Representing and Reasoning with Defaults For Learning Agents

Benjamin N. Grosof
IBM T. J. Watson Research Center
P.O. Box 704, Yorktown Heights, NY 10598
(914) 784-7100 ; Internet: grosof@watson.ibm.com

Abstract

The challenge we address is to create autonomous, inductively learning agents that exploit and modify a knowledge base. Our general approach, embodied in a continuing research program (joint with Stuart Russell), is *declarative bias*, i.e., to use declarative knowledge to constrain the hypothesis space in inductive learning. In previous work, we have shown that many kinds of declarative bias can be relatively efficiently represented and derived from background knowledge. We begin by observing that the default, i.e., revisable, flavor of beliefs is crucial in applications, especially for competence to improve incrementally and for information to be acquired through communication, language, and sensory perception in integrated agents. We argue that much of learning in humans consists of "learning in the small" and is nothing more nor less than acquiring new plausible premise beliefs. Thus representation of defaults and plausible knowledge should be a central question for researchers aiming to design sophisticated learning agents that exploit a knowledge base. We show that such applications pose several representational requirements that are unfamiliar to most in the machine learning community, and whose combination has not been previously addressed by the knowledge representation community. These include: prioritization-type precedence between defaults; updating with new defaults, not just new for-sure beliefs; explicit reasoning about adoption of defaults and precedence between defaults; and integration of defaults with probabilistic and statistical beliefs. We show how, for the first time, to achieve all of these requirements, at least partially, in one declarative formalism: Defeasible Axiomatized Policy Circumscription, a generalized variant of circumscription.

1 INTRODUCTION; DECLARATIVE BIAS

Guide to Reader: In the main, this paper is at a fairly high, survey level; we discuss the importance and the integration of a number of representational issues, rather than discussing any one in full technical detail. Thus the form of this paper is: observations, plus a set of pointers to other papers that elaborate on various aspects. If you are already familiar with our idea of declarative bias, its motivations, and its non-monotonic aspects in our previous work, you can skip to section 4 without much loss, using just the abstract as your introduction. Sections 1.1, 2, 3, and 6 are drawn from previously published work; the rest of this paper is new.

The challenge we address is to create autonomous, inductively learning agents that exploit and modify a knowledge base. We are especially concerned in this paper with *competence* (what the program is able to infer, unlimited by computation time)[1], as opposed to pure speed-up (how fast the program can infer), learning. By "inductive", we mean involving inductive leaps that go beyond what is known for-sure and posit a falsifiable hypothesis. By "autonomous", we mean that manually-programmed bias is not required for each new inductive learning problem, and, more generally, programmer intervention / interaction is minimized.

In our view, the greatest difference in inductive learning (competence) capability between humans and cur-

[1]What we are calling "competence" is very similar to "the knowledge level"; but some, e.g., Dietterich [1986], define the knowledge level as monotonic. Many have also leveled the criticism that organization and form that makes for more computational efficiency is also a kind of knowledge, albeit more "heuristic" in flavor than "epistemological", in the terms of [McCarthy and Hayes, 1969]. (We sympathize with this criticism.) We therefore use a more neutral term, common in psychology.

rent automatic methods is that humans are able to bring to bear a great deal of background knowledge to formulate and decompose their learning problems, and to reduce the space of hypotheses for each learning problem. This saves in example complexity and computational complexity.

Our general approach, embodied in a continuing research program (joint with Stuart Russell), is *declarative bias*. In the declarative bias approach, the system's current knowledge base serves not only for performance and prediction (i.e., non-learning activities), but also to identify and constrain the instance language and the hypothesis space that the inductive component of a learning agent must search. This background knowledge is represented declaratively. By "declarative" in this paper, we mean formulable, more or less, in terms of classical logic or its extensions that, for example, handle default reasoning and probabilistic reasoning.

In previous work [Russell and Grosof, 1987] [Grosof and Russell, 1990] [Russell and Grosof, 1990] [Grosof and Russell, 1989], we have shown that many kinds of declarative bias can be relatively efficiently represented and derived from background knowledge that embodies a partial domain theory. A key kind of background knowledge that implies bias is about *relevant features*. Knowledge about such relevancy can be captured in the classical logical form of *determinations*. E.g., size, speed and weaponry determine whether or not an animal is a predator.

1.1 DECLARATIVISM

The declarative approach to AI is just one broad methodological approach among several in AI. Briefly, in it, one views the sequence of transformations that a program performs on its data structures as inference steps in a logical system. The logical sentences involved in these steps are the attributed meanings of the corresponding data structures.

We are attracted by the general advantages of the declarative approach. Some of those advantages are that declarativism:

- cleanly separates beliefs and premises from inferential processing;
- helps unify and integrate both procedures and data across systems;
- aids concise communication and understanding;
- provides semantics — therefore, a check on intuition, and formal guarantees enabling partial validation;
- facilitates incremental modifiability of programs, due to the existence of standard implemented inference systems and re-usable canonically-interpreted knowledge bases. One may update a program with new axioms, or with new inference mechanisms.

