Applying Theory Formation to the Planning Problem

Randy G. Goebel
Scott D. Goodwin

Logic Programming and AI Group
Department of Computer Science
University of Waterloo
Waterloo, Ontario
N2L 3G1

Research Report CS-87-02
January 1987

Applying Theory Formation to the Planning Problem

Randy G. Goebel
Scott D. Goodwin

Logic Programming and AI Group
Department of Computer Science
University of Waterloo
Waterloo, Ontario, CANADA N2L 3G1
rggoebel@waterloo.CSNET
sdgoodwin@waterloo.CSNET

Abstract

There are at least three aspects to the problem of generating plans. The problem of specifying how relations persist over time has been named the "frame problem," in comparison to motion picture film where most things do not change from frame to frame. Closely related is the "qualification problem," which is that of specifying the conditions underwhich events or actions can actually take place. In addition, the "prediction problem" is that of making and verifying predictions, then somehow revising potentially incorrect specifications.

In the frame and qualification cases, the problem is not just that of specifying preconditions and effects of actions but that of drawing conclusions about future states, based on the available specifications. As it seems that *any* specification of action preconditions and effects can only approximate the world they model, certain assumptions are required in order to predict which relations will hold in future states. Therefore, both problems require one to make appropriate assumptions about action preconditions and effects, then use those assumptions to predict future states.

Here we cast this problem of reasoning from assumptions in a general framework, and propose that current non-monotonic reasoning systems can be usefully compared along a continuum defined in terms of how assumptions are expressed and interpreted. At one end of the proposed continuum, sets of plausible assumptions are ordered by explicitly rendered preference heuristics. At the other end, preferred assumptions are expressed as part of the precondition and effect specifications, so that preferences are "compiled in," rather than explicit.

We explain the genesis of this continuum thesis, provide some evidence for its existence, and explain results that are necessary for its support. We conclude with an example that shows where defaults and theory formation are on this continuum, and compare it with other similar work on planning.

1. Introduction

In many logical representations (e.g., [Green69, McCarthy69, Hayes71, Kowalski79]), the dynamic world is modelled as a sequence of discrete states, each named by an individual term composed of a sequence of functions. For example, we might note that blocks a and b are on a table in state s, and that the action sequence $pickup(a)$, $stack{-}on(a,b)$ will stack up the blocks. As explained by Green [Green69], we can use a deductive theorem prover to synthesize this sequence of actions by expressing actions as functions from state to state. After expressing initial facts, action preconditions, and action effects as a set Γ of formulae, a constructive proof of an existential formula $\exists S.on(a,b,S)$ from Γ, viz.,

$$\Gamma \vdash \exists S.on(a,b,S)$$

will produce the name of a state in which a is on b, e.g.,

$$S = stack{-}on(a,b,pickup(a,s)).$$

The difference between "$pickup(a)$" and "$pickup(a,s)$" indicates how actions are viewed as functions from state to state, so that the composite function term $stack{-}on(a,b,pickup(a,s))$, evaluated in applicative order, will *apparently* transform the world state named by s into one in which $on(a,b)$ is true.

The problem remains, however, that the world is its own best representation. Without the guarantee of a complete and accurate axiomatization, there may be arbitrarily many reasons why a deduced sequence of actions might *not* produce the desired effect, or why a desired sequence of actions cannot be deduced (cf. [Hayes71, Hayes73]).

As an example of the former, we may suppose that some set of blocks world axioms Γ are true, but it may be that executing, say $stack{-}on(a,b,pickup(a,s))$ doesn't make $on(a,b)$ true because b turns out to be a pyramid. In other words, the *assumption* that our intended interpretation was a model of Γ was false.

In the latter case, there might simply be things that we wish to conclude but cannot. For example, we might expect, after executing $pick{-}up(a,s)$, that a is no longer on the table. We might not, however, anticipate the gluing of the block to the table so that we end up holding both the block and the table. Similarly, we might not anticipate a lot of other potential qualifications (cf. [McCarthy80]). One solution is to simply *assume* that the block will be off the table, unless we have some reason to believe otherwise (cf. [Fikes71]). Although plans synthesized under such assumptions are subject to error, one hopes that, by making assumptions explicit, they can be revised to produce improved plans.

When we make assumptions as part of a planning problem specification, there may be multiple ways in which such assumptions will support the desired goals. For example, if two different actions a and b both are assumed to accomplish the same goal and both are assumed to have their preconditions satisfied, we must prefer one over the other to actually accomplish the goal. In other words, we must somehow select the most reasonable set of assumptions that supports the action we want to apply.

Here we cast this problem of reasoning from assumptions in a very general framework, and propose that current non-monotonic reasoning systems can be usefully compared along a continuum defined in terms of how assumptions are expressed and interpreted. At one end of the proposed continuum, sets of plausible assumptions are ordered by explicitly rendered preference heuristics. At the other end, preferred assumptions are expressed as part of the precondition and effect specifications, so that preferences are "compiled in," rather than explicit.

In the following sections, we provide a general view of the planning problem. Some proposed solutions, together with their need for assumptions, are examined. The general need for approximate reasoning based on assumptions is addressed and the inevitable need for non-monotonic reasoning systems is explained. A method of comparing several existing systems is suggested before posing desired theorems of all such systems. Our favoured method of using assumptions in plan synthesis is illustrated and its relationship to other recent ideas is acknowledged. Finally, we present some tentative conclusions.

2. Temporal Representation and Reasoning

The problem of predicting the future arises in *any* axiomatization of a changing world. Of the many ways of conceiving of time in order to describe a changing world (e.g., points, intervals, points and intervals, before and after modalities),[1] any that permits variables to range over time points or intervals can express the question about whether or not certain relations hold at future times. As we have seen, in such representations any request for a deductively justified prediction *is* planning. For example, for any world description Γ, *propositional fluent* R [McCarthy69], time variable T, and constant *now* denoting the present, we can always pose a question like

$$\Gamma \vdash \exists T.R(T) \wedge T \geq now$$

As explained above, we can express change as functions of time so that a composition of functions substituted for T provides the name for a time at which the propositional fluent is a consequence of Γ.

2.1. The Frame Problem

Acknowledging some notion of changing time entails the need for specifying action preconditions and effects. The *frame problem* arises in any situation like the following: let P and Q be arbitrary relations, s a state, and a an action. Suppose we are given that

$$\Gamma = \{\forall X.P(a(X)) \subset Q(X),\ Q(s)\},$$

which we can read as "P will be true after action a, if Q is true before," and "Q is true in state s." We might expect, at least intuitively, to be able to ask if there is some state in which both P and Q are true, i.e., to ask whether

$$\Gamma \vdash \exists X.P(X) \wedge Q(X) \tag{1}$$

However, the relation expressed in (1) is false unless Γ includes the *frame axiom*

---

[1] See [Kowalski86b] for a nice discussion of various conceptions of time. See Ladkin [Ladkin86b, Ladkin86a] for an algebraic characterization of time intervals. Reichgelt provides an interesting first order axiomatization of the model theory of a modal logic [Reichgelt86].

$$\forall X.Q(a(X)) \subset Q(X) \tag{2}$$

which says that "if Q is true, then performing action a doesn't change Q." Adding formula (2) to Γ makes (1) true, however one such axiom will be required for every combination of action and relation.

