in this section, we give the definition of DAP circumscription for the *blunt* (non-"point"-y) case.

[9] See [Etherington, 1988] (Theorem 8.3) for a precise equivalence result about the above view. [Grosof, 1992] (especially, sections 8.5 and 3.3) generalizes this and discusses other related equivalences.

6.1 Concept and Overview

In order to express reasoning about prioritization and adoption (the policy), we need to represent prioritization and adoption within the language. That is, we need to be able to specify, and represent, the policy via axioms within the language. Yet above, N and R were meta-linguistic (relative to $\mathcal{L}$). Our first step, then, is to bring them into $\mathcal{L}$: we make them into new predicate symbols. This trick is taken from Lifschitz' scheme (discussed in section 2), which makes the policy predicate V into a new predicate symbol.

In prioritized predicate circumscription, N can be viewed as simply selecting among symbols in Z. To define its activity appropriately, we introduce into $\mathcal{L}$ the name (i.e., index in the sense above) of each predicate in $\mathcal{L}$. I.e., we introduce the new object symbol $'P$ for every predicate symbol P in Z (again, taken from Lifschitz). This expansion of the language is fairly modest: it is only linear in the number of symbols in the original base language. Essentially, we are introducing an extremely limited form of quotation (reification). $N('ab3)$, then, expresses the minimization of the predicate $ab3$. Asserting $N('ab3)$ as part of the base sentence of the circumscription will then assert the adoption of the corresponding open default $\neg ab3(x)$. Similarly, $R('ab3, 'ab6)$ says that the default corresponding to (the minimization of) $ab3$ has higher priority than the default corresponding to $ab6$.

Our next step is to enable reasoning about prioritization and adoption to be *defeasible*. Our method is to permit the policy predicates N and R to be affected by prioritized default reasoning at a higher precedence. More generally, we permit any finite number n of (totally-ordered) levels of this. I.e, we permit a *reflective tower* of arbitrary height. Each level i has its own tuple of policy predicates: $POLi \stackrel{\text{def}}{=} \langle Ni, Ri \rangle$. The prioritized defaults at each level are permitted to non-monotonically affect the policy predicates at lower levels. (Note that Lifschitz' formalism only represents one level of policy; this is much of the reason why it is limited to only monotonic reasoning about policy.)

We write N for the tuple of the Ni's, and R for the tuple of the Ri's. When $n = 1$ (e.g., for DAP1 defined below), we will drop the i subscripts.

In the next section, we will define *Defeasible Axiomatized Policy* circumscription with n (e.g., $n = 3$) levels (DAPn):

$DAPCn(B; N; R; Z)$

as a special case of generalized prioritized circumscription. DAPn circumscription includes prioritized predicate circumscription as a special case. Its definition uses a *meta-prioritization* prioritization partial order MR that is simply a finite total order. The meta-prioritization essentially says that level

$i+1$ has greater priority than level i, for $i = 1, \ldots, n-1$.

There is one further consideration. In applying our approach, we have found that it is also a useful (convenient and natural) specification convention to apply a kind of Closed World Assumption to the policy predicates at each level, after the higher levels have had their way. More precisely, we minimize the policy predicates for level i with meta-prioritization in-between level i and the immediately higher level $i+1$.

Monotonic Case; BAP: DAP circumscription includes monotonic as well as defeasible reasoning about the policy. However, to **define** *purely* monotonic reasoning about the policy requires a slight variant: of DAP1 circumscription. In this variant, which we call *Basic Axiomatized Policy (BAP)* circumscription, the policy predicates are fixed rather than minimized. But we have found that minimizing, rather than fixing, the policy is more specificationally convenient in all applications we have tried, including where there is only one level of reasoning about policy: DAP1 appears more useful than BAP for knowledge representation. BAP is nevertheless useful as a tool for analyzing DAPn's behavior (see section 9), and is interesting theoretically in some other regards.

6.2 Detailed Definition of DAP Circumscription

In this sub-section, we give full details of the definitions of DAP circumscription for the blunt (non-pointwise) case. There are several technical subtleties involved.

Let $Z \stackrel{\mathrm{def}}{=} \langle Y, POL \rangle$ stand for the tuple of all predicate symbols in $\mathcal{L}$, where Y is the tuple of all non-policy predicates in the original base language, and $POL \stackrel{\mathrm{def}}{=} \langle POL1, \ldots, POLn \rangle$. We also introduce, as new object (0-ary function) symbols in $\mathcal{L}$, the names of all predicate symbols (i.e., of Z), as discussed in sub-section 6.1.

We represent each level's activity with a preference relation $STDPOLi$.

$$Z^1 \preceq_{STDPOLi} Z^2 \stackrel{\mathrm{def}}{\equiv}$$
$$\forall a.\ Ni^1(a) \wedge [\forall b.\ Ni^1(b) \wedge Ri^1(b,a) \supset (Zb^1 = Zb^2)] \supset (Za^1 \leq Za^2)$$

Here, a and b are each a predicate name. Za stands for the predicate whose name is a; likewise Zb stands for the predicate whose name is b.

Note the similarity to the definition of prioritized predicate circumscription that we gave in section 5. Here, Ni plays the role of N there; Ri plays the role of R there. (Ni is viewed here as a unary predicate, rather than as a set or index selector, however.) A subtlety is that there are two alternative "candidates" to play the role of Ni: Ni^1 and Ni^2; likewise, there are two candidates to play the role of Ri: Ri^1 and Ri^2. Arbitrarily, we chose the candidates Ni^1 and Ri^1, respectively. Our definition will ensure that this arbitrary choice makes no difference; we discuss why below.

Recall that, as discussed in sub-section 6.1, we minimize the policy predicates for level i with meta-prioritization in-between level i and the immediately higher level $i+1$. We express the minimization of each $POLi$ via the pre-order $MINPOLi$.

$Z^1 \preceq_{MINPOLi} Z^2 \stackrel{\text{def}}{\equiv} [(Ni^1 \leq Ni^2) \wedge (Ri^1 \leq Ri^2)]$

The overall minimized preference relation in DAP circumscription is $MVSPn$. "$MVSPn$" is mnemonic for Minimized Variable Standard Policy, with n levels. $MVSPn$ is defined as the result of prioritizing the $STDPOLi$'s and the $MINPOLi$'s according to the meta-prioritization. That is, we use prioritization to aggregate these starting pre-orders (technically, preference relations), one for each level, into the overall pre-order (preference relation).