(For some more discussion of the declarative approach in AI and its advantages, see [Genesereth and Nilsson, 1987] (chapter 1) [Nilsson, 1983] [Doyle, 1985] [McCarthy, 1987].)

Declarativism has not heretofore been greatly accepted in machine learning; its impact has been rather decidedly less than in many other areas of AI. We suspect that much of the reason is socio-cultural: many researchers in the machine learning community do not feel comfortable with logic, just as many in the AI community do not feel comfortable with probability and statistics. Explanation-based generalization (EBG) [Mitchell *et al.*, 1986] can be viewed, however, as an early, relatively simple, successful penetration of declarativism to the realm of machine learning. We aim to extend the scope (and, hopefully, the successes!) of declarativism's use for learning.

One specific advantage of our declarative approach is that it facilitates combining the results of learning with other kinds of background knowledge. This enables "prior" knowledge (i.e., background knowledge that is present prior to a particular, given learning problem) to be accumulated, in part, from past learning. Thus as the agent progresses in time, it *modifies its own bias*. A positive feedback loop is created for learning. More generally, the declarative approach makes it easier for learning programs and their designers to take advantage of declaratively-formulated knowledge and inference methods developed for other purposes in the AI community. Part of our aim is agents that integrate several kinds of learning, both inductive and non-inductive (e.g., EBG-type speed-up), along with non-learning activities, using a common knowledge base.

1.2 SUMMARY

In this paper, we identify several representational issues that arise in this declarative approach to knowledge-based, autonomous, inductively learning agents, and give some partial solutions. Rafting along the main stream of knowledge representation in AI, we concentrate here mostly, but not entirely, on the making by the program (*learning agent*) of true-false distinctions, skirting the complexities of probability and similar kinds of uncertainty. (But see sections 7 and 8.)

We begin by observing that the *default*, i.e., revisable, flavor of beliefs is crucial in inductive learning applications, especially for competence to improve incrementally and for information to be acquired through communication, language, and sensory perception in integrated agents. We argue that much of learning in humans consists of "learning in the small" and is nothing more nor less than acquiring new plausible premise beliefs. Thus: *Representation of defaults and plausible knowledge should be a central question for researchers aiming to design sophisticated, inductively learning agents that exploit a knowledge base.*

We show that such applications pose several *representational requirements that are unfamiliar to most in the machine learning community, and whose combination has not been previously addressed by the knowledge representation community.* These include: prioritization-type *precedence* between defaults; updating with new defaults, not just new for-sure beliefs; explicit reasoning about adoption of defaults and precedence between defaults; and integration of defaults with probabilistic and statistical beliefs.

To reap the full advantages of a declarative approach requires a strong semantics: so that reasoning steps can be described as inferences that are semantically valid with respect to the logical system. We show how, for the first time, to achieve all of the above requirements, at least partially, in one declarative formalism: Defeasible Axiomatized Policy (DAP) Circumscription, a generalized variant of circumscription [McCarthy, 1980] [McCarthy, 1986] [Lifschitz, 1984]. DAP circumscription extends the capabilities of the formalism that we previously developed, the Circumscriptive Language of Defaults (CLD) [Grosof and Russell, 1990], a tool for specifying prioritized circumscriptions [McCarthy, 1986] [Lifschitz, 1985] [Grosof, 1991]. Prioritized circumscription and CLD are studied in [Grosof, 1992c]. CLD is just a syntactically sweetened special case of DAP circumscription.

2 INDUCTIVE LEARNING IS NON-MONOTONIC REASONING

In [Grosof and Russell, 1990] (elaborating the second half of [Russell and Grosof, 1987]), we analyzed inductive leaps, and the shifts of bias underlying them (e.g., cf. Utgoff's STABB [1984] [1986]), as logically non-monotonic from the declarative viewpoint. The following paragraph from [Grosof and Russell, 1990] (sub-section 11.1) recapitulates that argument.

One of the most fundamental characteristics of empirical inductive learning, e.g. the process of experimentally-grounded science, is that the agent must be prepared to retract working hypotheses as it accumulates new observations (or otherwise acquires information about which it is highly confident). When we ascribe a declarative status to these working hypotheses as sentences that the agent believes about the external world, then the agent's believed set of sentences evolves, in general, *non-monotonically*. If an agent is completely sure of its initial bias, no "inductive leap" is required to reach a definition for the target concept. In such a case, though the agent's starting (bias) theory may be incomplete, once the agent has observed enough instances, the concept's definition follows as a purely deductive conclusion from the bias plus the observations. From the viewpoint of logical semantics, the change in the agent's knowledge at any point in the learning process is no more than the information in the instances already observed. All the agent's predictions are correct. Additional observations simply narrow the zone of instance space where the agent is predictively uncertain. By contrast, in the most interesting kinds of empirical learning, the agent risks error in its predictions; it adopts beliefs that are *semantically novel*, i.e., that are not entailed simply by the knowledge in its initial certain beliefs plus its observations. For an agent to manifest such potential for retraction, for it to be capable of drawing conclusions that are merely plausible rather than completely certain, means that its inductive learning must be treated as a process of *non-monotonic* inference.

