Logic-Based Abductive Inference[*]

Sheila A. McIlraith
Knowledge Systems Laboratory
Stanford University
Stanford, CA 94305-9020
sam@ksl.stanford.edu

July 6, 1998

Abstract

This paper surveys the work on abductive inference within the field of artificial intelligence (AI), with particular attention to logic-based abduction. The paper commences with a formal description of three popular characterizations of abductive inference. This is followed by an examination of several specific logic-based abductive frameworks, each of which applies syntactic restrictions to the formulation of the abductive reasoning problem and the resultant explanation. Mechanisms for computing logic-based abductive explanations, and the complexity of variants of the abduction task are examined in the sections to follow. This paper also surveys different applications of abduction in AI, and the connections between abduction and other types of nonmonotonic reasoning. The paper concludes with a discussion of potential future research areas.

[*] Revision of an earlier draft written while the author was a doctoral candidate at the University of Toronto.

Contents

1 Introduction

"'Dr. Watson, Mr. Sherlock Holmes,' said Stamford, introducing us.
"'How are you?' he said cordially, ...'You have been in Afghanistan, I perceive.'
"'How on earth did you know that?' I asked in astonishment."
Later, Sherlock answers the question:
"'You appeared to be surprised when I told you, on our first meeting, that you had come from Afghanistan.'
"'You were told, no doubt.'
"'Nothing of the sort. I knew you came from Afghanistan. From long habit the train of thought ran so swiftly through my mind that I arrived at the conclusion without being conscious of intermediate steps. There were such steps, however. The train of reasoning ran, Here is a gentleman of a medical type, but with the air of a military man. Clearly an army doctor, then. He has just come from the tropics, for his face is dark, and that is not the natural tint of his skin, for his wrists are fair. He has undergone hardship and sickness, as his haggard face says clearly. His left arm has been injured. He holds it in a stiff and unnatural manner. Where in the tropics could an English army doctor have seen such hardship and got his arm wounded? Clearly in Afghanistan. The whole train of thought did not occupy a second....'"
"'It is simple enough as you explain it,' I said, smiling."

– A Study in Scarlet [22]

Unbeknownst to either Sherlock Holmes, or to his creator A. Conan Doyle, Holmes was a master of abductive reasoning, as he illustrates in the excerpt above. For abduction is "the process of forming an explanatory hypothesis" from a set of observations. The term abduction was first introduced by American philosopher Charles S. Peirce in the late 1800s [26] and reintroduced to the artificial intelligence (AI) literature in the 1970s by Harry Pople Jr. [58]. Eugene Charniak popularized the term abduction in his AI textbook [6] by ascribing to Peirce's trichotomy for inference, which includes abduction. In attempting to define a basis for scientific inquiry, Peirce distinguished between three forms of inference: deduction, induction and abduction. Simply described, given the major premise $a \supset b$, the minor premise a, and the conclusion b Peirce asserted that: a *deduction* reasons from $a \supset b$ and a to produce the conclusion b; an *induction* reasons from a and b to produce the plausible rule $a \supset b$; and finally, an *abduction* infers the plausible explanation a, from $a \supset b$ and b. Holmes demonstrates his mastery of abductive inference in the excerpt above by conjecturing that Watson had injured his left arm from the observation $holds\_left\_arm\_stiff\_and\_unnatural$ and the premise $left\_arm\_injured \supset holds\_left\_arm\_stiff\_and\_unnatural$. From this rather simplistic description, abduction looks like a reverse modus ponens inference rule. Peirce viewed it as the selection of a preferred probationary explanation for the occurrence of b that would subsequently be confirmed by scientific process. Within the field of artificial intelligence, abduction is defined as *inference to the best explanation*, without subsequent confirmation. Further, its characterization is not limited to the formulation provided by Peirce.

Definition 1 (Abductive Inference) *Abductive inference is inference to the best abductive explanation.*

In the AI literature, there are three predominant characterizations of abductive inference: a logic-based account (e.g., [57], [24], [11]), a set-covering account (e.g., [48] [1]) and a probabilistic account (e.g., [47], [48]). They differ both is their definition of what constitutes an abductive explanation and consequently what constitutes a best explanation. By far the most prevalent definition of abductive inference is the logic-based characterization of abduction as *theory formation.* Given a background theory and an observation to be explained, an abductive inference conjectures one or more best explanations for the observation from the background theory. The abductive explanation must be *consistent* with the theory and when conjoined to the theory must *entail the observation*. Defining what constitutes a best explanation is a source of debate, often depending upon the specific application domain. Generally, at least some notion of simplicity or minimality is incorporated into the preference criterion. From this description, we see that abduction is unsound, defeasible inference. Explanations are conjectured and may not be true. While consistent with the theory at that point, they may become inconsistent with the addition of further observations. After all, Watson may *not* have come from Afghanistan!