We define the meta-prioritization order MR as follows.[10] Let the index of $STDPOLi$ be the number $2i-1$. Let the index of $MINPOLi$ be the number $2i$. Let MR be ordinary arithmetic strictly-greater-than, defined over the indices $MN \stackrel{\text{def}}{=} \langle 1, \ldots, 2n \rangle$. Let $SMPOL$ stand for the tuple formed from all the $STDPOLi$'s and the $MINPOLi$'s. MN thus indexes $SMPOL$.

Then the overall preference relation $MVSPn$ is defined as the result of prioritizing the starting preference relations $SMPOL$ according to the prioritization order MR:[11]

$MVSPn \stackrel{\text{def}}{=} (SMPOL^{MN}; MR)$

This overall n-level policy first minimizes, then uses, the policy predicates at each level. This overall policy is "standard" in a particular sense which we will discuss in section 9: it is essentially the same for every DAP circumscription.

Note that the above meta-prioritization is a novel application of prioritization which exploits the generality of its definition in [Grosof, 1991]: the starting pre-orders are *not* prioritized *predicate* pre-orders, nor are they prioritized "formula" (cf. [McCarthy, 1986]) / "default" pre-orders.[12]

Recall our earlier observation that in defining $\preceq_{STDPOLi}$, there was a subtlety: we had to decide whether to use Ni^1 versus Ni^2, and whether to use Ri^1 versus Ri^2. Arbitrarily, we chose the former. But it made no difference to the overall $MVSPn$ because, effectively these "candidates" are fixed in a particular sense, which we will call, for now, "the context of their use". (Re-

[10]Note that MR is meta-linguistic relative to the base language $\mathcal{L}$.

[11]Recall the definition of prioritized pre-order in section 5.

[12]Technically speaking, this application of prioritization also involves a slight generalization of the idea of prioritization as defined in [Grosof, 1991] and studied in [Grosof, 1992], since the starting relations and the resulting relation are not unconditionally guaranteed to be pre-orders. See the discussion of relative circumscription below.

call the concept of fixing from section 5.) That is, in that context, Ni^1 is everywhere equivalent to Ni^2, and Ri^1 is everywhere equivalent to Ri^2. This fixing (in context) of Ni and Ri emerges from the minimization of $POLi$ by $\preceq_{MINPOL}$ at higher meta-prioritization. In section 9, we will give results that more precisely characterize this fixing and what we mean by the context of their use. [We discuss this point in more detail.]

There is a further subtlety involved in ensuring that $STDPOLi$ is really a well-defined pre-order (in the relativized context of its use): we need to enforce our intended interpretation that each $POLi$ be well-defined as a policy for a prioritized predicate circumscription. The most important aspect of this interpretation is that Ri be well-defined as a prioritization partial order. That is, we need to guarantee that Ri is indeed a well-founded strict partial order, when restricted to (the domain) Ni.

To accomplish this, we require that the overall base sentence of the DAPn circumscription always include an extra base axiom ($WELLDEFN[POL]$ below) that enforces our intended interpretation. To properly define this method of enforcement by inclusion of extra base axioms, however, we find it convenient to generalize a bit further the previous notion of "general" circumscription.

In general circumscription as defined by [Lifschitz, 1984], the minimized binary relation was required to *unconditionally* obey the reflexivity and transitivity properties (recall section 5). We find it useful to generalize this notion so as require only that the minimized binary relation obey those properties *relative to the base sentence*.

Definition 1 (Relative Circumscription)
A *relative* circumscription $RC(B;G;Z)$ is the minimization in $B[Z]$ of a binary *preference relation* G which, *relative* to the base sentence $B[Z]$, is guaranteed to be a pre-order. (Here, G is defined over Z, as in the definition of general circumscription in section 5.) That is:

$$RC(B;G;Z) \stackrel{\text{def}}{\equiv} B[Z] \wedge \neg\exists Z'.\, B[Z'] \wedge Z' \prec_G Z$$

where:

$$\forall Z.\, B[Z] \supset G(Z,Z)$$
$$\forall Z^1, Z^2, Z^3.\, B[Z^1] \wedge B[Z^2] \wedge B[Z^3] \supset \{G(Z^1,Z^2) \wedge G(Z^2,Z^3) \supset G(Z^1,Z^3)\}$$

The extra base axiom that accomplishes the enforcement is $WELLDEFN[POL]$, where:

$$WELLDEFN[POL] \stackrel{\text{def}}{\equiv} PRIORIT[R,N] \wedge UNANAMES$$

$$PRIORIT[R,N] \stackrel{\text{def}}{\equiv} \bigwedge_{i=1}^{n} PRIORIT[Ri,Ni]$$
$$PRIORIT[Ri,Ni] \stackrel{\text{def}}{\equiv} STRICTPO[Ri] \wedge WELLFND[Ri,Ni]$$
$$STRICTPO[Ri] \stackrel{\text{def}}{\equiv}$$
$$\forall x.\, \neg Ri(x,x) \;\wedge\; \forall x,y,z.\, Ri(x,y) \wedge Ri(y,z) \supset Ri(x,z)$$
$$WELLFND[Ri,Ni] \stackrel{\text{def}}{\equiv}$$
$$\forall Q.\, [(Q \leq Ni) \wedge (\exists x.\, Q(x))] \supset [\exists x.\, Q(x) \wedge \forall y.\, \neg Ri(y,x)]$$

$UNANAMES$ is defined as the Uniqueness of Names Axiom for the names of the predicates. $PRIORIT[Ri,Ni]$ expresses the property that Ri qualifies as a prioritization partial order (i.e., is a well-founded strict partial order) over the "domain" where Ni is true. $WELLFND[Ri,Ni]$ expresses the property that Ri is well-founded over Ni.

At last, we are ready to define DAP circumscription: as a relative circumscription.

Definition 2 (Defeasible Axiomatized Policy (DAP))
The *Defeasible Axiomatized Policy* circumscription with n (e.g., $n=3$) levels (DAPn), of any base sentence $B[Z]$ is defined as:

$$DAPCn(B;N;R;Z) \equiv RC(BB;MVSPn;Z)$$

where BB is defined as:

$$BB[Z] \stackrel{\text{def}}{\equiv} B[Z] \wedge WELLDEFN[POL]$$

and where $MVSPn$, Z, N, R, and $WELLDEFN[POL]$ are defined as discussed earlier in this section. [13]

For brevity, when we have no particular n in mind, we will often speak simply of DAP, rather than DAPn, circumscription.

Fixing: For the sake of simplicity of presentation, in our earlier definitions of circumscriptions, we neglected the ability to directly specify some subset of predicates to be fixed in the circumscription. Our definition of DAP (and BAP, below) circumscription straightforwardly generalizes to express fixing directly, however.