As noted by others (e.g., [Kowalski86b, p. 11]), the frame problem has both an *epistemological* and *heuristic* component (cf. [McCarthy69, McCarthy77]). Its epistemological component is the difficulty of expressing the required collection of axioms in an intuitively straightforward way. Many proposals seek to express frame axioms in a conceptually economical way, and still use deduction to synthesize plans (e.g., [Kowalski79, McCarthy80, Lifschitz86a]). This requires a major shift in the way that one perceives the dynamic world. For example, Kowalski [Kowalski79, Kowalski86a, Kowalski86b] transforms a formula like $P(\bar{x},s)$, where P is a relation, $\bar{x}$ a tuple of individuals, and s a state, into $holds(P(\bar{x}),s)$. The semantic domain now consists of relations, functions, and world states, in addition to those things normally considered individuals. This perception offers a certain economy of expression, as we can write one general frame axiom, viz.,

$$\forall X,Y,Z.holds(X,do(Y,Z)) \subset holds(X,Z) \wedge preserves(Y,X) \tag{3}$$

and simply add, as facts, appropriate formulae defining the *preserves* relation, e.g., $\forall X.preserves(pickup(X),blue(X))$ says that picking up any X preserves the property of being blue.[2] Another interesting epistemological manoeuvre is that of Lifschitz [Lifschitz86a], who proposes that the frame problem be solved by explicitly defining a relation $causes(X,Y,Z)$, which can be read as "Action X causes relation Y to have truth value Z." With this perception of causation, one can then express the intent of formula (3) as

$$\forall X,Y,V. \neg \exists Z\; causes(X,Y,Z) \supset [holds(Y,do(X,V)) \equiv holds(Y,V)] \tag{4}$$

Or, in other words, if there is nothing that causes a relation to change truth value, then it must stay the same after an action.

In the case of both Kowalski and Lifschitz, assuming a *minimal extension* of the *preserves* and *causes* allows one to conclude that relations continue to hold unless otherwise stated. In an earlier formalization [Kowalski79], Kowalski provides this assumption as a implicit interpretation of the *preserves* relation [Kowalski79, p. 136-137]. In a more recent formalization based on intervals instead of states, he provides an explicit *persistence axiom* that uses negation-as-failure to minimize a *terminates* relation which asserts conditions underwhich relations no longer hold [Kowalski86b, p. 20]. Lifschitz uses circumscription [McCarthy80, McCarthy86, Lifschitz86a] to specify additional axioms that restricts one to the minimal interpretation of the *causes* relation.

Other approaches to the frame problem still propose further epistemological modifications, including an altered form of reasoning. We consider the proposals of Reiter [Reiter80] and Nute [Nute86] as examples in this category. In both approaches, the important epistemological change is to classify statements about action effects and

---

[2] Things are more complex than they appear here, as the meaning of variables inside of terms is not straightforward. See, for example [Bowen82, Bowen85a, Bowen85b].

preconditions into categories that distinguish facts from assumptions. In Reiter's terminology, we express frame axioms as *default rules* of the form:

$$\frac{r(\bar{X},Y):\mathbf{M}\,r(\bar{X},f(\bar{X},Y))}{r(\bar{X},f(\bar{X},Y))} \tag{5}$$

which can be read as "If relation r is true on tuple $\bar{X}$ in state Y, and it is *consistent* to believe that r is true on tuple $\bar{X}$ after applying action f, then one can infer that r is true on tuple $\bar{X}$ in the state $f(\bar{X},Y)$ which results from performing action f." Note that this is a default rule *schema* for which appropriate values of r and f must be provided to express the actual default rules. Most important is how such default rules are applied: instead of trying to establish a relation like (1) above, Reiter provides a new definition of $\vdash$, say $\vdash_D$, where instances of default rules like (5) are used to support a derivation *as long as they are consistent* with the known facts of Γ, viz.,

$$\Gamma \cup \Delta \vdash_D F$$

or formula F is "default derivable" from facts Γ and defaults Δ for some consistent subset of instances of Δ. This means that, in addition to modifying the more naive perception of the domain, the reasoning strategy that leads to a conclusion depends on the consistency of assumptions made along the way.

Nute's system of defeasible reasoning [Nute86] has a similar flavour to Reiter's in that the ability to express different kinds of facts (*absolute* and *defeasible*) leads to conclusions of varying strength. Nute's system is different from Reiter's in that assumptions are sentences of the formal language, rather than inference rules.

These brief caricatures of work by Kowalski, Lifschitz, Reiter, and Nute, are intended to suggest the existence of a trade off between the complexity of expression (cf. circumscription and deduction) and complexity of reasoning (cf. default rules and extended deduction).[3]

2.2. The Qualification Problem

Like the frame problem, the qualification problem arises in any situation where the specification of action effects and preconditions are used to predict future states of the world. The problem, however, is not that of somehow asserting an exasperating list of non-effects, but of somehow asserting an even more exasperating list of non-preconditions. For example, consider the simple axioms

$$\Gamma = \{\forall X.P(a(X)) \subset Q(X),\ Q(s),\ \forall X.Q(a(X)) \subset Q(X)\}.$$

Note that frame axiom (2) is now included in Γ so that we can derive the relation expressed in (1), with $X{=}a(s)$. Now suppose someone notes that some new relation, say R, would prevent action a from taking place, e.g., a might be the action of picking up a block, which would be difficult if the block were glued down. We can deal with a qualification, say that R is "block glued down," by rewriting the relevant axiom of Γ as

---

[3] This tradeoff is similar to that described by Brachman and Levesque [Levesque85].

$$\forall X.P(a(X)) \subset Q(X) \wedge \neg R(X)$$

But our antagonist can point out arbitrarily many such qualifications. The *qualification problem* (cf. [McCarthy80]) is that of somehow expressing this potentially infinite number of situations that precludes the application of an action.

Again, this problem has both an epistemological and heuristic component, (although intuitively, the proportions seem different than those of the frame problem). McCarthy, Kowalski, and Lifschitz' treatment of the qualification problem are related in that their solutions seek to express the intuitive statement "all action preconditions are those stated" within the formal language that specifies action preconditions and effects. For example, McCarthy proposes an extremely powerful new predicate "abnormal," which he motivates in the following quotation:

> Many people have proposed facts about what is "normally" the case. One problem is that every object is abnormal in some way, and we want to allow some aspects of the object to be abnormal and still assume the normality of the rest. We do this with a predicate *ab* standing for "abnormal." ... The argument of *ab* will be some aspect of the entities involved. Some aspects can be abnormal without affecting others. The aspects themselves are abstract entities, and their unintuitiveness is somewhat a blemish on the theory [McCarthy86, p. 93].

Regardless of McCarthy's uneasiness about "abnormal," it allows one to express statements that address the qualification problem. For example, we might now write the preconditions of action a as

$$\forall X.P(a(X)) \subset Q(X) \wedge \neg abnormal(X)$$

By using the predicate *abnormal* we acknowledge that certain abnormal situations might prevent action a from taking place. Now we can merely make individual statements about such abnormalities, e.g.,

$$abnormal(X) \subset R(X)$$

where R is "block glued down." We can now use circumscription to minimize *abnormal* and prevent action a only in those abnormal situations X that we have been told about.[4] Lifschitz [Lifschitz86a] proposes a more intuitive solution based on a relation $precond(X,Y)$ which is true when X is a propositional fluent that must be true in order for action Y to be executed. We can then provide a contextual definition for another relation $success(X,Y)$ which is true when an action X's preconditions are satisfied by situation Y:

$$\forall X,Y,Z.success(X,Y) \equiv precond(Z,X) \supset holds(Z,Y)$$

Circumscribing *precond* will maximize the situations underwhich an action can be successfully performed.