3 USEFULNESS OF DEFAULTS FOR REPRESENTING SHIFT-ABLE VERSION-SPACE-TYPE BIAS IN INDUCTIVE RULE AND CONCEPT LEARNING

In [Grosof and Russell, 1990], we showed how to derive many kinds of inductive leaps, and the shifts of bias underlying them, from *default* background knowledge, i.e, as deductions (formally, entailments) in a non-monotonic logic of prioritized defaults, based on circumscription (CLD, discussed below). In particular, we demonstrated that "version space" bias can be represented, via default (first-order) formulas, in such a way that it will be weakened when contradicted by observations. This built on a first-order formulation [Russell and Grosof, 1987] of the biases in the Version Space method [Mitchell, 1978]. Implementation of inference in the non-monotonic logic then enables the principled, automatic modification of the description space employed in a concept learning program, which Mitchell [1982] remarked "would represent a significant advance in this field", and which Bundy *et al.* [1985] named as "the most urgent problem facing automatic learning".

We also showed how to formulate with prioritized defaults two kinds of "preference" biases previously regarded as "syntactic" or "symbol-level" (e.g. by [Dietterich, 1986]): maximal specificity and maximal generality.

In [Russell and Grosof, 1990] and [Grosof and Russell, 1989], we showed that our approach was adequate to represent and derive almost all of the bias employed in Meta-DENDRAL [Mitchell, 1978], one of the most substantial inductive concept learning programs in the literature to date.

4 MORE GENERAL ROLE FOR DEFAULTS IN LEARNING AGENTS: REPRESENTING KNOWLEDGE INPUTS AND ERROR-CORRECTION

A Broader, More Ambitious, Picture of a Learning Agent:
Picture an agent that learns as it goes, while acquiring information from various sources, especially perception. The current knowledge base feeds learning, and learning modifies that knowledge base. The agent picks its learning problems (e.g., to maximally improve some measure of performance) and modifies its own bias(es) as it goes, while it improves its competence, making errors then later correcting them. Much of humans' competence learning has, as an information source, communication and language use. We expect future highly competent integrated learning agents to oftentimes have these properties.

So, how to get closer to this kind of *correcting, sophisticated learning agent*? We believe that the declarative approach holds promise. Next, we discuss a number of representational issues in pursuing it. (For convenience, we will speak of this aimed-at sophisticated learning agent in the present tense; this is, naturally, to be interpreted by you as really subjunctive or in the future.)

Defaults Are Needed For Sophisticated Learning Agents, e.g., Their Knowledge Inputs:
We begin by arguing for the thesis that: Defaults are not just useful, but very much needed, in a declarative approach to sophisticated learning agents. This goes beyond their ability to represent shift-able version-space of the kind we investigated previously. The ability to correct errors means that the agent's reasoning is non-monotonic, when viewed declaratively (by an argument very similar to section 2). Defaults are the most widely accepted declarative tool for describing (e.g., specifying) non-monotonic belief and reasoning. They are "bite-size" chunks of information, similar to a well-formed formula in first-order logic. Almost every non-monotonic logical formalism has some analogue of a default.

Many inputs to a sophisticated learning agent's knowledge base, when viewed declaratively, take the form of defaults:

- Initial background knowledge entered by a designer, may employ defaults, e.g., default inheritance. Many kinds of background knowledge need defaults to represent them: e.g., about commonsense or prototypes.
- More fundamentally, reasoning about the effects of actions, including frame assumptions, is non-monotonic; to express knowledge concisely in this realm needs defaults.
- Empirical information, whether or not arrived at by the agent's own inductions, is typically revisable, i.e., plausible, in status.
- Defaults are needed to faithfully represent input knowledge gleaned from natural language and communication (e.g., reading), either by being told or by observing. For example, a word sense or anaphoric (pronoun) reference must be guessed at; that guess is essentially a default. Speech acts also involve default reasoning. [Perrault, 1987] [Appelt and Konolige, 1988] [McRoy, 1989] [Morgenstern, 1990] [Morgenstern and Guerriro, 1991] [Zadrozny, 1988].
- Defaults are also needed to faithfully represent interpreted sensory perceptions, e.g., that $Ball61$ is on $Table302$. These perceptions may be errorful; and the situation that the perceptions describe may change. One can view the perception as having had revisable status, or else the reasoning based on it as having had revisable status; either way, defaults are needed. (One may reply that perceptions are better represented via probabilistic-type uncertainty than by a definitive true-false belief. However, even that does not, essentially, remove the need for default reasoning: see section 7.)

Representing Preference-Type Bias:
A default p can be viewed as a preference for believing p over believing $\neg p$. One can thus view a default as a preference-type belief. As such, it is able to represent some preference-type inductive biases, i.e., preferences among inductive hypotheses. By contrast, restriction-type bias (on a hypothesis space) corresponds to for-sure (i.e., purely monotonic) beliefs.

Match Between Defaults and Errorful Competence-Level Learning:
To summarize, then, there is a good qualitative match between defaults and the kind of knowledge-based competence learning that involves the correction of errors, especially in sophisticated agents. Indeed, one can view the evolution under updating of any expressive default theory (or, more generally, non-monotonic theory) as the evolution of competence with correction; i.e., one can view it as competence learning with correction.