In contrast to the logic-based characterization of abduction is the set-covering account of abduction, which is best represented by the *Parsimonious Covering Theory* (PCT) [48]. PCT uses causal networks to represent the relationship between *disorders* (potential explanation primitives) and *manifestations* (potential observations). Given an observation (one or more manifestations), PCT infers explanations – sets of disorders which account for the observation. Each set of disorders covers the observation and is parsimonious.

The probabilistic account of abduction characterizes abductive inference as the task of finding the most probable explanation for the evidence observed. There are several probabilistic accounts of abductive inference (e.g., [47] [48]). They combine causal networks with some notion of plausibility. Accounts differ in their definition of plausibility.

At an intuitive level, abduction is a form of hypothetical reasoning and as such aids in the characterization of many human reasoning tasks. In fact, it was early work on diagnostic problem solving [58] which kindled interest in abduction within the AI community. In addition to diagnosis (e.g., [6], [58], [57]), abductive inference has been applied to the problems of image understanding (e.g., [14], [54]), plan formation [23], plan recognition [6], temporal reasoning [67], natural language understanding (e.g., [33], [5]), database updates [35] and nonmonotonic reasoning (e.g., [25], [50]). To illustrate the application of abductive inference, consider the problem of medical diagnosis. A background theory may be created to capture the relationship between diseases and their manifested symptoms. Given, the observation of certain symptoms, a disease can be abduced which accounts for the occurrence of those symptoms. Similarly, tasks such as vision, plan recognition and aspects of natural language understanding may all be conceived as abductive reasoning problems. From the variety of applications, we see that the term *explanation* and for that matter *observation* are used loosely, reflecting the origins of abductive inference in diagnostic problem solving. If logical implication (or in the case of causal networks, the link) is used to capture the notion of causation, then the term explanation is apparent, reflecting the cause of the observations. However, an abductive explanation goes beyond the notion of a strictly causal account. For example, a theory about our intrepid detective might state that Holmes either smokes a pipe or a cigar. If we observe that Holmes smokes a pipe, an abductive explanation for this observation is that Holmes does not smoke a cigar, even though we would not take this to be an adequate explanation in the colloquial sense of the word.

This paper surveys the explicit work on abduction within artificial intelligence. Although a survey of the abduction literature would not be complete without a description of all three popular characterizations of abduction and their variants, the focus of this paper will be on logic-based abduction. The notion of abduction was first addressed in the AI literature some twenty years ago, but it has only become widespread within the last ten. Initial work on abduction addressed the mechanization of abductive inference ([58], [14], [61]). Many of these systems were applied to the task of diagnosis. Subsequent work attempted to define and formalize the notion of abductive inference [52], [48] by proposing frameworks and formal characterizations. These formalizations led to the realization that some previous work in AI, de Kleer's ATMS [17] for example, and Genesereth's DART system [30], incorporated procedures from abductive inference. Following formalization of the task of abduction, complexity results were established to confirm informal suspicions that the general task of abductive inference was intractable ([4], [66]). At the same time, abductive inference was being applied to many of the hypothetical reasoning tasks listed above. This resulted in more formal characterization of certain tasks as well as further domain specific definitions of *best* explanation. More recent work on abduction has established the relationship between abduction and other areas of AI including nonmonotonic reasoning.

This survey attempts to identify and synthesize the important contributions of this body of literature. We commence with a more formal presentation of the various characterizations of abductive reasoning. In Section 3, we examine several specific logic-based abductive frameworks. Each framework incorporates some syntactic restrictions on the formulation of the abductive reasoning problem and on the resultant explanation. By restricting the form of the abductive reasoning problem, some interesting results have been proven. Section 4 examines the problem of actually computing logic-based abductive explanations, while Section 5 examines complexity results on generating abductive explanations. Applications of abduction in artificial intelligence are surveyed in Section 6. Abduction has overlaps with many other areas of AI. Section 7 briefly notes some of the fundamental relationships. The survey concludes with a discussion of potential future research areas.