[13] Note that, for each level i, $WELLFND[Ri,Ni]$ is a second-order sentence.

We observe that if Ni is true only on the names of predicates, and this set of names is finite, then Ri is guaranteed to be well-founded on Ni. This is the manner in which we intend Ni to be used in the blunt (non-pointwise) case: e.g., in Theorem 8 and Theorem 13 (section 9); and in the Fiscal Agenda Example (section 7). In such a situation, or if BB otherwise implies it, $WELLFND[Ri,Ni]$ can equivalently be omitted in the definition of $WELLDEFN[R,N]$, and, therefore, in the definition of BB. BB is then expressible in first-order if B is.

In the pointwise case, the well-foundedness of Ri on Ni is more of an issue, however, since there may be infinitely many points which satisfy Ni.

6.3 Detailed Definition of BAP Circumscription

BAP circumscription is defined most easily as a modification of the one-level case of DAP circumscription. Next, we give this definition for the blunt (non-pointwise) case.

Definition 3 (Basic Axiomatized Policy (BAP))
To arrive at BAP circumscription: in the definition of DAP1 circumscription (Definition 2), replace $MINPOL$ by $FIXPOL$, where:

$$Z^1 \preceq_{FIXPOL} Z^2 \quad \stackrel{\text{def}}{\equiv} \quad [(N^1 = N^2) \ \wedge \ (R^1 = R^2)]$$

Notationally, we write the resulting modification of DAPC1 as:

$$BAPC(B; N; R; Z)$$

We defer further formal discussion of DAP / BAP circumscription (pointwise version, and analysis) until after an example.

7 Fiscal Agenda Example

Consider the following (of course, extremely fictional) story about political advice-taking, which illustrates how DAPn circumscription enables defeasible reasoning about prioritization. In this version of this paper, we formalize it using the blunt (non-pointwise) version of DAPn. (The pointwise version is representationally more concise and natural, but is a bit harder to follow technically, what with all the quantifiers and variables flying around.)

The agent is the U.S. President. It is a year before she comes up for re-election, and the U.S. economy is mired in a recession. She is reasoning about her biggest problem: what ought to be her agenda (try below) of fiscal actions to help improve the economy in the short term. She has two trusted advisers L (Larry) and Q (Quincy). By default, she takes their advice about fiscal matters. Sometimes, but not always, their advice conflicts. E.g., Larry's advice to try a 10% cut in the Federal personal income tax conflicts with Quincy's advice to try a 20% increase in revenue sharing to state and local governments, since it is impossible to get Congress to agree to the degree of budget-busting required to do both. To begin with, the President has no basis for precedence (prioritization) between their recommendations when they conflict.

Below, "Jawboning" means exhorting using the President's "bully pulpit" for general moral suasion. The advisers' recommendations about jawboning (to Wall Street to be optimistic investors, and to consumers to buy lots at Christmastime) do not conflict, since talk is cheap.

We represent the President's premise beliefs as an accumulating conjunction of base axioms in DAP2 circumscription as follows.

Notation: All of the axioms are implicitly universally quantified. We assume all terms are distinct. $\approx\!\!\!\!\!\mid\;\; p$ stands for p being non-monotonically entailed by the circumscription.

$$\neg ab1(x) \;\supset\; \neg try(x)$$
$$N1('ab1)$$
$$\neg ab2(x) \;\supset\; [advise(L,x) \wedge subject(x, Fiscal) \;\supset\; try(x)]$$
$$\neg ab3(x) \;\supset\; [advise(Q,x) \wedge subject(x, Fiscal) \;\supset\; try(x)]$$
$$N1('ab2) \;\wedge\; N1('ab3) \;\wedge\; R1('ab2,'ab1) \;\wedge\; R1('ab3,'ab1)$$
$$advise(L, JawBone(Consumers, Xmas) \;\wedge$$
$$subject(JawBone(Consumers, Xmas), Fiscal)$$
$$advise(Q, JawBone(WallSt, Optimism) \;\wedge$$
$$subject(JawBone(WallSt, Optimism), Fiscal)$$
$$advise(L, TaxCut10\%) \;\wedge\; subject(TaxCut10\%, Fiscal)$$
$$advise(L, RevenueShare20\%) \;\wedge\; subject(RevenueShare20\%, Fiscal)$$
$$\neg[try(TaxCut10\%) \;\wedge\; try(RevenueShare20\%)]$$
$$\mid\!\approx\; try(JawBone(Consumers, Xmas)) \;\wedge$$
$$try(JawBone(WallSt, Optimism))$$
$$\mid\!\approx\; try(TaxCut10\%) \;\vee\; try(RevenueShare20\%)$$
$$\not\mid\!\approx\; try(TaxCut10\%) \quad ; \quad \not\mid\!\approx\; try(RevenueShare20\%)$$
$$\mid\!\approx\; \neg R1('ab2,'ab3) \;\wedge\; \neg R1('ab3,'ab2)$$
$$\mid\!\approx\; \neg try(DoublePublicWorks) \;\wedge\; \neg try(InvadeLibya) \;\wedge\; \ldots$$

Now suppose also that she has the default beliefs that Larry is more expert on fiscal matters, and that relative fiscal expertise is a basis for prioritization between their recommendations. She then concludes non-monotonically to try the tax cut.

$$\neg ab4 \;\supset\; [moreexpert(L, Q, Fiscal) \;\supset\; R1('ab2,'ab3)]$$
$$\neg ab5 \;\supset\; [moreexpert(Q, L, Fiscal) \;\supset\; R1('ab3,'ab2)]$$
$$N2('ab4) \;\wedge\; N2('ab5)$$
$$\neg ab6(x,y,z) \;\supset\; \neg moreexpert(x,y,z)$$
$$N2('ab6) \;\wedge\; R2('ab6,'ab4) \;\wedge\; R2('ab6,'ab5)$$
$$\neg ab7 \;\supset\; moreexpert(L, Q, Fiscal)$$
$$N2('ab7) \;\wedge\; R2('ab7,'ab6)$$
$$\mid\!\approx\; try(TaxCut10\%)$$
$$\mid\!\approx\; R1('ab2,'ab3)$$

The justifying conclusion ($R1('ab2,'ab3)$) about relative prioritization is itself defeasible. Suppose she, further, has the default beliefs that Quincy is more expert at marketing political ideas, and that that is more important than relative fiscal expertise in prioritizing their recommendations. She then *reverses* *her previous non-monotonic conclusion about the prioritization* between the recommendations, and, therefore, concludes non-monotonically to try the revenue sharing instead.