Principles for Learning New Defaults:
Two oft-asked questions about defaults, both in the non-monotonic reasoning community, and in the machine learning community, are: "What are the normative principles for learning them?"; and: "Why should we use them if we don't know how they arise epistemologically?". These are interesting and legitimate

concerns. However, we do not believe they should be obstacles to using defaults for learning agents. We point out, firstly, that defaults are useful as input to a learning agent even if they are not being learned "from scratch" via induction. For example, they are useful, and, we argued above, crucial, to represent ongoing input information from language and communication. Secondly, we point out that the normative principles for learning "for-sure" beliefs are far from a settled question. Indeed, there are deep philosophical objections to the learnability of a universally-quantified statement over an infinite domain. A case can be made that inductive learnability of a default is more justifiable than inductive learnability of a for-sure empirical statement. Thirdly, we point out that a general normative basis for adopting a default, like for adopting any kind of belief, may be found in decision theory applied to some evaluation criterion over the learner's behavior.

Why Be Declarative About Defaults?:
Another question / criticism in many researchers' minds, is: "Why bother with declarative formalism for defaults and non-monotonic reasoning"? E.g., "Why can't I just treat default reasoning as simply the replacement, under direct updating, of the truth value (say: true, false, or unknown) associated with a data structure representing a statement p?" I.e., to do what is the current least-common-denominator style of default reasoning in practical systems, e.g., in AI frame-based systems with default inheritance.

One good reason is that an update (knowledge input) may not be directly about the same proposition(s) as previous beliefs. E.g., when an agent believes p and q and $p \wedge q \supset r$, and then an update asserts $\neg r$. This cannot be handled by the direct-replacement method alluded to above. In general, there are subtleties and trickinesses in default reasoning. It is easy to get global inconsistency if one is not careful. Declarative formalisms for default reasoning bring with them an evolving body of understanding about ways to manage belief change, and conflict between beliefs, in a rationalized fashion.

A second good reason is that incorporating (assimilating, integrating) new input knowledge is aided by non-monotonic logical systems. They distinguish premises (axioms) from conclusions, and eliminate the need to revise premises. When this distinction is not made, previously acquired default rules may have to be modified immediately with qualifications in order to deal with interactions, to preserve consistency. By contrast, in a non-monotonic logical system, the ramifications of new update knowledge can be worked out later. The working theory is at risk of inconsistency with that update (i.e., with the new global set of premises) until the ramifications are worked out, but a kernel basis for the belief revision is always present, in the meantime: the premises.

Another style of non-declarative treatment of defaults is to do highly limited, incomplete consistency checking, e.g., in a negation by failure to prove. This method produces unsoundness, and, often, global inconsistency.

5 IMPORTANCE OF PRIORITIZATION-TYPE PRECEDENCE

Conflict:
In general, a set of two or more defaults may *conflict*, in the sense that they contradict each other, i.e., cannot consistently all "go through", given the for-sure beliefs. A classic example is Quakers-and-Republicans. One default says that, typically, Quakers are Pacifists. Another default says that, typically, Republicans are not Pacifists. For-sure, Nixon is a Quaker and a Republican. The two defaults conflict about Nixon (more precisely, their Nixon instantiations conflict). Neither the conclusion that Nixon is a Pacifist, nor the conclusion that Nixon is not a Pacifist, is sanctioned (in the sense of skeptical / cautious entailment).

Precedence:
An important concept in representation of defaults is prioritization-type *precedence*. [2] Precedence is a kind of information that may be specified about defaults. For one default to have precedence over a second means that in cases of conflict between them, the first "wins": its conclusion goes through. E.g., if the Republican default has precedence over the Quaker default, then the conclusion that Nixon is not a Pacifist is sanctioned. Precedence resolves the conflict.

Precedence can be viewed as a kind of confidence information about a default belief.

Bases for Precedence:

1. One well-known kind of precedence is the specificity dominance principle employed in inheritance with exceptions, e.g., cf. [Touretzky, 1986]. To represent default inheritance requires precedence for more specific rules over less specific rules. E.g., the more specific default rule that the New York subway does not run when there is a electrical power blackout takes precedence over the default rule that the New York subway runs at any time.

However, we showed previously [Grosof, 1991] that there are several bases for precedence information

[2] By "prioritization" here, we mean the concept of prioritization developed in [Grosof, 1991] and studied in [Grosof, 1992c], that generalizes the concept of prioritization in circumscription developed by [McCarthy, 1986] and [Lifschitz, 1985] and many other kinds of precedence developed by others in non-monotonic reasoning.

other than specificity.

Need for Non-Layered Precedence:
We also showed there that, in most useful default theories the partial order of precedence among the global set of defaults does *not* obey the condition of being *stratified* or *layered*. (By "stratified" or "layered", we mean isomorphic to the system of military rank: generals are all higher-precedence than colonels, who are all higher than majors, etc.; there is no precedence between two members of the same rank.) This goes even for simple default inheritance theories with precedence only on the basis of specificity. Most AI researchers who do know about precedence are familiar only with this layered case of precedence, to which the original definition of prioritization in [Lifschitz, 1985] was restricted.