2 Characterizations of Abduction

In this section we expand upon discussion in the introduction and provide a more detailed characterization of predominant accounts of abductive reasoning. We include a logic-based account, a set-covering account and a probabilistic account. The focus of this survey paper is on logic-based abductive reasoning. As a result, subsequent sections will focus on issues related to the logic-based account.

2.1 Logic-based Accounts of Abduction

Abductive inference is inference to the best explanation. The original philosophical formulation coined by Peirce resembled an unsound reverse modus ponens inference rule. Over the years, the AI community has converged upon *theory formation* as the accepted logical characterization of abduction. In this section we present this and other noteworthy logical definitions of abduction.

2.1.1 Abduction as Theory Formation

Abduction is generally defined with respect to an abductive framework. Many such frameworks have been proposed (e.g., [9], [38], [52], [24]). Motivated by the trade-off between expressiveness and ease of computation, they differ with respect to syntactic restrictions on the various components of the framework and on the syntactic form of the resultant abductive explanations.

Below we take liberty with logic and define an intentionally vague *generic* abductive framework from which we can define the concept of logic-based abductive explanation. In Section 3, we provide more explicit definitions with respect to specific abductive frameworks.

A language $\mathcal{L}$ is assumed.

Definition 2 (Generic Abductive Framework)
A generic abductive framework is a pair $(\Sigma, \mathcal{E})$ *where:*

- Σ *is a background theory,*

- *$\mathcal{E}$ is a distinguished set from which explanations are drawn. The elements of $\mathcal{E}$ are sometimes referred to as "abducibles".*

Note that $\mathcal{E}$ can be the entire language $\mathcal{L}$, but is generally restricted to a subset of distinguished literals of the language.

Definition 3 (Abductive Explanation as Theory Formation)
Given an abductive framework $(\Sigma, \mathcal{E})$ and observation O, E, drawn from $\mathcal{E}$ is an abductive explanation for O iff

- *$\Sigma \cup E \models O$, and*
- *$\Sigma \cup E$ is satisfiable.*

The definition of *best* abductive explanation differs from application to application, but at very least is comprised of some expression of minimality or simplicity. Such definitions are syntactic in nature and consequently must be defined in the context of a language. For example,

Definition 4 (LITS (following [41]))
Let $\mathcal{L}$ be a propositional language, p range over the propositional letters of $\mathcal{L}$ and α and β range over the propositional sentences of $\mathcal{L}$.

The literals of α, $LITS(\alpha)$, is defined by: $LITS(p) = \{p\}$; $LITS(\neg\alpha) = \{\overline{m} \mid m \in LITS(\alpha)\}$; $LITS(\alpha \wedge \beta) = LITS(\alpha \vee \beta) = LITS(\alpha) \cup LITS(\beta)$.

Definition 5 (Simpler $\prec$)
E is simpler than E' i.e., $E \prec E'$, iff $LITS(E) \subset LITS(E')$.

Intuitively, this says that E is simpler than E' if it contains a subset of the literals of E'. From these definitions, we define the notion of a minimal abductive explanation, which for many applications suffices as a definition of best abductive explanation. Of course, there may be several "best" explanations in this context. We do not favour the shortest explanations, but favour all explanations that do not contain superfluous literals.

Definition 6 (Minimal Abductive Explanation)
E is a minimal abductive explanation for O iff there is no E' which is an abductive explanation for O and which is simpler than E.

Some specific abductive reasoning frameworks have extended their notion of best to include other preference criteria such as probabilities or priority rankings. These will be presented in the context of the specific frameworks.

2.1.2 Abduction in Terms of Models of Belief

Levesque provides a knowledge-level account of abduction [41] [1]. A knowledge-level account [45] enables abduction to be characterized without concern for the manner in which knowledge is represented, providing flexibility in representation and manipulation techniques at the symbol level. Drawing from previous work on logics of implicit and explicit belief [40], Levesque characterizes

[1] As noted in the paper, the account is not strictly at the knowledge level. The notion of simplicity employed in defining minimal abductive explanations is necessarily syntactic and therefore not at the knowledge level.

abduction in terms of a formal model of belief. Then, by altering the underlying notion of belief, he provides several different forms of abductive reasoning. One of the resultant forms of abduction is equivalent to the theory-formation characterization above. The other provides a limited notion of abductive reasoning which enables computational machinery to find a tractable subset of the theory-formation abductive explanations. The strength of this account over the theory-formation account provided above is the simplicity with which different forms of abductive reasoning may be characterized by changing the notion of belief. Furthermore, this account is very general. It does not suffer the syntactic restrictions of so many of the formulations.