$$\begin{array}{l}\neg ab8 \;\supset\; moreexpert(Q, L, PoliticalMarketing) \\ \neg ab9 \;\supset\; [moreexpert(Q, L, PoliticalMarketing) \;\supset\; R1('ab3, 'ab2)] \\ N2('ab8) \;\wedge\; N2('ab9) \;\wedge\; R2('ab8, 'ab7) \;\wedge\; R2('ab9, 'ab7) \\ \qquad \approx\!\!\!\!| \;\; try(RevenueShare20\%) \\ \qquad \approx\!\!\!\!| \;\; R1('ab3, 'ab2)\end{array}$$

8 Pointwise

Definition 4 (Point)
A *point* is a tuple (e.g., $\langle 'ab2, Nixon\rangle$) that corresponds to an atom of the base language $\mathcal{L}$ (e.g., $ab2(Nixon)$). The first member of the tuple is a predicate name (e.g., $'ab2$). The remaining members of the tuple correspond to arguments of the predicate named by the first member of the tuple. E.g., let $Loves$ be a binary predicate. Then $\langle 'Loves, Jack, Jill\rangle$ is a point.

Definition 5 (Extension of DAP / BAP to Pointwise)
The definition of DAP / BAP circumscription in section 6 extends straightforwardly to the pointwise version. Definitions 2 and 3 are modified in two regards.

1. The definitions of the policy predicates (N and R) are generalized: to take *points* as their arguments, rather than simply predicate names. (This is similar to Lifschitz' [1988] formalism[14].) Each Ni takes a single point as its argument. Each Ri takes a pair of points as its argument tuple.

   For example, suppose $ab3$ and $ab4$ are predicate symbols; i.e., $ab3, ab4 \in Z$. Suppose further that $ab3$ has arity 2, while $ab4$ has arity 3. Then
   $N1('ab3, x, y)$
   $R1('ab4, Joe, Karen, Larry)$
   are well-formed atoms.

2. In the definition of $STDPOLi$ (early in sub-section 6.2), a and b are generalized to be points; Za and Zb thus each stand for an atom. E.g., if a is $\langle 'ab2, Nixon\rangle$, then Za stands for $ab2(Nixon)$.

[14]Recall our discussion of it early in section 2 and sub-section 6.1.

First Formal Definition of Pointwise Prioritization: Interestingly, Lifschitz in his previous work on pointwise circumscription [1987] [1988] never formally defined pointwise prioritization; he just gave examples. With the BAP case of Definition 5, we have thus provided, for the first time, a formal definition of pointwise prioritization [We elaborate.]: unlike Lifschitz' scheme, it is a direct generalization of the blunt version of prioritization (cf. [Lifschitz, 1985] [Grosof, 1991]). Alternatively, the DAP1 case of Definition 5 can be used as a somewhat more complex definition of pointwise prioritization; we find its minimization of $N1$ and $R1$ is often representationally convenient. Later, we will clarify the relationship between DAP1 and BAP: see Theorem 6 in section 9.

9 Analysis: Expressive Power and Inference

Standard Policy: Note that DAP circumscription employs a *standard* policy: every DAP circumscription is defined as the minimization of, essentially, the *same* pre-order (same up to: syntactic re-naming of symbols; plus, how many symbols are in $\mathcal{L}$; plus, n, i.e., how many levels are considered) in which the policy predicates have a special role. In the case of DAP1, this standardization makes it possible to represent equivalently within $\mathcal{L}$ the specification of the policy of any particular prioritized predicate circumscription (see Theorem 8 below). By contrast, the policies of various prioritized predicate circumscriptions may differ greatly.

Likewise, BAP circumscription also employs a standard policy. It too can represent any particular PPC (see Theorem 7 below).

Review: Series Circumscription and Serial Decomposition: We **define** a **series** of circumscriptions (more briefly, a "series circumscription") to be a cascade in which each *phase* circumscription takes the result of the previous as its base sentence. A series circumscription is specified by (1) the *starting* base sentence, which is the base sentence used by the first phase; and (2) the sequence (tuple) of phase policies (by "policy" here, we mean in the sense of general (relative) circumscription). [Grosof, 1992] defines and studies "serial decompositions" of prioritized circumscriptions. For our purposes in this paper, we **define** a **serial decomposition** to be an equivalence of a circumscription to a series of other circumscriptions.

Theorem 6 (Serial Decomposition of DAP1 to PPC and BAP)

Every DAP1 circumscription is serially decomposable into two phases. In the first phase (a PPC), the policy predicates $N1$ and $R1$ are minimized in parallel. The second phase is BAP circumscription.

Intuitively, the second phase simply uses the policy predicates that result from the first phase.

More formally:
Since $n = 1$, let us notationally omit the level index 1 when writing the policy predicates. The theorem can then be stated as follows.
For every base B,

$$DAPC1(B; N; R; Z) \;\equiv\; BAPC(BPM; N; R; Z)$$

where

$$BPM \;\stackrel{\text{def}}{\equiv}\; PPC(B; N, R; \emptyset; Z)$$

That is, BPM is the prioritized predicate circumscription in which the policy predicates N and R are minimized in parallel (i.e., with empty priority between their minimizations).

Proof Sketch: Uses Theorem 5.32 from [Grosof, 1992] about serial decomposition of general prioritized circumscriptions on the basis of descending prioritization.

Theorem 7 (Correspondence of PPC to BAP)
Consider an arbitrary prioritized predicate circumscription (PPC), as defined in section 5, whose base is $B[Y]$, and where Y is the tuple of all predicate symbols in its ("original") base language:

$$PPC(B; (Y^{NM}; RM); Y)$$

Let N and R be new predicates in the base language, cf. the discussion at the beginning of sub-section 6.1, that correspond to NM and RM, respectively. (Remember from section 5: NM and RM are defined meta-linguistically in terms of indices of the tuple Y.) Let $AC[N, R]$ be a complete axiomatization of $\langle N, R\rangle$ that makes N correspond to NM and R correspond to RM. Thus AC says that N is true exactly on the names of (more generally, on all the points belonging to) the predicates Y^{NM} (recall the notation Z^N in section 5). Let BAC be defined as the conjunction of B and AC.
Then the BAP circumscription (BAPC) whose base is BAC is essentially equivalent to the PPC. More precisely, the BAPC is a conservative extension of the PPC: the BAPC is equivalent on the original Y sub-language in which the PPC is defined; however, it also mentions the policy predicates.

$$BAPC(BAC; N; R; Z) \;\equiv\; AC[N, R] \wedge PPC(B; Y^{NM}; R^{NM}; Y)$$

$$PPC(B; Y^{NM}; R^{NM}; Y) \;\equiv\; \exists N, R.\; BAPC(BAC; N; R; Z)$$

Proof Sketch: Fairly direct from the definition of BAP circumscription.