Need for Non-Specificity Precedence in Sophisticated Learning Agents:
We observe here that several kinds of precedence, other than specificity, are especially important for representing defaults in sophisticated learning agents:

2. *Reliability* of sources. A source might be a communicating agent. E.g., from a child's point of view, "things my mother told me" have precedence over "things my brother told me", which in turn have precedence over "things the school bully told me". Or a source might be a sensor.
3. *Temporal precedence*, which is closely related to reliability. For example, fresher reports are regarded as more reliable. We observed in the last section that to represent over-ride-able true-false input beliefs (e.g., percepts) in sophisticated learning agents requires them to be treated as defaults. But for a new input belief (e.g., $Likes(She, Me)$) to be part of the agent's current theory, i.e., to "go through" as a default, means that it must over-ride any temporally earlier information (e.g., $\neg Likes(She, Me)$) that conflicts with it. This needs precedence: for the new over the old.

   Another example is disambiguation of context-dependent meaning in natural language and communication. Often one must guess, only to have the guess corrected by a later clarification. Again, this needs precedence: for the new over the old.
4. *Authority* in the legal and organizational senses. For example, federal law takes precedence over state law; policy directives from the head of an organization take precedence over those from subordinates; holiday parking rules take precedence over weekday parking rules. Sophisticated agents will have to deal with such information.
5. To represent the combination, i.e., aggregation, of different *preference-type inductive biases*. For example, maximal specificity / maximal generality bias (see section 3) is typically applied after (most) other preference-type biases have affected ("had their say about") the hypothesis space in concept / rule learning. This inferential ordering, i.e., the application "after", needs precedence to represent it.
6. Difference in (probabilistically) expected utility of actions based on defaults. For example, an action policy oriented toward an emergency situation has precedence over one oriented towards a routine situation, since in the case when both apply (i.e., in an emergency), following the first policy but not the second has higher expected utility than vice versa. This basis for precedence is very related to the process of inducing defaults and precedences from statistical and utility information.

Need for Explicitness of Precedence:
In many formalisms (e.g., Touretzky's inheritance system and its successors), specificity precedence is implicit. However, we observe here that it is important that one be able to specify precedence *explicitly* in the declarative formalism: firstly, because there are so many different bases for it; and secondly, because it is necessary to make precedence explicit to be able to reason about it. For example, an agent may have the default that default information gleaned from a message that originates later should take precedence over default information gleaned from a message that originated earlier from the same source. But, often, the agent may know only the order of arrival of messages; the order of arrival is not known for-sure to be the order of origination. Rather, often, one applies the default that, typically, later arrival indicates later origination. We would like an agent to be able to engage in such basic kinds of reasoning about precedence, e.g., to reason by default about time-stamping of defaults. (See section 8 for more.)

6 A CIRCUMSCRIPTIVE LOGIC OF DEFAULTS

We have developed a logical formalism for default reasoning with explicit non-layered precedence: the Circumscriptive Logic of Defaults (CLD). CLD was developed specifically with learning applications in mind. CLD is a meta-language for specifying prioritized default or predicate circumscriptions. [Grosof and Russell, 1990] and [Grosof, 1991] defined CLD. [Grosof, 1992c] elaborates and studies CLD and the prioritized default theories that it specifies. CLD was the first non-monotonic formalism to represent explicit non-layered precedence as premise information. It also represents updating with new premise defaults, not just new for-sure premises, unlike previous variants of circumscription and unlike many non-monotonic formalisms.

In core CLD, there are three types of axioms (premises). A *base* axiom is prefixed by $\bullet\!>$ and specifies a for-sure premise belief (a closed first-order formula):

$$\bullet\!> \ Quaker(Nixon) \wedge Republican(Nixon)$$

A *default* axiom is prefixed by $:>$ and specifies a default premise belief (an open first-order formula). In addition, a default axiom may, optionally, have a label (e.g., $d1$), which is used as a name-tag to specify its precedence via prioritization axioms (see below):

$$(d1) :> \ Quaker(x) \supset Pacifist(x)$$

$$(d2) :> \ Republican(x) \supset \neg Pacifist(x)$$

A *prioritization* axiom specifies a pairwise strict precedence between two default axioms, via their labels. E.g.,

$$\mathcal{PREFER}(d2, d1)$$

says that the default with label $d2$ has higher precedence than the default with label $d1$.

A CLD axiom set $\mathcal{A}$ non-monotonically entails a CLD theory, which is the set of all conclusions entailed (model-theoretically, in second-order logic) by the prioritized default circumscription that is specified by that axiom set:

$$\mathcal{PDC}(\mathcal{A}) \stackrel{\text{def}}{\equiv} B[Z] \ \wedge \ \neg\exists Z'. \ B[Z'] \wedge Z \prec_{(D;R)} Z'$$