---

[4] The minimization afforded by circumscribing the theory Γ with respect to the predicate "abnormal" is not, as McCarthy notes, as straightforward as suggested here. In particular, there may be required several different forms of abnormality, in which case something more like ...*abnormal aspect*1, *abnormal aspect*2, etc. might be required. See McCarthy [McCarthy86, pp. 92-95] and [Lifschitz86a, p. 2-3] for further details.

Note that the frame problem is that of specifying a large collection of frame axioms, while the qualification problem is that of specifying a large collection of negative qualifications. The former requires a shorthand notation for a collection of axioms, while the latter requires a shorthand for making the meta statement that "nothing else can prevent the action." This view of a qualification is not unlike an explicit statement such as "I didn't know about it, so it doesn't apply" which has the flavour of expressions formalized in Moore's autoepistemic logic [Moore83]. McCarthy has noted autoepistemic reasoning as one kind of non-monotonic reasoning, and has speculated that "...Perhaps this kind of reasoning can be handled by circumscription" [McCarthy86, p. 92]. We note only that Moore insists that the consequences of any autoepistemic theory are not open to question [Moore83, pp. 274-275]. This is because they are intended to model the beliefs of the agent embracing the theory, rather than conjectures made on the basis of assumptions. The important question is whether there is any fundamental difference between autoepistemic theories and other forms of non-monotonic theories. At least part of the answer is given by Brown [Brown86], who provides a modal logic in which consistency and provability are represented as modal operators, and in which several non-monotonic reasoning systems, including Moore's, have been axiomatized. Brown's point is that first order non-monotonicity can be formalized as monotonic inference in a modal logic. We have yet to exploit this claim to provide further insight on whether autoepistemic reasoning is fundamentally different from other forms of non-monotonic reasoning.

In the case of Reiter's default logic and Nute's logic for defeasible reasoning, it is not clear how the qualification problem would be solved. One can clearly solve the problem by some epistemological gyrations, e.g., by introducing predicates like $applicable(X,Y)$, meaning that action X is applicable in situation Y, and $prevents(X,Y)$, meaning that situation X prevents action Y. For action a, we might then write a default rule like

$$\frac{applicable(a,X) : \mathbf{M}\,\neg prevents(X,a)}{\neg prevents(X,a)}$$

We are unaware of any such proposal from either Reiter or Nute, so we acknowledge that they may have some other solutions in mind. Regardless, by somehow rendering qualifications in either system, we find that the qualification problem solution depends on how the non-monotonic proof system interprets the default and defeasible assumptions. This again suggests a trade off between the complexity of expression and complexity of reasoning.

2.3. The Prediction Problem

Just as we sometimes debug programs by validation (i.e., verify the correctness of outputs based on test inputs), we can debug action specifications by making predictions, then comparing those predictions with the actual domain. When predictions are refuted, the specification is somehow incorrect; the problem is then to adjust the specification to conform to newly observed properties of the domain. Hayes [Hayes71, p. 496ff.] has called this the *prediction problem.*

Israel has suggested that all non-monotonic reasoning can be considered in a framework of scientific theory formation [Israel80]. The prediction problem is then cast as that of hypothesis generation, testing and revision, which provides a coherent framework for viewing both theoretical and practical work on inconsistency detection and consistency restoration.

For example, the formalization of logic database updates has received some attention, especially in logic programming, where Kowalski and Bowen have explained how a database of definite clauses should be updated [Bowen82]. This formalization of "theory revision" lacks some detail (e.g., there is a coarsely defined predicate *AnalyseFailureRestoreConsistency*), but provides insight into combining a logical meta and object language to express theory revision axiomatizations.

Also related to the idea of theory formation is some work of a more abstract nature under the description "logical evolution of theories" (e.g., [Gumb77, Gumb78]), whose application is not yet clear.

Practical issues in solving the prediction problem include the identification of inconsistencies, and the subsequent repair of inconsistent assumptions. Here there are at least two issues. One is to identify predictions that will provide information about competing sets of assumptions. Algorithms for finding "crucial literals" whose validation will distinguish competing theories have been specified both by Shapiro and by Seki et al. [Shapiro82, Seki85]. Crucial literals are like scientific experiments, whose results can refute some theories and help confirm others. Equally important is the recording and use of derivation information so that derivations that support certain predictions can be efficiently modified (e.g., [deKleer85, Greiner86]).

3. Approximate Representation and Reasoning

Recent formalizations of non-monotonic reasoning systems directly address the question of drawing conclusions based on assumptions. By specifying a language of facts and assumptions, and a procedure for both drawing and retracting assumption-based conclusions, non-monotonic reasoning systems provide a natural tool for investigating possible solutions to the frame, qualification, and prediction problems.

Since our theories can only approximate the domain of interest and we still have to draw conclusions to synthesize plans (or more generally, to take action), we must exercise assumptions about the approximation, e.g., the STRIPS assumption [Fikes71]. We would rather that our language of assumptions be such that we can formally define the consequences of assumption-based reasoning in such a way as to enable their retraction upon detecting inaccurate assumptions.

3.1. Multiple Models and Theory Preference

Even without a concrete syntax, we can discuss the relationship between an *approximate theory* T, defined as a collection of *facts* F and a collection of *assumptions* A so that $T = F \cup A$. The set of models of a consistent approximate theory T is a subset of the models of F. If there are sets of assumptions, e.g., $A1$, $A2$, ..., An, then each theory $F \cup Ai$ has as models a potentially different subset of the models of F.

For any goal description, we are interested in approximate theories (i.e., $F \cup Ai$) that support the derivation of a planning goal. We are further interested in which of the possible approximate theories is most likely. Note that some have severely criticized various non-monotonic theories because they admit multiple extensions (cf. [Reiter80]) or have non-constructible closures (e.g., [Hanks86, McDermott86]), but we contend that multiple possible approximate theories are an essential characteristic of scientific theory formation (cf. [Goebel86a]).

To acknowledge the possibility of multiple approximate theories is to acknowledge a requirement for an ordering of such theories. In fact, to actually act upon theorems of approximate theories, means to decide which is likely to accomplish the goal. The required ordering is defined on the set of models of $F \cup A1$, $F \cup A2$, ..., $F \cup An$. To actually compute the ordering by examining the $F \cup Ai$, some syntactic definition of the semantic ordering is needed. We would like any syntactic definition to satisfy the following theorem schema:

Theorem: (*model ordering preservation*): Let $\mathbf{M}(T)$ denote the models of theory T. Let P_M be the specification of a model ordering procedure and P be a theory ordering procedure. If P is an implementation of P_M then
for all theories T_i, T_j, $P(T_i,T_j)$ iff $P_M(\mathbf{M}(T_i),\mathbf{M}(T_j))$.

This merely says that any implementation of a theory ordering strategy based on symbolic manipulation must meet its semantic specification. Poole [Poole85] provides an example.

In our explanation of a theory formation approach to the planning problem given below, we use a theory ordering procedure that works on just the Ai rather than $F \cup Ai$, or its closure. We therefore further require the following theorem schema:

Theorem: (*theory ordering preservation*): For any ordering procedure P, $P(F \cup Ai, F \cup Aj)$ iff $P(\mathit{theorems}(F \cup Ai), \mathit{theorems}(F \cup Aj))$.[5]

We expect that any non-monotonic reasoning system based on approximate theories will have to satisfy these theorems in order to be well-defined and computationally feasible.

[5] $\mathit{theorems}(F)$ denotes the closure of axioms F under theoremhood.