Theorem 8 (Reduction of DAP1 to PPC)

I) Theorem 7 holds when the BAP circumscription is changed to a DAP1 circumscription.
II) For the DAP1 case, the condition on the base can be weakened. In particular, in the statement of Theorem 7, one may replace BAC by BAA (as well as $BAPC$ by $DAPC1$), where:

$$BAA[Z] \stackrel{\text{def}}{\equiv} B[Y] \wedge AA[N, R]$$

where AA is the positive part of AC. That is, let AA be the *atomic axiomatic representation* of N and R, by which we mean the conjoining of one N atom for each predicate name that satisfies N (i.e., for each predicate whose index is in NM), plus one R atom for each pair of predicate names that satisfies R. Then

$$DAPC1(BAA; N; R; Z) \equiv AC[N, R] \wedge PPC(B; (Y^{NM}; RM); Y)$$
$$PPC(B; Y^{NM}; R^{NM}; Y) \equiv \exists N, R.\ BAPC(BAC; N; R; Z)$$

Proof : **I)** follows from Theorems 6 and 7, plus the fact that PPC is ineffectual when the base is complete in the minimized predicates. By "ineffectual", we mean that the PPC is equivalent to its base.

II) follows from those Theorems plus the fact that the first-phase PPC resulting from serial decomposition according to Theorem 6 has precisely the effect of (the negative part of) AC. This fact can be shown in a number of different ways from the previous circumscription literature. In particular, it is an immediate special case of Theorem 6.18 of [Grosof, 1992], which generalizes the "solitary" case result of [Lifschitz, 1985], which in turn generalized the predicate-completion result of [Reiter, 1982]. □

PPC Is Special Case of DAP / BAP: Observe that Theorems 7 and 8 imply that PPC is a special case of BAP, of DAP1, and thus of DAPn, circumscription.

Fundamental Increase in Expressiveness: DAP circumscription's expressive power is fundamentally greater than prioritized *predicate* circumscription, in an important sense. Theorem 8 tells us that the bottom level in DAP circumscription often corresponds to ordinary prioritized predicate circumscription. This bottom level of non-monotonic reasoning employs a policy (i.e., preference relation) which we can view as (corresponding to) a preference criterion on models. The higher levels produce defeasible reasoning about this policy. Thus *the model-preference criterion itself is changing* as new base axioms are added; this is fundamentally different from prioritized predicate circumscription, and from all (that we are aware of) previous approaches to model-preference NMR in the literature. In terms of [Doyle and Wellman, 1991], both the aggregation method (parameterized by the prioritization partial order $R1$) and the starting set of preference criteria (the defaults, parameterized by the adoption selector $N1$) are being *modified* under updating with new base axioms. [We make this argument in more detail.]

DAP Is a Strict Generalization of PPC: From a mathematical perspective: DAP circumscription is a special case of *generalized* prioritized (relative) circumscription. The overall minimized preference relation $MVSPn$ is formed by prioritization of a tuple of starting preference relations: the $STDPOLi$'s and the $MINPOLi$'s. However, the $STDPOLi$'s are *not* prioritized predicate pre-orders, and thus neither is $MVSPn$. Thus, in general, a DAP circumscription is *not* a prioritized *predicate* circumscription (except in special cases, e.g., as in Theorem 8).

Representational Conciseness: Less formally, but probably more importantly for knowledge representation, we argue that DAP circumscription is a tool that can often yield at least one dimension (i.e., defeasible prioritization) of greater representational adequacy and natural-ness than prioritized predicate circumscription and all (that we are aware of) other current non-monotonic logical formalisms for specifying default reasoning, at least in their currently-conceptualized conventions / patterns of use for applications. DAP circumscription is often more representationally concise, in the sense of requiring less modification between members of a collection of alternative belief (conclusion) sets, than prioritized predicate circumscription, in essentially the same way that (prioritized) predicate circumscription is often more representationally concise than monotonic (classical, first-order) logic. [We illustrate by example.] Indeed, considerations of representational conciseness (e.g., in Closed World Assumption, frame problems, qualifications, ramifications) have been much of the motivation for the development and application of non-monotonic logics in general. [We make this argument in more detail.]

The Reducibility of Depth: AEL is probably the best-studied of the previous systems that can represent default beliefs about default beliefs. An intriguing set of results about it [Konolige, 1988a] [Marek and Truszczynski, 1989] are its reducibility, under certain technically subtle conditions and variations about its semantics, to Default Logic (DL); and, conversely, the reducibility of DL to AEL. Since AEL permits nested NM belief operators, while DL is defined in "depth-one" fashion, these equivalence results mean that either (the relevantly defined version of) AEL has only spurious depth, or that (the relevantly defined version of) DL hides "unsuspected" (i.e., from its depth-1 appearance) expressive richness. Which interpretation is more appropriate is still an open question.

Circumscription[15] has been viewed as in the "flat", sort of "depth-1" cate-

[15] general circumscription cf. [Lifschitz, 1984], as opposed to "introspective circumscription" [Lifschitz, 1989]

gory with respect to nesting of belief, e.g., cf. modal NM logics such as AEL. We showed how to overcome that seeming limitation.

However, we have as a **theorem** that one can express n-level defeasible adoption essentially equivalently in one level (i.e., even more simply, in ordinary prioritized predicate circumscription): by generalizing the by-now well-known method of cancellation of an abnormality default. The generalization is to make the cancellation be a default with higher ordinary (rather than meta-) prioritization. [16] Thus, in a particular sense, depth-n in our scheme reduces to depth-1, for adoption. **For prioritization, by contrast, we have not been able to find any such reduction.** [We give full details, and discuss context of related work.]

Advantages Compared to Lifschitz' Approach; Improved Approach to Pointwise Circumscription: Most importantly, our approach can represent defeasible prioritization, while Lifschitz' approach cannot, at least in any manner yet known. In the (pointwise or blunt) case of monotonic reasoning about prioritization and adoption, the expressive power of our approach (whether BAP or DAP1 is used) is similar to Lifschitz'. However, even in this case, our approach has several advantages over his.