Here, Z is the tuple of all predicate (and function) symbols in the first-order language used to express the base and default formulas. B is the conjunction of the base formulas from $\mathcal{A}$. D is the tuple of the default formulas from $\mathcal{A}$. N is the index tuple of D. Dk indicates the k^{th} default formula in the tuple D. The precedence partial order R, defined over domain N, is the transitive closure of the pairwise comparisons specified in the prioritization axioms from $\mathcal{A}$. $R(j, i)$ means that the j^{th} default has higher precedence than the i^{th} default. $\leq$ stands for (universally quantified) implication. $=$ stands for (universally quantified) equivalence, i.e., $\leq \ \wedge \ \geq$. $\prec_{(D;R)}$ is defined as the strict version $(\preceq_{(D;R)} \wedge \neg\succeq_{(D;R)})$ of the prioritized "formula" preorder $\preceq_{(D;R)}$:

$$\begin{aligned} Z \preceq_{(D;R)} Z' \ &\stackrel{\text{def}}{\equiv} \\ \forall i \in N. \ [\forall j \in N. \ R(j,i) \supset \ & (\forall x. \ Dj[Z,x] \equiv Dj[Z',x])] \\ & \supset \ (\forall x. \ Di[Z,x] \supset Di[Z',x]) \end{aligned}$$

Here Dj and Di refer to the j^{th} and i^{th} members, respectively, of the tuple D.[3] We **define** the corresponding circumscriptive prioritized default *theory* as the set of all conclusions entailed (model-theoretically, in second-order logic) by the prioritized default circumscription.

CLD extends in several directions. One is to express for-sure (base) and default beliefs in higher-order logic, not just first-order. Another is to specify the fixing of function and predicate symbols. These extensions are straightforward (see [Grosof, 1992c]).

More interestingly for our purposes here is that CLD extends to explicit, non-monotonic (default) reasoning about precedence (prioritization) among, and adoption of, defaults: see section 8. And CLD can represent default reasoning about probabilistic beliefs, as we show next.

7 INTEGRATION WITH PROBABILISTIC REASONING

Need for Integration of Probabilities and Statistics:
Probabilistic-type uncertainty is a vital aspect of the representation of plausible knowledge. For sophisticated learning agents, it is imperative to integrate probabilistic and statistical reasoning with true-false reasoning, including default true-false reasoning. For our declarative approach, that implies the requirement of integration in the declarative formalism. For example, a sophisticated learning agent must be able to reason about inducing the adoption (and precedence) of a default on the basis of probabilities that were, in turn, induced from a statistical history. This requires explicit reasoning about adoption (and precedence) of defaults, as well.

Monotonic Probabilistic Reasoning:
Monotonic reasoning about probabilities can be integrated with monotonic true-false reasoning by expressing probability statements in classical logic (e.g., cf. Probabilistic Logic [Nilsson, 1986]; extended to conditional probabilities in [Grosof, 1986]) or modal logic (e.g., cf. [Halpern, 1990]). Halpern's modal approach enables the reasoning about probabilities to be essentially first-order, rather than essentially propositional as in Probabilistic Logic.

Probabilistic and Statistical Reasoning Is Usually Non-Monotonic:
However (as we were the first to observe), probabilistic reasoning in practice is typically logically non-monotonic; indeed (as we first showed), the basic kind of Bayesian updating is equivalent to inheritance-type prioritized default reasoning about probabilities.[4] In [Grosof, 1988], we showed that prioritized default

[3] For notational simplicity, we ignore the potentially different arities of the various open formulas Di.

[4] The non-monotonic quality was much earlier observed by Henry Kyburg, in quite different terms, however, in his work on reference classes.

circumscription is able to represent the basic case of Bayesian updating. This built on Probabilistic Logic. The defaults are conditional independence assumptions.

Since then, Bacchus [1990] has shown how to extend our method of representing Bayesian reasoning so as represent first-order reasoning about probabilities, and some kinds of statistical information. This built on Halpern's above-mentioned logic; he defined his own new non-monotonic formalism for reasoning about the probabilities and statistics; however, the essence of the non-monotonic aspect in his system is, like ours, inheritance-type default conditional independence assumptions with specificity dominance-type precedence.

CLD Can Represent Basic Integration:
We observe here that CLD is able to express the prioritized default theories we used to represent Bayesian reasoning. CLD is thus able to integrate basic Bayesian probabilistic reasoning with true-false prioritized default reasoning.

Probabilities Cannot Replace Defaults:
We observe here, secondly, that our above-described results imply that probabilities cannot replace defaults, as some have argued that it should or might. Probabilistic / statistical reasoning is itself non-monotonic; the most common subset involves, essentially, default reasoning (with precedence) about the probabilities. Non-monotonicity and probabilistic-type uncertainty should be viewed as two different dimensions of knowledge representation and reasoning.

8 EXPLICIT REASONING ABOUT PRECEDENCE AND ADOPTION

Earlier, we discussed two needs for explicit reasoning about precedence among, and adoption of, defaults in sophisticated learning agents. One is to reason, by default, about precedence, e.g., for source reliability based on presumed temporal precedence. Another is to reason about induction of new defaults from probabilistic / statistical history.

We report here that we have developed, for the first time, a method to represent such explicit, non-monotonic (i.e., default) reasoning about precedence and adoption: via a new formalism, Defeasible Axiomatized Policy (DAP) circumscription, that generalizes CLD and prioritized predicate / default circumscription.