3.2. Using Theory Preference Knowledge

We believe that every non-monotonic reasoning system should be viewed as one that generates approximate theories and then applies some theory preference criteria to order those approximate theories. Furthermore, we hypothesize that all non-monotonic reasoning systems can be compared on a continuum defined in terms of whether theory preference is defined by axioms of the formal language or as a heuristic procedure for ordering multiple approximate theories.

As evidence, consider McCarthy's remark about various formal languages for expressing assumptions:

> We can ... regard the process of deciding what facts to take into account and then circumscribing as a process of compiling from a slightly higher-level non-monotonic language into mathematical logic, especially first-order logic [McCarthy86, p. 107].

The hint is that circumscription's semantic effect of minimizing the interpretation of certain predicates can be had in more direct ways. Note, for example, that negation-as-failure and circumscription are methods of non-monotonic reasoning where theory preference (i.e., preference for certain models of certain $F \cup Ai$) are "compiled in" as adjuncts to facts. For example, rather than actually produce the appropriate circumscription for *abnormal* in the specification of qualifications, we might employ Prolog's negation-as-failure to minimize *abnormal*:[6]

$$holds(P,do(A,S)) \leftarrow holds(P,S) \& not(abnormal\ldots))$$

This idea also derives from Reiter, who has shown an equivalence between predicate completion [Clark78] and the predicate circumscription of Horn-defined predicates [Reiter82]. Similar evidence is provided by Poole [Poole86a] who shows the relationship between negation-as-failure and the use of defaults in theory formation.

We propose that further elaboration, by way of transforming different non-monotonic theories into one another, will demonstrate that scientific theory formation and its requirement for preference amongst multiple competing approximate theories can, for interesting problem domains, be compiled into Prolog-like languages that serve as efficient machine languages. Circumscription will continue to have a role as a specification language,[7] just as Clark's completion is a specification for a partial[8] interpretation of negation-as-failure [Clark78].

The obvious method for actually proving the hypothesis, as opposed to providing evidence by example, is to define a series of logical transformations that formally derive one non-monotonic language theory from another. This kind of transformation has already

---

[6] Suggested by Vladimir Lifschitz, University of Toronto colloquium, October 21, 1986.

[7] Suggested by Vladimir Lifschitz, in a personal communication.

[8] See Lloyd [Lloyd84, pp. 82-87] for a discussion of the critical difference between the closed-world assumption and negation-as-failure.

been most successful in showing the relationship between MYCIN rules and backward reasoning with causal rules (e.g., see [Goebel86b, p. 218-219]), and for converting ordinary logic programs into equivalent and more efficient parallel logic programs (e.g., see [Seki86]). In fact, such results are consistent with Hayes GOLUX project, and its attempt to view computation as controlled deduction. Somewhere along the spectrum between theory preference heuristics and categorical theories is the need to effectively reduce the number of competing approximate theories by considering preferences as soon as possible — incrementally, in fact, as already suggested by Poole [Poole86b]. The incremental construction of approximate theories is merely the incremental application of theory preference as control information.

Regardless of the speculative nature of this continuum hypothesis, it is clear that its pursuit will provide further insight into the relationship between various kinds of non-monotonic reasoning systems.

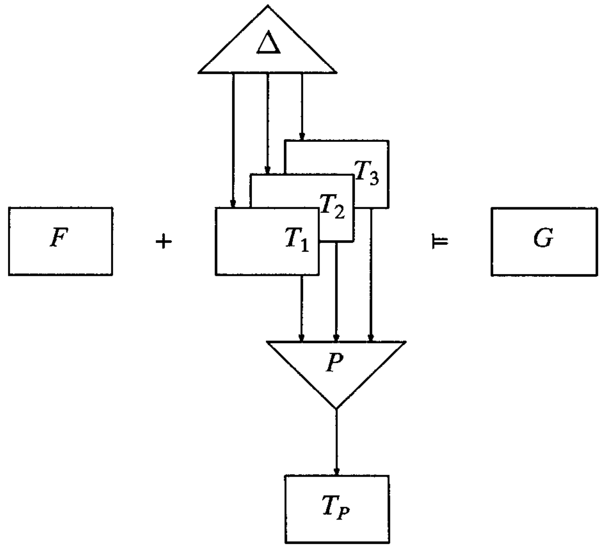
4. Using Theory Formation as a Framework for the Planning Problem

We are developing a framework for solving the frame, qualification and prediction problems, based on theory formation and an explicit preference heuristic (called chronological persistence) that provides one possible definition for a theory ordering procedure P on theories of a first order language of facts and assumptions.

Frame axioms are expressed as default assertions, consistent instances of which can be assumed to complete the derivation of plans. Each consistent set of assumptions describes a situation in which the goal is satisfied. The notion of persistence can be used to distinguish multiple competing situation descriptions and thereby determine whether the goal is predicted by the preferred situation description [Goodwin86]. The theory formation system underlying our planning framework is Theorist [Poole87]. Theorist views reasoning as scientific theory formation (rather than as deduction). Theorist reasons by building theories (henceforth *theory* means scientific theory — not logical theory) that explain a set of observations. A *theory*, consisting of instances drawn from a set of *possible hypotheses*, is said to *explain* a set of *observations* if the theory, together with the *facts*, logically implies the observations; it must also be consistent with the facts.

In theory formation, multiple theories that explain the observations correspond to the multiple minimal models in circumscription and to the multiple extensions in default logic. Given that it is difficult, if not impossible, to determine a unique theory by strengthening the underlying set of axioms, it seems natural instead to accept the approximate theories as possible explanations and then to rank them according to some suitable criterion and thereby determine a preferred explanation. (For example, the Copernican view of the solar system is not the result of the complete axiomatization of the properties of the universe, but rather it is a preferred explanation supported by a necessarily weak set of axioms.)

In the theory formation framework, planning problems are described (Fig. 1) using a combination of *facts* and *defaults* (possible hypotheses). In worlds with complete information, the initial situation is described solely by facts. The effects of an action partition relations describing a world into two groups. Relations that are known to be changed by the performance of the action form one group, and all other relations — those *presumed* to be unaffected by the performance of the action — form the other group. Laws of



Find $T_i \subseteq \Delta$ such that:

$F \cup T_i \models G$, and

$F \cup T_i$ is consistent

where Δ is a set of frame defaults

T_i is a set of instances of frame defaults

F is a set of facts describing the initial situation and laws of motion

G is a goal description (or its negation).

Then choose $T_P \in \{T_1, T_2, \ldots\}$

such that $\forall i \; T_P >^p T_i$

where $>^p$ is a theory preference heuristic

T_P is a preferred theory

P is a procedure for choosing a preferred theory.

Figure 1. Planning in a Theory Formation/Theory Preference Framework

motion [Hayes73] describing the relations that are known to change are expressed as facts, while the laws of motion for the relations presumed invariant are expressed as defaults. In our current work, we have shown how these defaults correspond to frame axioms. Collections of ground instances of these *frame defaults* form theories from which predictions can be made. In this theory formation framework, a situation is named by the sequence of actions from which it results and it is described by the facts together with a theory. Since there can be many theories describing a situation, the issue of theory preference arises. For example, an action is presumed applicable if its preconditions are predicted by the preferred theory. In this framework, a solution to a planning problem has two components: a sequence of actions (a plan) which achieves the goal, and a theory which predicts the goal description.

Since it is not known in advance whether the goal description or its negation holds, they are treated as competing *predictions*. From the theories predicting the goal or its negation, one seeks a preferred theory from which the truth of the goal description is determined. The above representation of planning problems and its use is considered in greater detail elsewhere [Goodwin87].