Firstly, Lifschitz' formalism permits *"nasty" unsatisfiability* of an extra kind. By nasty unsatisfiability, we mean the situation in which the circumscription is unsatisfiable, even though the base sentence is satisfiable; i.e., the circumscription operation itself introduces inconsistency. Circumscription and other model-preference NM formalisms are well-known to be susceptible to nasty unsatisfiability: when models are not "well-founded" in the sense defined by [Etherington *et al.*, 1985] [Lifschitz, 1986]. But Lifschitz' formalism permits a kind of incoherence in the policy which results in nasty unsatisfiability even in very simple, "well-founded" theories. Our approach does not.

Secondly, Lifschitz' formalism cannot directly represent prioritization that is not layered (i.e., not "stratified"). [Grosof, 1991] (elaborated in [Grosof, 1992]) showed that non-layered prioritization is very important for applications. While it is possible (using reduction theorems given in [Grosof, 1992]) to represent non-layered prioritization in Lifschitz' formalism in an indirect manner, it is often rather awkward, involving a substantial increase in the expressive complexity (e.g., clausality / universality may be lost) and size (e.g., squared blow-up) of the representation.

[We give more details, including examples, of these first two advantages.]

Thirdly, because our approach uses the more epistemogically natural con-

[16] We observe that this result gives some perspective on why abnormality form is so popular representationally in NMR: it makes retraction of *adoption* easy.

cept of prioritization, rather than variability, it facilitates knowledge acquisition.

Fourthly, because our approach is a special case of single general circumscription, it is easier to apply results available from the body of work on model-preference logics (e.g., [Shoham, 1988] [Brown and Shoham, 1989] [Boddy *et al.*, 1989] [Selman and Kautz, 1989]), and on preferential logics (e.g., [Kraus *et al.*, 1990] [Lehmann, 1989])

Theorem 9 (Serial Decomposition of DAPn Into DAP1)
Every DAPn circumscription is equivalent to a series of DAP1 circumscriptions: one for each level, in descending meta-prioritization order. More precisely, let:

$$B_0 \stackrel{\text{def}}{\equiv} B \qquad ;$$

and let, $\forall i = 0, \ldots, n-1.$:

$$B_{i+1} \stackrel{\text{def}}{\equiv} DAPC1(B_i; N_{n-i}; R_{n-i}; Z)$$

(Here, the i^{th} phase corresponds to the $(n-i+1)^{th}$ level, and results in B_i.)
Then:

$$DAPCn(B; N; R; Z) \equiv B_n.$$

Proof Sketch: Uses Theorem 5.3.2 from [Grosof, 1992] about "serial decomposition" of general prioritized circumscriptions on the basis of descending prioritization.

Theorem 10 (Serial Decomposition of DAPn Into BAP and PPC)
Theorems 9 and 6 together imply that any DAPn circumscription can be serially decomposed into a series in which PPC and BAP phases alternate. There are n BAP phases: one per level.

Proof : Apply Theorem 6 to each of the DAP1 phases in the result of decomposing according to Theorem 9. □

Our earlier explanation (sub-section 6.2) about the fixing of $POLi$ in the context of its use is most easily thought of in terms of the above serial-decomposition form of the $DAPCn$. The context of use of $POLi$ is the BAP for phase i.

Theorem 11 (Well-Definition of DAP Circumscription)
For any n and B, DAP circumscription is well-defined: i.e., it is indeed a relative circumscription.

Likewise, for any B, BAP circumscription is well-defined as a relative circumscription.

Proof Sketch: The proof for the BAP case follows fairly directly from the definition of BAP. To prove the DAP case, the main issue is to show that $MVSPn$ is indeed guaranteed to be a pre-order, i.e., transitive (and reflexive), given the base (BB from Definition 2). This follows fairly directly from Theorem 10, combined with the BAP case.

Meta-Prioritization Implies Prioritization: One property to note about the meta-prioritization is that it makes each default specified at a given level i, even if not "about" policy predicates, have higher precedence (prioritization) than every default specified at any lower level.

Inference in DAPn: Next, we show that, for a large class of theories, including the Fiscal Agenda Example in section 7, the problem of inference in DAPn circumscription reduces to the problem of inference in prioritized predicate circumscription.

Definition 12 (Positive-Definite)
A sentence $B[Z]$ is *positive-definite* in the predicates $Q \subseteq Z$ iff there exist a pair sentences BE and BF such that:

1. $$B[Z] \;\equiv\; BE[Q] \,\wedge\, BF[Z]$$
2. All predicates appearing in BE are members of Q.
3. BE is a (finite) conjunction of positive unit clauses.
4. All appearances of Q in BF are syntactically negative.

Theorem 13 (Reduction of DAPn to Series PPC)
Suppose that a blunt $DAPCn(B; N; R; Z)$ has the property that the base in each phase of its serial decomposition, cf. Theorem 9, is (monotonically) equivalent to a formula that is *positive-definite* about the policy predicates used in that phase. I.e., suppose that, in that serial decomposition, for each level $i = 0, \ldots, n-1$, B_i is positive-definite in $\langle N_{n-i}, R_{n-i}\rangle$ after the $MINPOL\{n-i\}$ phase.

Then the $DAPCn$ is equivalent to a *series* of prioritized predicate circumscriptions, 2 per meta-prioritization level, again in descending meta-prioritization sequence. The starting base sentence of the series is just B. For each level i, the first phase policy is, simply, to minimize the policy predicates $POLi$; the second phase policy is, simply, to employ $POLi$ (as defined, in the manner of Theorem 8, by the $POLi$-part of the first phase's result).

Proof Sketch: From Theorem 9 and a more general version of Theorem8.

(This result generalizes to the pointwise case, but gets more complicated.)

Thus for a large class of theories, including the Fiscal Agenda Example above, the problem of inference in DAPn circumscription reduces to the problem of inference in prioritized predicate circumscription, for which inference procedures are currently available (e.g., [Przymusinski, 1989]). [We give more discussion of inference.]

More generally, though, the policy in even DAP1 (or BAP) circumscription may be indefinite. [We discuss how the DAP1 circumscription (e.g., phase of DAPn) can then be equivalent to a disjunction of prioritized predicate circumscriptions with differing policies.]

Theorem 13 is quite useful for proving entailments in hand-worked examples (e.g., Fiscal Agenda) of DAPn circumscription.