8.1 DEFEASIBLE AXIOMATIZED POLICY (DAP) CIRCUMSCRIPTION

DAP circumscription 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 defeasible reasoning about prioritization, *without* resorting to a more powerful logical language; i.e., without promiscuously proliferating any additional meta-languages to represent these meta-levels. We have shown that this representational generalization 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.)

DAP circumscription is (defined as) a special case of general circumscription cf. [Lifschitz, 1984]. However, DAP circumscription's expressive power is fundamentally greater than prioritized *predicate* circumscription, which is the kind almost all previous work on circumscription has studied. DAP circumscription includes CLD and prioritized predicate circumscription as special cases. Thus, like them, it is able to represent basic integration of Bayesian probabilistic reasoning with true-false prioritized default reasoning.

Unfortunately, we do not have space here for the full technical details of these results about DAP circumscription (nor even for its definition, which is a bit hairy); they are given in [Grosof, 1992a].

9 FUTURE WORK: MORE REPRESENTATIONAL CHALLENGES

There are a number of further challenges in representing and reasoning with defaults for sophisticated learning agents, that we plan to address in future work.

The computational complexity of default reasoning, and, especially, of belief revision in default theories: The complexity results currently available tell us that default reasoning is significantly harder than monotonic reasoning, for reasonably expressively rich cases of most non-monotonic formalisms of interest. For example, even for query-answering in propositional default theories without priorities, current results show worst-case is **exponential** (NP-hard) [Selman and Kautz, 1989] [Kautz and Selman, 1989] [Selman and Levesque, 1989]. For current expressively rich default formalisms (e.g., circumscription, Default Logic, Autoepistemic Logic), the only procedures available for forward inference and belief revision are *exhaustive* and thus impractical, especially for creating and maintaining a working body of conclusions. These procedures compute the entire non-monotonic theory, then re-compute from scratch upon updating with new axioms (e.g., some new premises arriving from senses or communication). In [Grosof, 1992c] and [Grosof, 1992b], we address this problem, and give theoretical results for prioritized circumscription that show how to reformulate default theories so as to "divide and conquer": enabling selective forward inference and restricted (yet sound) belief revision.

The non-monotonic normative principles guiding more complex kinds of probabilistic and statistical reasoning, beyond basic Bayesian updating: In the probability and statistics literature (e.g., see [Loui, 1988], which builds on Kyburg's earlier work), a key problem is how to select a "reference class" when specificity dominance is not present.

For example, suppose one is interested in whether cars are lemons, in the sense that they will break down in their first 5000 miles, and suppose one has (as the only information relevant to lemon-ness) statistics on lemon-ness for all American cars made by Ford (say, frequency 0.1). Suppose one is looking at a particular American car (let us call it Betsy) that is a Ford, was made on a Monday, is a blue Taurus station wagon with serial number 2346791, has a ding in the front left fender, and so on). Then in this basic case of Bayesian updating, the usually-followed normative principle is to conclude that the probability that this car (Betsy) is a lemon is 0.1. That this is a default conclusion (see section 7) rather than a monotonic conclusion is illustrated by the following: suppose one receives the additional information that the statistical frequency of lemon-ness for the class of Ford Tauruses is 0.05; then the usually-followed normative principle is to conclude that the probability that Betsy is a lemon is 0.05, justified by choosing the reference class to be the more specific class of Ford Tauruses, rather than the less specific class of all Fords. The previous conclusion that the probability of Betsy being a lemon is 0.1 has been retracted: it was defeasible.

Now suppose one also receives the additional information that the statistical frequency of lemon-ness for American cars made on Mondays is 0.2. Should one choose the reference class of Fords Tauruses, leading to the (default) conclusion that Betsy has probability 0.05 of being a lemon? Or should one instead choose the reference class of Monday-mades, leading to the conclusion that Betsy has probability 0.2 of being a lemon? Perhaps one should compromise on some value between 0.05 and 0.2? The above usual normative principle does not apply: neither class (Ford Tauruses, Monday-mades) is more specific than the other. What normative principles should guide such choices of reference class are currently unclear. Moreover, how should the sample sizes and confidence intervals associated with the statistics affect matters? E.g., suppose one class has only a small sample available.

How to do explanation-based learning from default inferences: A default conclusion is justified not only by what rules and facts appeared positively in its proof, but also by the lack (omission) of any conflicting overriding rules and facts. This makes explanation-based generalization very much trickier than in the case of monotonic reasoning: there may be a huge or infinite space of such omission conditions.

Conditionality of defaults, in the sense of distinguishing asymmetrically between antecedent and consequent in default rules, e.g., in the manner of a conditional probability: We neglected that issue in this paper, but it appears important, especially in learning sets of defeasible rules from statistics, where rules with overlappingly applicable antecedents may have conflicting consequents.

10 CONCLUSIONS

Even without declarative formalism, the representational issues we have raised here are important: designers of correcting learning agents should keep them in mind. The subject of non-monotonic reasoning deserves much more attention from the machine learning community; it is peculiarly important for knowledge-based learning that involves inductive leaps and correction.

Summary: see sub-section 1.2 and/or the abstract.

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.