$F = \{$	The set of facts:
	Initial Situation:
$\neg loaded(0)$,	The gun is not loaded
$alive(0)$,	John is alive
	Action: load
$loaded(do(load, S))$,	The gun is loaded after the action load
	Action: wait (no known changes)
	Action: shoot
$\neg alive(do(shoot, S)) \leftarrow loaded(S)$,	John dies when shot with a loaded gun
$\neg loaded(do(shoot, S))\}$	After shooting, the gun is not loaded
$\Delta = \{$	The set of Frame Defaults:
$[A, S]\ loaded(do(A, S)) \leftrightarrow loaded(S)$,	
$[A, S]\ alive(do(A, S)) \leftrightarrow alive(S)\}$	

Figure 2. Yale Shooting Scenario

Situations:
$1 \equiv do(load, 0)$
$2 \equiv do(wait, do(load, 0))$
$3 \equiv do(shoot, do(wait, do(load, 0)))$

Intended Model Features	0	1	2	3
loaded	F	T	T	F
alive	T	T	T	F

T_1 Model Features	0	1	2	3
loaded	F	T	T	F
alive	T	T	T	F

T_2 Model Features	0	1	2	3
loaded	F	T	F	F
alive	T	T	T	T

Figure 3. Essential Features of Models for the Shooting Example

As an example of the representation framework, consider the axiomatization of the Yale shooting scenario (Fig. 2) [Hanks86]. In this example, two sets of statements are given. One set, F, describes the relations that are accepted as true in the world. Within this set, there are two kinds of axioms: those that describe the initial situation, and those that describe the changes caused by the performance of actions. A second set of statements, Δ, contains a frame default for each *primitive* [Fikes71] relation occurring in F. Collections of instances (the variables in the square brackets are to be instantiated) of these defaults form theories. An instance of a frame default asserts that the truth-value of the corresponding relation is preserved when performing the particular action in the particular situation.

In the above example, consider whether John will be alive after the actions *load*, *wait*, and then *shoot*. Two of the theories which describe the invariance of relations for this sequence of actions are

$$T_1 = \{loaded(do(wait,do(load,0))) \leftrightarrow loaded(do(load,0)),$$
$$alive(do(load,0)) \leftrightarrow alive(0),$$
$$alive(do(wait,do(load,0))) \leftrightarrow alive(do(load,0))\} \text{ and}$$

$$T_2 = \{alive(do(load,0)) \leftrightarrow alive(0),$$
$$alive(do(wait,do(load,0))) \leftrightarrow alive(do(load,0)),$$
$$alive(do(shoot,do(wait,do(load,0)))) \leftrightarrow alive(do(wait,do(load,0)))\}.$$

The statements in the theories are instances of frame defaults that record two different sets of assumptions about the action sequence. Note that these two theories conflict since

$$F \cup T_1 \models \neg alive(do(shoot,do(wait,do(load,0)))) \text{ while}$$
$$F \cup T_2 \models alive(do(shoot,do(wait,do(load,0)))).$$

In our intended model (i.e., the one that corresponds to our intuitions), however, $\neg alive(do(shoot,do(wait,do(load,0))))$ is true (Fig. 3).

In view of the possibility of having multiple and potentially conflicting theories, how can a selection be made between them? The notion of persistence is intended to reflect what McCarthy calls the "common sense law of inertia;"[9] that is, when an action is performed, most things remain unchanged. When presented with competing theories, each making different predictions, we desire a heuristic that prefers the theory that corresponds to our intuition about persistence. The heuristic should simultaneously satisfy three criteria:

1) *accuracy* — it should select a theory which makes predictions that correspond to our expectations;
2) *sufficiency* — if the goal description (or its negation) is expected, then it should be predicted by the selected theory;
3) *resource conservatism* — it should select a theory with maximal obtainable *accuracy* for minimal computational effort.

In order to formalize this intuition, we propose the following semantic account of persistence.

Definition 1.

Let $s_n = do(a_{n-1},do(a_{n-2},\ldots,do(a_1,s_1))\ldots)$ be a situation in the domain. The situations s_1 to s_n determine a *path* which we write as $<s_1,s_n>$. Furthermore, the *length*

---

[9] Cf. [Lifschitz86b, p. 408].

of a path $<s_1,s_n>$ is defined to be $n-1$, one less than the number of situations on the path from s_1 to s_n. A *unit path* is a path of length one.

Definition 2.

Given a consistent theory T

a) A *primitive* propositional fluent $R(\bar{x})$ is said to *persist* in T over the unit path $<s,do(a,s)>$ if
$\forall S \in <s,do(a,s)>, T \models R(\bar{x},S)$ or $\forall S \in <s,do(a,s)>, T \models \neg R(\bar{x},S)$;

b) A *persistence set* P_T for theory T is a set of propositional fluents indexed by the unit path over which they persist in T. The persistence set P_T^i is the subset of P_T containing the propositional fluents indexed by unit path $<s_i,s_{i+1}>$;

c) A domain-dependent partial ordering of persistence sets (denoted by $\geq^p$) can be defined to reflect the relative likelihoods of each persistence set. Thus if two persistence sets differ on a highly persistent propositional fluent $R(\bar{x})$, the persistence set which includes it would be higher ranked. We can define $=^p$ and $>^p$ in the usual way.

In definition 2a persistence is defined in relation to primitive [Fikes71] propositional fluents. This ensures that the truth-value of a defined propositional fluent is preserved only when its associated primitives persist. Definition 2c leaves open the question about how $\geq^p$ should be defined. As indicated, the partial ordering $\geq^p$ is a domain dependent ranking of the likelihood of persistence sets. If all we know about the likelihood of a propositional fluent persisting is that it is more likely to persist than not (i.e., the common sense law of inertia), then we can define $\geq^p$ to be $\supseteq$. When we know more about the persistence characteristics of various propositional fluents, then this information can be included by providing an arbitrarily complex definition of $\geq^p$. For example, if we know that all propositional fluents in the domain of interest are equally likely to persist, then we can define $\geq^p$ to be a comparison of the cardinality of persistence sets. In this case, the partial ordering $\geq^p$ is defined as the total ordering $>$ on the cardinality of persistence sets. When the domain knowledge about persistence probabilities is complex, Neufeld's work on common sense probabilistic theory comparators may be useful [Neufeld87].

This semantic definition of persistence can be used to distinguish theories. This observation motivates the following definitions.

Definition 3.

A theory T_1 is said to be *chronologically more persistent* than theory T_2 over the path $<s_1,s_n>$ if

a) $\exists j < n,\ P_{T_1}^j >^p P_{T_2}^j$

b) $\forall i < j,\ P_{T_1}^i =^p P_{T_2}^i$

where P_T^i is the persistence set for T over $<s_i,s_{i+1}>$. The partial ordering $\geq^p$ on persistence sets defines a partial ordering $\geq^{cp}$ on theories. Thus, $T_1 >^{cp} T_2$ means T_1 is chronologically more persistent than T_2.

Definition 4.

A theory is said to be *chronologically maximally persistent* (CMP) over a path if there does not exist a theory which is more persistent over that path.

In the Yale shooting scenario (Fig. 2), if $\geq^p$ is taken to be $\supseteq$ then

$$T_1 = \{loaded(do(wait,do(load,0))) \leftrightarrow loaded(do(load,0)),$$
$$alive(do(load,0)) \leftrightarrow alive(0),$$
$$alive(do(wait,do(load,0))) \leftrightarrow alive(do(load,0))\}$$

is the unique chronologically maximally persistent theory over the path $<0,do(shoot,do(wait,do(load,0)))>$.