In defining abduction, Levesque provides a standard propositional language $\mathcal{L}$, with the extra constant $\Box$ for falsity. All beliefs are expressed in $\mathcal{L}$. A second language $\mathcal{L}^*$ is employed to talk about what is believed or not believed. It is identical to $\mathcal{L}$ except that atomic sentences are of the form $B\alpha$, where α is a sentence in $\mathcal{L}$. The subscript λ is used to indicate different types of belief, $B_\lambda\alpha$. Both languages are interpreted in the standard fashion. In particular, atomic sentences of $\mathcal{L}^*$ are interpreted with respect to an *epistemic state*, e. Thus $e \models B_\lambda\alpha$ says that $B_\lambda\alpha$ is true at epistemic state e.

The general notion of abductive explanation is defined as follows.

Definition 7 (Abductive Explanation in Terms of Beliefs)
Given belief type λ, α is an abductive explanation for β with respect to epistemic state e (α $expl_\lambda$ β wrt e) iff $e \models [B_\lambda(\alpha \supset \beta) \wedge \neg B_\lambda \neg\alpha]$.

As with the theory-formation definition of abduction minimal, abductive explanations are preferred. The best abductive explanation is the minimal abductive explanation.

Definition 8 (Minimal Abductive Explanation)
Given belief type λ, α is a minimal abductive explanation for β with respect to epistemic state e (α $min\text{–}expl_\lambda$ β wrt e) iff

α $expl_\lambda$ β wrt e and for no $\alpha^* \prec \alpha$ is it the case that α^* $expl_\lambda$ β wrt e,
where $\prec$ is as defined in Definition 5 and $LIT(\Box) = \emptyset$.

Abduction under Implicit and Explicit Belief

Following these general definitions, two specific notions of belief are considered: implicit belief where $B_\lambda = B_I$ and explicit belief where $B_\lambda = B_E$ [39]. Intuitively, something is explicitly believed if it is actively held to be true by an agent. Something is implicitly believed if it follows from what is actively held to be true by an agent.

Implicit belief is the classical notion of belief where beliefs are closed under logical consequence; i.e., $\models (B_I\alpha \wedge B_I(\alpha \supset \beta)) \supset B_I\beta)$. The definition of abductive explanation under implicit belief reduces to the theory-formation characterization of abductive explanation. Specifically,

Proposition 1
α $expl_I$ β wrt e iff $\Sigma \cup \{\alpha\} \models \beta$ and $\Sigma \cup \{\alpha\}$ is consistent, where Σ is some representation of the epistemic state at the symbol level.

Explicit belief is a weaker sense of belief, originally conceived as a more computationally tractable form of belief than implicit belief. It is not closed under logical consequence.

From characterizing abduction in terms of explicit belief, we get a limited notion of abductive explanation which is tractable in certain instances where the theory-formation characterization is not. This will be discussed in further detail in the section on computational complexity.

2.2 Set-covering Accounts of Abduction

The set-covering account of abduction ([48], [1]) is best represented by the Parsimonious Covering Theory (PCT), developed at the University of Maryland [48]. PCT claims to be restricted to explanation problems in the diagnosis domain. Consequently, the abductive framework is defined using diagnostic terminology. PCT uses causal networks to represent the relationship between *disorders* (potential explanation primitives) and *manifestations* (potential observations). The simplest PCT framework contains no intermediate states in the causal network and no numeric notion of plausibility.

Definition 9 (Simple PCT Framework)
A diagnostic problem P is a 4-tuple $< D, M, C, M^+ >$ where:

- $D = \{d_1, d_2, \ldots, d_n\}$ *is a finite, non-empty set of objects called disorders;*
- $M = \{m_1, m_2, \ldots, m_k\}$ *is a finite, non-empty set of objects called manifestations;*
- $C \subseteq D \times M$ *is a relation with* $domain(C) = D$ *and* $range(C) = M$*, called causation; and*
- $M^+ \subseteq M$ *is a distinguished subset of M which is said to be present.*

Definition 10
For any $d_i \in D$ and $m_j \in M$ in a diagnostic problem $P =< D, M, C, M^+ >$,
$effects(d_i) = \{m_j \mid < d_i, m_j > \in C\}$, the set of objects directly caused by d_i; and
$causes(m_j) = \{d_i \mid < d_i, m_j > \in C\}$, the set of objects which can directly cause m_j.