[Bundy *et al.*, 1985] Alan Bundy, Bernard Silver, and Dave Plummer. An analytic comparison of some rule-learning programs. *Artificial Intelligence*, 27, 1985.

[Dietterich, 1986] Thomas G. Dietterich. Learning at the knowledge level. *Machine Learning*, 1(3), 1986.

[Doyle, 1985] J. Doyle. Expert systems and the "myth" of symbolic reasoning. *IEEE Transactions on Software Engineering*, 11:1386–1390, 1985.

[Genesereth and Nilsson, 1987] M. Genesereth and N. Nilsson. *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann, Los Altos, CA., 1987.

[Grosof and Russell, 1989] Benjamin N. Grosof and Stuart J. Russell. Declarative bias for structural domains. In Alberto M. Segre, editor, *Proceedings of the Sixth International Workshop on Machine Learning*, San Mateo, CA, 1989. Morgan Kaufmann. Held Cornell University, Ithaca, NY. Also available as IBM Research report RC 14614.

[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, 1986] Benjamin N. Grosof. An inequality paradigm for probabilistic knowledge. In John F. Lemmer and Laveen Kanal, editors, *Uncertainty in Artificial Intelligence (1)*, pages 259–275, Amsterdam, 1986. North-Holland. Volume containing revised versions of papers appearing in the Proceedings of the AAAI/RCA First International Workshop on Uncertainty in Artificial Intelligence, held Los Angeles, CA, Aug. 1985.

[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, 1992a] Benjamin N. Grosof. Defeasible and pointwise prioritization. Submitted., 1992.

[Grosof, 1992b] Benjamin N. Grosof. Reformulating non-monotonic theories for inference and updating. In Michael Lowry, editor, *Proceedings of the 1992 Workshop on Change of Representation and Problem Reformulation*, Moffett Field, California, April 1992. NASA Ames Research Center Technical Report. Workshop held Asilomar, California, April 28 – May 1, 1992. Also available as IBM Research Report RC17955.

[Grosof, 1992c] Benjamin N. Grosof. *Updating and Structure in Non-Monotonic Theories*. PhD thesis, Computer Science Dept., Stanford University, Stanford, California 94305, 1992.

[Halpern, 1990] Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 1990.

[Kautz and Selman, 1989] H. Kautz and B. Selman. Hard problems for simple default logics. *Proceedings of the First International Conference on Principle of Knowledge Representation and Reasoning*, pages 189–197, Toronto, Ontario, 1989.

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

[Loui, 1988] Ronald P. Loui. Non-monotonicity in probabilistic reasoning. In J. Lemmer and L. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 273–290. 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.

[McCarthy and Hayes, 1969] J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Mitchie, editors, *Machine Intelligence 4*, pages 463–502. American Elsevier, New York, 1969.

[McCarthy, 1980] J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

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

[McCarthy, 1987] J. McCarthy. Generality in artificial intelligence. *Communications of the ACM*, 30, 1987.

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

[Mitchell *et al.*, 1986] T.M. Mitchell, R.M. Keller, and S.T. Kedar-Cabelli. Explanation-based generalization: A unifying view. *Machine Learning*, 1:47–80, 1986.

[Mitchell, 1978] Tom M. Mitchell. *Version Spaces: An Approach to Concept Learning*. PhD thesis, Computer Science Dept., Stanford University, Stanford, California 94305, 1978.

[Mitchell, 1982] Tom M. Mitchell. Generalization as search. *Artificial Intelligence*, 18(2):203–226, 1982.

[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 Mateo, California, 1990. Morgan Kaufmann.

[Nilsson, 1983] Nils J. Nilsson. Ai prepares for 2001. *AI Magazine*, 4(4), 1983. Based on AAAI-83 Presidential Address.

[Nilsson, 1986] Nils J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28(1):71–88, Feb 1986.

[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.

[Russell and Grosof, 1987] Stuart J. Russell and Benjamin N. Grosof. A declarative approach to bias in concept learning. *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, San Mateo, CA, 1987. Morgan Kaufmann. Held Seattle, Washington.

[Russell and Grosof, 1990] 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.

[Selman and Kautz, 1989] B. Selman and H. Kautz. The complexity of model preference default theories. In M. Reinfrank *et al.*, editor, *Proceedings of the Second International Workshop on Non-Monotonic Reasoning*, pages 115–130, Berlin, Germany, 1989. Springer Lecture Notes on Computer Science.

[Selman and Levesque, 1989] B. Selman and H. Levesque. The tractability of path-based inheritance. *Proceedings IJCAI-89*, pages 1140–1145, Detroit, MI., 1989.

[Touretzky, 1986] D. Touretzky. *The Mathematics of Inheritance Systems*. Pitman, London, 1986.

[Utgoff, 1984] Paul E. Utgoff. *Shift of Bias for Inductive Concept Learning*. PhD thesis, Rutgers University, New Brunswick, NJ, 1984.

[Utgoff, 1986] Paul E. Utgoff. Shift of bias for inductive concept learning. In J.G. Carbonell, R. Michalski, and T. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach, Volume II*, San Mateo, CA, 1986. Morgan Kaufmann.

[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.