The theory T_1 predicts $\neg alive(do(shoot,do(wait,do(load,0))))$. There are other consistent theories, some of which even make the opposite prediction, for instance

$$T_2 = \{alive(do(load,0)) \leftrightarrow alive(0),$$
$$alive(do(wait,do(load,0))) \leftrightarrow alive(do(load,0)),$$
$$alive(do(shoot,do(wait,do(load,0)))) \leftrightarrow alive(do(wait,do(load,0)))\}.$$

This theory predicts $alive(do(shoot,do(wait,do(load,0))))$, but it is not chronologically maximally persistent. The CMP theory T_1 corresponds to the intended model (cf. Fig. 3). It accurately reflects our expectations about the stability of the world. Therefore, at least in this case, chronological maximal persistence satisfies the accuracy criterion and thus could serve as a heuristic preference measure. It also satisfies the sufficiency criterion since it predicts the goal (or its negation depending on which is expected). CMP theories, however, are usually not the most economical choice — they do not satisfy the conservatism criterion. In a CMP theory, many of the propositional fluents that persist are simply irrelevant to the problem at hand. A better theory would exclude all irrelevant instances of frame defaults. The preference heuristic described here can be modified to select theories which exclude irrelevant defaults and still meet the sufficiency criteria. A semantic account of this modified preference heuristic and a procedure to compute preferred theories is presented elsewhere [Goodwin87].

The preference measure, "chronological persistence," is only a heuristic as is evident from an example given by Kautz [Kautz86, p. 404]. In that example, a car is observed to be missing some time after it was parked. The CMP theory would predict the car was still in the parking lot until it was observed to be missing, but there is no reason to prefer this prediction — it could have been stolen any time between when it was parked and when it was observed missing. Though our preference measure is only a heuristic, it is useful nevertheless, since it seems to correspond to our intuition about the invariance of relations much of the time.

5. Related Work

Hanks and McDermott [Hanks85, Hanks86] show that using default (or other non-monotonic) reasoning to deal with the frame problem inevitably results in the need to choose between multiple models. Because the popular forms of non-monotonic reasoning don't seem to provide a mechanism for choosing between models, they come to the discouraging conclusion that logic is inadequate as an AI representation language. They turn to a direct procedural characterization to describe default reasoning processes and give an algorithm that generates their intended model for a set of axioms. In our terms, this model is a model of $F \cup T_{CMP}$. We have demonstrated how this theory can be arrived at in a simple, intuitive theory formation framework through a semantically well-defined theory preference heuristic. Hopefully, this disspells, to some extent, the perception of inadequacy of logic stemming from Hanks' and McDermott's work.

Another related idea is that of Lifschitz [Lifschitz85, Lifschitz86c, Lifschitz86b]. He makes the observation that the usual forms of circumscription [McCarthy80] are inadequate for dealing with axiomatizations of planning problems (cf. [McCarthy86]). To overcome this inadequacy, he introduces *pointwise circumscription*. Our notion of chronological maximization of persistence is analogous to a form of prioritized pointwise circumscription which prefers "minimization at earlier moments of time". More recently, Lifschitz has devised a formalism which uses ordinary circumscription to minimizes causes and preconditions [Lifschitz86a]. Apparently, minimizing causes allows minimization of uncaused changes which captures the intuition behind the common sense law of inertia. Minimizing preconditions minimizes the qualifications for the performance of action. Thus, Lifschitz' recent work address both the frame and qualification problem. Lifschitz is able to avoid the multiple model problem because the *causes* relation has no situation parameter. This is similar to Kowalski's *preserves* relation [Kowalski79, pp. 136-137] which is minimized through negation-as-failure. While not having a situation parameter in causes (or preserves) does avoid the multiple model problem, it seems to do this by sacrificing expressiveness. For instance, the changes induced by flipping a toggle switch depend on the previous situation; so it seems in this case, the causes relation for the action *flip* needs a situation parameter.

The work of Shoham [Shoham86a, Shoham86b] is related to the work presented here. His work on the *initiation problem*[10] led him to the idea of *chronological maximization of ignorance*. While the initiation problem is different from the frame problem, solutions to each problem reflect the need to maximize (or minimize) step by step (i.e., chronologically).

Recent work by Kowalski and Sergot [Kowalski86a, Kowalski86b] also addresses the frame problem. More specifically, Kowalski [Kowalski86b] proposes a first order *persistence* axiom that specifies how one can deduce whether a relation holds at a given time in a particular temporal database. A database is formalized as the ground terms of a logical theory specified in *event calculus*; the first order ground atomic formula $holds(r,t)$ is true when r is an instance of a relation, and t is a time interval over which the relation is true. Database update constraints are specified in terms of *terminate* and *initiate*

[10] This is the qualification problem in a temporal setting.

conditions on relations, events and actions. The epistemological aspect of the frame problem is claimed to be solved by axiomatizing the database in terms of a single relation *holds* (cf. [Kowalski79]), and relying on the persistence axiom's use of negation-as-failure to assume that nothing affects the truth of a relation unless explicitly declared. The heuristic aspect of the frame problem is viewed as efficiently using the persistence axiom to answer the general question of whether an instance of a relation r holds at time t. This problem is solved by describing an efficient method for implementing the use of the persistence axiom.

This notion of persistence, rendered as a first order axiom that specifies what "holds" in terms of how explicit initiate and terminate conditions affect the truth of relations, is only weakly related to the notion of persistence described here. Kowalski's formulation has no explicit concept of the update constraints being contingent or assumable if consistent, consequently there is no possibility of multiple conflicting answers to the question "does r hold at time t?" The persistence axiom relies on negation-as-failure while isolating those constraints that affect the truth of a relation so that different answers to the same question can be had due to intervening updates. However, there is no possibility of uncertain or multiple possible responses to questions of a relation's future persistence.

It seems that Kowalski assumes that frame axioms as defaults are unnecessary, as the concept of non-monotonicity can be handled more generally by using negation-as-failure within the persistence axiom. However, the burden to explicitly assert the affect of actions on relations remains; a default-like statement of the form "normally relation R persists" is not possible. Kowalski's alternative definition might be rephrased as "relation R persists, until I tell you otherwise."

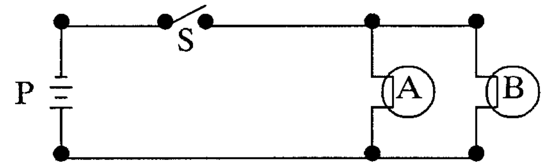
Finally, Kautz [Kautz86] proposes a solution to the frame problem using a generalization of circumscription. He defines the following partial ordering of models:

$$M1 \leq M2 \text{ if and only if}$$
$$\forall t,f \,.\, (t,f) \in M1[Clip] \supset ((t,f) \in M2[Clip]) \vee$$
$$\exists t2,f2 \,.\, t2 < t \;\&\; ((t2,f2) \in M2[Clip]) \;\&\; ((t2,f2) \notin M1[Clip])$$