There is an implicit independence assumption, so the effects of a set of disorders is the union of the effects of individual disorders in the set.

Definition 11 (Cover) *The set $D_I \subseteq D$ is said to be a cover for $M_J \subseteq M$ if $M_J \subseteq effects(D_I)$.*

From these definitions, we can define the notion of a PCT abductive explanation.

Definition 12 (PCT Abductive Explanation) *A set $E \subseteq D$ is said to be an abductive explanation for M^+ for a problem $P =< D, M, C, M^+ >$ iff E covers M^+ and E satisfies a given parsimony criterion.*

Like the notion of "best", the definition of parsimony is contentious. The simple PCT framework has settled on the notion of *irredundancy* as the parsimony criterion.

Definition 13 (Parsimony Criterion) *A cover D_I of M_J is said to be irredundant if none of its proper subsets is also a cover of M_J; it is redundant otherwise.*

This simple PCT model has been extended in several ways. There is a corresponding probabilistic account which is described in the following section. Aside from adding plausibility measures, the simple PCT framework has also been extended to include intermediate states between manifestations and disorders.

The simple PCT characterization of abductive reasoning is equivalent to the logic-based theory-formation characterization for a syntactically restricted theory Σ. M^+ corresponds to O, the observation. D, the set of disorders corresponds to $\mathcal{E}$, the set from which explanations are drawn. Finally C, the causal relation between manifestations and disorders corresponds to a syntactically restricted theory Σ. Furthermore, the parsimonious criterion of irredundancy is equivalent to the theory-formation notion of minimality. One drawback of the set-covering approach is that it is limited in its expressive power and it is difficult to measure what impact a change in the theory will have on the explanations.

2.3 Probabilistic Account of Abduction

A potential drawback of both the logic-based and set-covering accounts of abduction is that they can generate a large number of minimal/parsimonious explanations, providing no other means of preference. Probabilistic accounts of abduction address this problem by characterizing the best abductive explanation as the instantiation of explanatory variables that attains the highest plausibility, given the observations.

Probabilistic accounts integrate causal networks with some notion of plausibility. Specific accounts differ in the expressiveness of their causal networks and in their conception of plausibility.

The simple PCT framework described above has been extended to incorporate plausibility. The resultant characterization is referred to as a *probabilistic causal model.* In this model, a prior probability p_i is associated with each disorder d_i. Each causal link from d_i to m_j has a causal strength c_{ij}, associated with it. c_{ij} represents how frequently d_i causes m_j. The relative likelihood $L(D_I, M^+)$ of any potential abductive explanation D_I given the presence of M^+ can be calculated using the relevant p_i's and c_{ij}'s. Although this probabilistic causal model was only developed for simple PCT causal networks, it has been extended by other researchers to apply to causal networks with intermediate states.

Pearl's belief networks [47] provide an alternative probabilistic account of abduction. Simple belief networks are directed acyclic graphs (DAGS) where each node represents a proposition (or variable) and the arcs represent direct dependency whose strength is captured by conditional probabilities. Belief nets differ from the PCT extension in that they use probabilities and Bayes Theorem to calculate the most likely abductive explanation. The interesting thing about Pearl's approach is that instead of making independence assumptions about the relationship between propositions, the probabilistic dependencies are explicitly represented in the structure of the causal network. A common criticism of probabilistic accounts of any sort is that the various probability values are difficult to accurately estimate.

3 Frameworks for Logic-based Abduction

In the previous section we reviewed several different characterizations of abductive inference. In this section we examine specific logic-based abductive reasoning frameworks. These frameworks all view abduction as theory-formation, differing with respect to the syntactic restrictions they place on the various components of the framework and on the form of the resultant best abductive explanations. There are computational advantages to these restrictions which will be investigated in a subsequent section. Here, we focus on several frameworks for abductive inference, interesting properties of those frameworks, and noteworthy contribution towards semantics for abduction or towards definitions of preferred abductive explanations.