Prioritized predicate circumscription has a number of inference procedures currently available for broad classes: e.g., [Przymusinski, 1989] [Ginsberg, 1989] [Baker and Ginsberg, 1989] [Inoue and Helft, 1990] [Helft *et al.*, 1991] [Grosof, 1992] (chapter 7). However, they are for query-answering only. This stymies the direct application of Theorem 13 to them to generate an inference procedure for DAPn circumscription. The problem is that each phase needs, as its "input" base sentence, a finite axiomatization of the *entire* circumscriptive theory from the previous phase. Exhaustive *forward* inference is required to generate such an entire theory. [Grosof, 1992] studies inference in serial prioritized predicate circumscriptions.

Nevertheless, Theorem 13 suggests that DAPn circumscription will often not require much greater computational complexity than prioritized predicate circumscription.

10 Conclusions and Future Work

Conclusions: We motivated our effort with a discussion of applications in learning agents. We argued that there the capability for reasoning defeasibly about prioritization appears desirable. We illustrated this point with the Fiscal Agenda example. ([Grosof, 1993] gives another detailed example, in the domain of bureaucracy.) **For rest of conclusions:** see the abstract; and see section 3 for discussion of contributions and limitations.

A word on methodology: We identified a new representational issue (defeasible prioritization) and proposed an approach (DAP circumscription) that is formally somewhat hairy, but can do the job (on at least several example applications). Several readers of earlier drafts posed the question of whether all this mathematical complexity is really necessary. Of course, it is entirely possible that it is not, in an ultimate sense. Our defense, however, is that it is rare in research to show that there does *not exist* a better solution to a representational issue. We believe the above (identification, proposal, examples) is an appropriate kind of contribution; we leave it up to you to find a solution that is more elegant or powerful, or to show how our scheme is actually representationally reducible to a simpler one.

Future Work: We would like to combine the advantages of our approach with those of modal and conditional NM logics. We plan first to investigate the applicability of our scheme to AEL and introspective circumscription [Lifschitz, 1989]. Our scheme can also be extended straightforwardly to make the meta-prioritization be non-layered (non-stratified), to bring the meta-prioritization into the language and enable defeasible reasoning about it, and so on. More experience with applications is needed first, though. An important issue is whether it is representationally awkward to have to partition defaults by meta-prioritization level. Other issues we plan to investigate are: putting limited quotation into the language; generalizing to multiple agents; and, last but not least, developing inference methods.

References

[Appelt and Konolige, 1988] Douglas Appelt and Kurt Konolige. A practical nonmonotonic theory for reasoning about speech acts. Technical Report Technical Note 432, SRI International, 333 Ravenswood Ave., Menlo Park, California 94025, Apr 1988.

[Bacchus, 1990] Fahiem Bacchus. *Representing and Reasoning with Probabilistic Knowledge.* MIT Press, Cambridge, Mass., 1990.

[Bacchus, 1991] Fahiem Bacchus. Default reasoning from statistics. *Proceedings of AAAI-91*, pages 392–398, 445 Burgess Drive, Menlo Park, CA 94025, 1991. AAAI Press.

[Baker and Ginsberg, 1989] A. Baker and M. Ginsberg. A theorem prover for prioritized circumscription. *Proceedings IJCAI-89*, pages 463–467, Detroit, MI., 1989.

[Boddy *et al.*, 1989] Mark Boddy, Robert P. Goldman, Keiji Kanazawa, and Lynn A. Stein. Investigations of model-preference defaults. Working paper, submitted to Computational Intelligence journal., 1989.

[Brown and Shoham, 1989] Allen L. Brown, Jr. and Yoav Shoham. New results on semantical nonmonotonic reasoning. In M. Reinfrank *et al*, editor, *Proceedings of the Second International Workshop on Non-Monotonic Reasoning*, Berlin, Germany, 1989. Springer Lecture Notes on Computer Science.

[de Kleer and Konolige, 1989] Johann de Kleer and Kurt Konolige. Eliminating the fixed predicates from a circumscription. *Artificial Intelligence*, 39:391–398, 1989.

[Delgrande, 1987] James P. Delgrande. A first-order conditional logic for prototypical properties. *Artificial Intelligence*, 33:105–130, 1987.

[Doyle and Wellman, 1991] Jon Doyle and Michael P. Wellman. Impediments to universal preference-based default theories. *Artificial Intelligence*, 49:97–128, 1991. Revised and expanded from version in *Proceedings of the First International Conference on Principles of Knowledge Representation*, pp. 94-102, Toronto, Ontario, Morgan Kaufmann, 1989.

[Etherington *et al.*, 1985] D. Etherington, R. Mercer, and R. Reiter. On the adequacy of predicate circumscription for closed-world reasoning. *Computational Intelligence*, 1:11–15, 1985.

[Etherington, 1988] D. Etherington. *Reasoning with Incomplete Information.* Pitman, London, 1988.

[Ginsberg, 1986] Matthew L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–80, 1986.

[Ginsberg, 1989] M. Ginsberg. A circumscriptive theorem prover. *Artificial Intelligence*, 39:209–230, 1989.

[Grosof and Russell, 1990] Benjamin N. Grosof and Stuart J. Russell. Shift of bias as non-monotonic reasoning. In Pavel Brazdil and Kurt Konolige, editors, *Machine Learning, Meta-Reasoning, and Logics.* Kluwer Academic, 1990. Based on the workshop held in Sesimbra, Portugal, in Feb. 1988. Also available as IBM Research Report RC14620.

[Grosof, 1984] Benjamin N. Grosof. Default reasoning as circumscription. *Proceedings of the First AAAI Non-Monotonic Reasoning Workshop*, pages 115–124, Oct 1984. Held New Paltz, NY.

[Grosof, 1988] Benjamin N. Grosof. Non-monotonicity in probabilistic reasoning. In J. Lemmer and L. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 237–249. Elsevier Science Publishers, 1988. Volume containing revised versions of papers appearing in Proceedings of the Second International Workshop on Uncertainty in Artificial Intelligence, held Philadelphia, PA, August 1986.

[Grosof, 1991] Benjamin N. Grosof. Generalizing prioritization. *Proceedings of the Second International Conference on Principle of Knowledge Representation and Reasoning*, pages 289–300, April 1991. Also available as IBM Research Report RC15605, IBM T.J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598.

[Grosof, 1992] Benjamin N. Grosof. *Updating and Structure in Non-Monotonic Theories.* PhD thesis, Computer Science Dept., Stanford University, Stanford, California 94305, October 1992. Published by University Microfilms, Inc.. Also available as Research Report from Stanford University Computer Science Dept. and/or IBM T.J. Watson Research Center.