and

$$M1 < M2 \text{ if } M1 \leq M2 \text{ and not } M2 \leq M1$$

The predicate $Clip(t,f)$ is true when the persistence of a fact f ceases at time t. From this definition, a model $M1$ is strictly better than $M2$ (i.e., $M1 < M2$) when they are identical (in terms of Clip) up to some time t at which some fact f changes in $M2$ but not in $M1$ (i.e., $(t,f) \in M2[Clip]$ but $(t,f) \notin M1[Clip]$). This model ordering corresponds to the chronological maximization of persistence when $\geq^p$ is taken to be $\supseteq$. To illustrate this, consider the example in figure 4 and two of the possible models corresponding to it (Fig. 5). Here we have two lights (A and B) connected in parallel through a switch (S) to a power source (P). For the period of interest, the switch and the two lights have an equivalent failure rate. The power source and the wiring are completely reliable. Under Kautz's model ordering, M_1 and M_2 are incomparable (both are minimal). When $\geq^p$ is taken to be $\supseteq$ then a theory T_1 corresponding to M_1 has the same chronological persistence as T_1 (corresponding to M_2) over a path corresponding to the action between time 0 and time 1, but if $\geq^p$ is taken to be a comparison of the cardinality of the persistence



$$
\begin{array}{ll}
F = \{ & \\
\neg on(s,0), & \neg hold(0,on(s)) \\
\neg on(a,0), & \neg hold(0,on(a)) \\
\neg on(b,0), & \neg hold(0,on(b)) \\
ok(s,0), & hold(0,ok(s)) \\
ok(a,0), & hold(0,ok(a)) \\
ok(b,0), & hold(0,ok(b)) \\
 & \\
on(s,1), & hold(1,on(s)) \\
\neg on(a,1), & \neg hold(1,on(a)) \\
\neg on(b,1), & \neg hold(1,on(b)) \\
 & \\
on(a,T) \leftarrow & hold(T,on(a)) \leftarrow \\
\quad on(s,T) \wedge ok(s,T) \wedge ok(a,T), & \quad hold(T,on(s)) \wedge hold(T,ok(s)) \wedge hold(T,ok(a)) \\
on(b,T) \leftarrow & hold(T,on(b)) \leftarrow \\
\quad on(s,T) \wedge ok(s,T) \wedge ok(b,T)\} & \quad hold(T,on(s)) \wedge hold(T,ok(s)) \wedge hold(T,ok(b))
\end{array}
$$

$$
\begin{array}{ll}
\Delta = \{ & \\
[X,T]\ ok(X,T{+}1) \leftrightarrow ok(X,T), & hold(T{+}1,F) \oplus clip(T{+}1,F) \leftarrow hold(T,F) \\
[X,T]\ on(X,T{+}1) \leftrightarrow on(X,T)\} &
\end{array}
$$

Figure 4. Comparison with Kautz´ Model Ordering

Model M_1		
	0	1
ok(s)	T	T
ok(a)	T	F
ok(b)	T	F
on(s)	F	T
on(a)	F	F
on(b)	F	F

Model M_2		
	0	1
ok(s)	T	F
ok(a)	T	T
ok(b)	T	T
on(s)	F	T
on(a)	F	F
on(b)	F	F

Figure 5. Two of the Models for the Light Circuit Example

sets corresponding to the above theories then T_2 has more chronological persistence than T_1. Thus, incorporating the assumption that all persistences are equally likely by defining $\geq^p$ to be a comparison of cardinalities enables us to distinguish the two models. The preference heuristic reflects our expectation that the switch failing alone is a better explanation (theory) than both lights failing.

6. Conclusions

We have attempted to sketch the background to a thesis that suggests a very general relationship between all forms of non-monotonic reasoning. By casting the planning problem in a general way, we have pointed out the inevitability of using some kind of approximate theories, and for requiring methods for distinguishing and preferring one approximate theory over another.

Our own approach to the planning problem uses an approximate reasoning system based on theory formation. We believe that our approach is consistent with the use of logic, but is procedural in the sense that the specification of theory preference is outside the logic of generated theories. The appeal of this point on the continuum is its basis in the philosophy of science and so-called "deductive nomological reasoning," [Hempel65] as well as the ease with which it can be applied to complex reasoning problems like the planning problem.

Acknowledgements

The Philosophical Engineering and Trendy Formalisms division of the Computer Science department's Logic Programming and AI group has helped mould these ideas. Vladimir Lifschitz's colloquium at the University of Toronto on October 21, 1986, and subsequent discussions contributed to the hypothesis about compiled versus interpreted theory preference. We also benefitted from discussions with David Poole about his ideas on how the Theorist framework is related to negation-as-failure and to circumscription. Special thanks to David Poole, Denis Gagné, Eric Neufeld, André Trudel, and Kelly Arrey for their careful reading and helpful comments on an earlier draft of this paper. This research was supported by Natural Sciences and Engineering Research Council of Canada grants A0894 and G1587.

References

[Bowen82]

K. Bowen and R.A. Kowalski (1982), Amalgamating language and metalanguage in logic programming, *Logic Programming*, A.P.I.C. Studies in Data Processing 16, K.L. Clark and S. A. Tarnlund (eds.), Academic Press, New York, 153-172.

[Bowen85a]
K.A Bowen and T. Weinberg (1985), A meta-level extension of Prolog, *IEEE 1985 Symposium on Logic Programming*, July 15-18, Boston, Massachusetts, 48-53.

[Bowen85b]
K.A. Bowen (1985), Meta-level programming and knowledge representation, *New Generation Computing* **3**(4), 359-383.

[Brown86]
F.M Brown (1986), A comparison of the commonsense and fixed point theories of nonmonotonicity, *Proceedings of the National Conference on Artificial Intelligence (AAAI-86)*, August 11-15, University of Pennsylvania, Philadelphia, Pennsylvania, 394-400.

[Clark78]
K.L. Clark (1978), Negation as failure, *Logic and Data Bases*, H. Gallaire and J. Minker (eds.), Plenum Press, New York, 293-322.

[deKleer85]
J. deKleer (1985), Problem solving with the ATMS, Intelligent Systems Laboratory, Xerox PARC, Palo Alto, California, September.

[Fikes71]
R.E. Fikes and N.J. Nilsson (1971), STRIPS: a new approach to the application of theorem proving in problem solving, *Artificial Intelligence* **2**(3&4), Winter, North-Holland, Amsterdam, 189-208.

[Goebel86a]
R. Goebel (1986), Exhuming the criticism of the logicist, Research report CS-86-33, Department of Computer Science, University of Waterloo, Waterloo, Ontario, September, 9 pages.

[Goebel86b]
R.G. Goebel, K. Furukawa, and D.L. Poole (1986), Using definite clauses and integrity constraints as the basis for a theory formation approach to diagnostic reasoning, *Third International Conference on Logic Programming*, July 14-18, Imperial College of Science and Technology, London, England, 211-222.

[Goodwin86]
S.D. Goodwin and R.G. Goebel (1986), Theory Preference Based on Persistence, CS-86-34, Department of Computer Science, University of Waterloo, Waterloo, Ontario, September, 21 pages.

[Goodwin87]
S.D. Goodwin (1987), Representing Frame Axioms as Defaults, Master's dissertation, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada [In preparation].

[Green69]
C.C. Green (1969), Application of Theorem Proving to Problem Solving, *Proceedings of the First International Joint Conference on Artificial Intelligence (IJCAI-69)*, May 7-9, Washington, D.C., 219-239.

[Greiner86]
R. Greiner (1986), Using inferential support information for all logic-based inference tasks, Department of Computer Science, University of Toronto, Toronto, Ontario, September, 37 pages.

[Gumb77]
R.D. Gumb (1977), A mechanized proof procedure for free intensional logics, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI-77)*, August 22-25, MIT, Cambridge, Massachusetts, 567.

[Gumb78]
R.D. Gumb (1978), Computation aspects of evolving theories, *ACM SIGART Newsletter* **67**, 13.

[Hanks85]
S. Hanks and D. McDermott (1985), Temporal reasoning and Default Logics, TR YALEU/CSD/RR# 430, Computer Science Department, Yale University.