The semantics of abduction is not well defined. Abductive inference is nonmonotonic. An inferred explanation may no longer hold with the addition of new information. Our current logic-based theory-formation account of abduction is meta-logical. The nonmonotonicity combined with the meta-logical characterization does not simplify the task of defining a semantics. As we will see, some steps have been made towards defining a semantics for syntactically restricted abductive frameworks.

Another issue that relates to abduction is the lack of a rigorous domain-independent comparator for determining a best abductive explanation. Logic-based characterizations of abduction can produce a large number of minimal abductive explanations and there is a need for further discrimination. In this section we will see several alternative comparators for abductive explanations.

We examine three abductive frameworks: a causal framework, a logic programming framework and the Theorist framework. In so doing, we adopt the notation used by the researchers. The correspondence to our generic logic-based framework is noted in each case. Our objective is to assist the reader when referring to the cited literature.

3.1 Causal Framework

A number of researchers (e.g., Poole ([51], [56]), Console [9], Konolige [38]) have investigated properties of abductive inference for a syntactically restricted class of theories referred to in the literature as *causal theories* (or in some cases, more specifically, *fault theories*). Although very limited in their expressive power, causal abductive frameworks are sufficient for some applications, particularly in the area of diagnosis.

Let $\mathcal{L}$ be a standard propositional language.

Definition 14 (Causal Abductive Framework) *The causal abductive framework is a triple (Causes,Effects,Σ) where:*

- *Causes, a set of atomic sentences of $\mathcal{L}$, is the set of causes;*
- *Effects, a set of atomic sentences of $\mathcal{L}$, is the set of effects we might observe and whose causes we seek as explanations;*
- *Σ, the causal theory is the background theory. It consists of a set of nonatomic definite clauses whose directed graph is acyclic. Causes do not appear at the head of the definite clauses; clauses of Σ are of the form $\neg c_1 \vee \neg c_2 \ldots \vee \neg c_n \vee e$, where $c_i \in$ Causes and $e \in$ Effects.*

This framework differs from the generic logic-based framework not only in the syntactic restrictions it places on Σ, but also in the designation of a distinguished set *Effects* from which all observations O are drawn. *Causes* corresponds to $\mathcal{E}$ in the generic framework.

Definition 15 (Abductive Explanation) *Given a causal framework (Causes,Effects,Σ), an abductive explanation for O, a conjunction of literals drawn from Effects, is C, a conjunction of literals drawn from Causes, such that $\Sigma \cup C \models O$ and $\Sigma \cup C$ is consistent.*

As in Section 2, preference is given to minimal abductive explanations.

Interesting results have come out of investigation of this restricted abductive framework, namely, a correspondence between abduction in the causal framework and consistency-based reasoning when completion axioms augment the background theory of the causal framework. This has enabled abductive inference in causal abductive frameworks to be related to completion semantics, as discussed in the subsection to follow.

Another interesting extension to this (approximate) framework is the incorporation of probabilities enabling a best abductive explanation to be defined probabilistically. Poole has proposed an account of probabilistic horn clause abduction. Further, he has shown a correspondence between his work and Pearl's belief nets. Although beyond the scope of this paper, the interested reader is referred to [55], [56].

3.1.1 Relating Closure to Abduction for Causal Frameworks

The relationship between abduction, closure, consistency-based reasoning and deduction was originally observed by Reiter [63] and described by Poole [51] and Console [9], [10] in the context of diagnostic reasoning with causal theories. It was generalized slightly and stated more clearly by Konolige [38].

To properly describe the results, it is necessary to digress slightly and define consistency-based reasoning. Consistency-based reasoning ([63], [18]) is arguably the most popular alternative to abduction for defining a set of hypotheses which account for an observation. It is commonly applied to diagnosis problems. We contrast a consistency-based hypothesis with an abductive explanation.

Definition 16 (Consistency-based Hypothesis) *Given an abductive framework* $(\Sigma, \mathcal{E})$ *and observation* O*,* E_{cb} *drawn from* $\mathcal{E}$ *is a consistency-based hypothesis for* O *iff* $\Sigma \cup E_{cb} \cup O$ *is consistent.*

Note that the criterion defining a consistency-based hypothesis is less rigorous than the criteria defining an abductive explanation. Every abductive explanation is a consistency-based hypothesis.

As with abductive explanations, preference is given to minimal consistency-based hypotheses (see Section 2).