[Grosof, 1993] Benjamin N. Grosof. Prioritizing multiple, contradictory sources in common-sense learning by being told; or, advice-taker meets bureaucracy. In Leora Morgenstern, editor, *Proceedings of the Second Symposium on Logical Formalizations of Common-Sense Reasoning (Common-Sense '93)*, IBM T.J. Watson Research Center, P.O. Box 704, Yorktown

Heights, NY 10598, 1993. Available as an IBM Research Report. Copies of the Proceedings are available via the editor (wider publication being arranged). Held Guest Quarters Hotel, Austin, Texas, Jan. 11-13, 1993.

[Helft *et al.*, 1991] Nicolas Helft, Katsumi Inoue, and David Poole. Query answering in circumscription. *Proceedings of IJCAI-91*, pages 426–431, San Francisco, California, 1991. Morgan Kaufmann.

[Helft, 1989] Nicolas Helft. Induction as nonmonotonic inference. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, San Francisco, California, 1989. Morgan Kaufmann. Held Toronto, Ontario, Canada, May 1989.

[Inoue and Helft, 1990] Katsumi Inoue and Nicolas Helft. On theorem provers for circumscription. *Proceedings of the Canadian Conference on Computer Science and Artificial Intelligence '90*, 1990. Held Ottawa, Canada, May 1990.

[Konolige, 1988a] K. Konolige. On the relation between default logic and autoepistemic logic. *Artificial Intelligence*, 35:343–382, 1988.

[Konolige, 1988b] Kurt Konolige. Hierarchic autoepistemic theories for nonmonotonic reasoning. *Proceedings of AAAI-88*, pages 439–443. Morgan Kaufmann, 1988. Held Minneapolis, MN.

[Kraus *et al.*, 1990] S. Kraus, D. Lehmann, and M. Magidor. Preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Lehmann, 1989] D. Lehmann. What does a conditional knowledge base entail? *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 212–222, Toronto, Ontario, 1989. Morgan Kaufmann.

[Levesque, 1990] Hector Levesque. All i know: A study in autoepistemic logic. *Artificial Intelligence*, 42, 1990.

[Lifschitz, 1984] Vladimir Lifschitz. Some results on circumscription. *Proceedings of the First AAAI Non-Monotonic Reasoning Workshop*, pages 151–164, Oct 1984. Held New Paltz, NY.

[Lifschitz, 1985] V. Lifschitz. Computing circumscription. *Proceedings IJCAI-85*, pages 121–127, Los Angeles, CA, 1985.

[Lifschitz, 1986] V. Lifschitz. On the satisfiability of circumscription. *Artificial Intelligence*, 28:17–27, 1986.

[Lifschitz, 1987] Vladimir Lifschitz. Pointwise circumscription. In Matthew L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, San Francisco, CA, 1987.

[Lifschitz, 1988] V. Lifschitz. Circumscriptive theories: a logic-based framework for knowledge representation. *Journal of Philosophical Logic*, 17:391–441, 1988.

[Lifschitz, 1989] Vladimir Lifschitz. Between circumscription and autoepistemic logic. *Proceedings of the First International Conference on Principle of Knowledge Representation and Reasoning*, pages 235–244, Toronto, Ontario, 1989.

[Marek and Truszczynski, 1989] W. Marek and M. Truszczynski. Relating autoepistemic and default logics. *Proceedings of the First International Conference on Principle of Knowledge Representation and Reasoning*, pages 276–288, Toronto, Ontario, 1989.

[McCarthy, 1986] J. McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[McDermott and Doyle, 1980] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[McRoy, 1989] Susan Weber McRoy. Nonmonotonic reasoning in natural language. Dept. of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4., Nov 1989.

[Moore, 1985] R. Moore. Semantical considerations on non-monotonic logics. *Artificial Intelligence*, 25:75–94, 1985.

[Morgenstern and Guerriro, 1991] Leora Morgenstern and Ramiro Guerriro. Epistemic logics for multiple agent nonmonotonic reasoning I. Submitted. IBM T.J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598. Includes revised version of contents of paper appearing in Proceedings of AAAI-90, entitled "A Formal Theory of Multiple Agent Nonmonotonic Reasoning", by Leora Morgenstern., Sep 1991.

[Morgenstern, 1990] Leora Morgenstern. A formal theory of multiple agent non-montonic logics. *Proceedings of AAAI-90*, San Francisco, California, 1990. Morgan Kaufmann.

[Perrault, 1987] C. Raymond Perrault. An application of Default Logic to speech act theory. Technical Report CSLI-87-90, Center for the Study of Language and Information, Stanford University, Ventura Hall, Stanford University, Stanford, California 94305, Mar 1987.

[Przymusinski, 1989] T. Przymusinski. An algorithm for circumscription. *Artificial Intelligence*, 38:49–73, 1989.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 12:81–132, 1980.

[Reiter, 1982] Raymond Reiter. Circumscription implies predicate completion (sometimes). *Proceedings of AAAI-82*, pages 418–420. American Association for Artificial Intelligence, Menlo Park, CA, 1982. Held Pittsburgh, PA.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[Russell and Grosof, 1987] Stuart J. Russell and Benjamin N. Grosof. A declarative approach to bias in concept learning. *Proceedings of AAAI-87*, pages 505–510, San Francisco, CA, 1987. Morgan Kaufmann. Held Seattle, WA. The first and second halves are complementary but distinct papers. The first is a preliminary version of [Russell and Grosof, 1990b]. The second is a preliminary version of [Grosof and Russell, 1990].

[Russell and Grosof, 1990a] Stuart J. Russell and Benjamin N. Grosof. Declarative bias: An overview. In Paul Benjamin, editor, *Conference on Reformulation and Inductive Bias*. Kluwer Academic, 1990. Edited volume based on the Philips Workshop held in Briarcliff, NY, June 8-10, 1988. Also available as IBM Research report RC15023.

[Russell and Grosof, 1990b] Stuart J. Russell and Benjamin N. Grosof. A sketch of autonomous learning using declarative bias. In Pavel Brazdil and Kurt Konolige, editors, *Machine Learning, Meta-Reasoning, and Logics*. Kluwer Academic, 1990. Based on the workshop held in Sesimbra, Portugal, in Feb. 1988. Also available as IBM Research Report RC15020.

[Selman and Kautz, 1989] Bart Selman and Henry A. Kautz. Model-preference default theories. Technical Report KRR-TR-89-7, Dept. of Computer Science, Univ. of Toronto, Toronto, Ontario, Canada M5S 1A4, Apr 1989.

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

[Zadrozny, 1988] Wlodek Zadrozny. A three-level theory of reasoning. Unpublished, IBM T.J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598., 1988.