[Hanks86]
S. Hanks and D. McDermott (1986), Default Reasoning, Nonmonotonic Logic, and the Frame Problem, *Proceedings of the National Conference for Artificial Intelligence (AAAI-86)* **1**, August 11-15, University of Pennsylvania, Philadelphia, Pennsylvania, 328-333.

[Hayes71]
P.J. Hayes (1971), A Logic of Actions, *Machine Intelligence*, vol. 6, B. Meltzer and D. Michie (eds.), Edinburgh University Press, Edinburgh, Scotland, 495-520.

[Hayes73]
P.J. Hayes (1973), The frame problem and related problems in Artificial Intelligence, *Artificial and Human Thinking*, A. Elithorn and D. Jones (eds.), Jossey-Bass, London, England, 45-59.

[Hempel65]

C.G. Hempel (1965), *Aspects of scientific explanation and other essays in the philosophy of science*, The Free Press, New York.

[Israel80]

D.J. Israel (1980), What's wrong with non-monotonic logic, *Proceedings of the First Annual National Conference on Artificial Intelligence (AAAI-80)*, August 18-21, Stanford University, Stanford, California, 99-101.

[Kautz86]

H. Kautz (1986), The logic of persistence, *Proceedings of the National Conference on Artificial Intelligence (AAAI-86)*, August 11-15, University of Pennsylvania, Philadelphia, Pennsylvania, 401-405.

[Kowalski79]

R.A. Kowalski (1979), *Logic for Problem Solving*, Artificial Intelligence Series 7, Elsevier North Holland, New York.

[Kowalski86a]

R.A. Kowalski and M.J. Sergot (1986), A Logic-based calculus of events, *New Generation Computing* **4**(1), Ohmsha Ltd. and Springer Verlag, 67-95.

[Kowalski86b]

R.A. Kowalski (1986), Database updates in the event calculus, DOC 86/12, Department of Computing, Imperial College, London, England, July, 29 pages.

[Ladkin86a]

P. Ladkin (1986), Primitives and units for time specification, *Proceedings of the National Conference on Artificial Intelligence (AAAI-86)*, August 11-15, University of Pennsylvania, Philadelphia, Pennsylvania, 354-359.

[Ladkin86b]

P. Ladkin (1986), Time representation: a taxonomy of interval relations, *Proceedings of the National Conference on Artificial Intelligence (AAAI-86)*, August 11-15, University of Pennsylvania, Philadelphia, Pennsylvania, 360-366.

[Levesque85]

H.J. Levesque and R.J. Brachman (1985), A fundamental tradeoff in knowledge representation and reasoning, *Readings in Knowledge Representation*, R.J. Brachman and H.J. Levesque (eds.), Morgan Kaufmann, Los Altos, California, 42-70.

[Lifschitz85]
V. Lifschitz (1985), Circumscription in the Blocks World, Stanford University, Stanford, December, 17 pages [unpublished].

[Lifschitz86a]
V. Lifschitz (1986), Formal theories of action, Stanford University, Stanford, September, 13 pages [unpublished].

[Lifschitz86b]
V. Lifschitz (1986), Pointwise circumscription: preliminary report, *Proceedings of the National Conference on Artificial Intelligence (AAAI-86)*, August 11-15, University of Pennsylvania, Philadelphia, Pennsylvania, 406-410.

[Lifschitz86c]
V. Lifschitz (1986), Pointwise Circumscription, Stanford University, Stanford, March, 15 pages [unpublished].

[Lloyd84]
J.W. Lloyd (1984), *Foundations of logic programming*, Springer-Verlag, New York.

[McCarthy69]
J. McCarthy and P.J. Hayes (1969), Some Philosophical Problems from the Standpoint of Artificial Intelligence, *Machine Intelligence*, vol. 4, B. Meltzer and D. Michie (eds.), Edinburgh University Press, Edinburgh, 463-502.

[McCarthy77]
J. McCarthy (1977), Epistemological problems of Artificial Intelligence, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI-77)*, August 22-25, MIT, Cambridge, Massachusetts, 1038-1044.

[McCarthy80]
J. McCarthy (1980), Circumscription—a form of non-monotonic reasoning, *Artificial Intelligence* **13**(1&2), April, North-Holland, Amsterdam, 27-39.

[McCarthy86]
J. McCarthy (1986), Applications of circumscription to formalising common-sense knowledge, *Artificial Intelligence* **28**(1), February, North-Holland, Amsterdam, 89-116.

[McDermott86]
D. McDermott (1986), A critique of pure reason, Computer Science Department, Yale University, June, 25 pages.

[Moore83]
R.C. Moore (1983), Semantical considerations on nonmonotonic logic, *Proceedings of Eight International Joint Conference on Aritificial Intelligence (IJCAI-83)*, August 8-12, Karlsruhe, Germany, 272-279.

[Neufeld87]
E.M. Neufeld (1987), Thinking Logically about Probabilities, Department of Computer Science, University of Waterloo, Waterloo, Ontario, January.

[Nute86]
D. Nute (1986), LDR: a logic for defeasible reasoning, ACMC Research Report 01-0013, Advanced Computational Methods Center, University of Georgia, Athens, Georgia.

[Poole85]
D.L. Poole (1985), On The Comparison of Theories: Preferring the Most Specific Explanation, *Proceeding of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, August 16-18, UCLA, Los Angeles, California, 144-147.

[Poole86a]
D.L. Poole (1986), Negation as failure and defaults: a critical comparison, Department of Computer Science, University of Waterloo, Waterloo, Ontario, November, 8 pages.

[Poole86b]
D.L. Poole (1986), Building consistent theories, Department of Computer Science, University of Waterloo, Waterloo, Ontario, November, 22 pages.

[Poole87]
D.L. Poole, R.G. Goebel, and R. Aleliunas (1987), Theorist: a logical reasoning system for defaults and diagnosis, *The Knowledge Frontier: Essays in the Representation of Knowledge*, N.J. Cercone and G. McCalla (eds.), Springer-Verlag, New York, 331-352 [in press].

[Reichgelt86]
H. Reichgelt (1986), Semantics for reified temporal logic, Research Paper No. 299, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, Scotland.

[Reiter80]
R. Reiter (1980), A logic for default reasoning, *Artificial Intelligence* **13**(1&2), April, North-Holland, Amsterdam, 81-132.

[Reiter82]
R. Reiter (1982), Circumscription implies predicate completion (sometimes), *Proceedings of the National Conference on Artificial Intelligence (AAAI-82)*, August 18-20, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 418-420.

[Seki85]
H. Seki and A. Takeuchi (1985), An algorithm for finding a query that discriminates competing hypotheses, Institute for New Generation Computer Technology, Tokyo, Japan, December, 14 pages.

[Seki86]
H. Seki and K. Furukawa (1986), Compiling control by a program transformation approach, Institute for New Generation Computer Technology, Tokyo, Japan, November, 24 pages.

[Shapiro82]
E. Y. Shapiro (1982), *Algorithmic program debugging*, MIT Press, Cambridge, Massachusetts.

[Shoham86a]
Y. Shoham (1986), Chronological Ignorance: Time, Knowledge, Nonmonotonicity, and Causation, Computer Science Department, Yale University, April, 24 pages.

[Shoham86b]
Y. Shoham (1986), Chronological Ignorance: time, nonmonotonicity, necessity, and causal theories, *Proceedings of the National Conference for Artificial Intelligence (AAAI-86)* **1**, August 11-15, University of Pennsylvania, Philidelphia, Pennsylvania, 389-393.