Proposition 2 *Given a causal abductive framework (Causes,Effects,*Σ*) and observation* $O \in$ *Effects, the minimal abductive explanations for* O *with respect to (Causes,Effects,*Σ*) are equivalent to the minimal consistency-based hypotheses for* O *with respect to (Causes,Effects,*Σ^**), where* Σ^* *is the Clark completion [8] of the background theory* Σ*.*

Intuitively, by completing the background theory with the addition of closure axioms, we are saying that the causes of a particular effect, e are all and only the causes of that effect. For example, if $(c_1 \supset e) \wedge (c_2 \supset e) \wedge \ldots \wedge (c_n \supset e)$, then adding the closure axiom for e results in $e \equiv c_1 \vee c_2 \vee \ldots \vee c_n$.

We can also define the notion of a *cautious explanation*, as the disjunction of all of the minimal abductive explanations.

Definition 17 (Cautious Abductive Explanation) *Given a causal abductive framework (Causes, Effects,* Σ*), a cautious abductive explanation for* O *is a formula* $E_{cautious}$ *such that* $E_{cautious} = \bigvee_i E_i$*,* $\forall E_i$ *such that* E_i *is a minimal abductive explanation for* O*.*

Proposition 3 *Given a causal abductive framework (Causes,Effects,*Σ*),* Σ^**, the Clark completion of* Σ *and observation* $O \in$ *Effects,* $\Sigma^* \cup O \models E_{cautious}$*.*

This is an important result as we will see in the next section because it enables us to compute cautious abductive explanations for causal theories deductively.

Console [10] demonstrates that abductive inference is based on a completion semantics by relating abductive explanations to cautious abductive explanations.

Proposition 4 *Given a causal abductive framework (Causes,Effects,*Σ*), observation* $O \in$ *Effects and cautious abductive explanation,* $E_{cautious}$*. Let* $E \subseteq$ *Causes and* ν *be an assignment of truth values to the abducible atoms, i.e. the elements of Causes, such that*

$\nu(\alpha) = true$ *iff* $\alpha \in E$

Then, E *is an abductive explanation for* O *iff* $\nu \models E_{cautious}$*.*

Stated more clearly, the individual abductive explanations for O are the abducible atoms contained in the different minimal Herbrand models [42] of $\Sigma^* \cup O$.

3.2 Logic Programming Framework

Researchers associated directly or indirectly with Imperial College have defined and studied a logic programming framework for abductive reasoning ([35], [24], [25]). The framework is distinguished from the generic logic-based framework not only by the syntactic restrictions it places on the components of the framework, but by the explicit mention of integrity constraints on the abducibles of the framework. The correspondence between the generic logic-based framework notation described in Section 2 and the one employed by the logic programming framework is as follows: $\Sigma \rightarrow P$; $\mathcal{E} \rightarrow A$; $E \rightarrow \triangle$; $O \rightarrow q$.

Let $\mathcal{L}$ be a first-order language.

Definition 18 (Logic Programming Abductive Framework) *The logic programming abductive framework is a triple (P, A, IC), where:*

- *P is a set of clauses of the form $H \leftarrow L_1, \ldots L_k$, $k \geq 0$ where H is an atom and L_i is a literal.*
- *A is a set of predicate symbols, the abducible predicates. The abducibles, are then all ground atoms with predicate symbols in A.*
- *IC, the integrity constraints, is a set of closed formulae.*

This framework in essence extends a logic program with the inclusion of integrity constraints and distinguished abducibles. The abducibles are the primitives from which explanations are drawn. The integrity constraints are relations on the abducibles.

Definition 19 (Abductive Explanation) *Given a logic programming abductive framework (P, A, IC), an abductive explanation for q is $\triangle \subseteq atoms(A)$ such that $P \cup \triangle \models q$ and $P \cup \triangle$ is consistent and does not violate IC.*

They adopt this definition, but claim it is not satisfactory because the semantics is unclear. Logic programming uses negation-as-failure, not classic negation. Furthermore, the handling of integrity constraints is not sufficiently defined.

To address this, a model-theoretic semantics is given for the logic programming abductive framework and abductive explanation is defined in terms of this semantics. Just as stable model semantics (SMS) can provide a semantics for logic programs with negation-as-failure [29], generalized stable model semantics [36], an extension of stable model semantics provides a semantics for our abductive framework.

Definition 20 (Stable Model) *Let P be a logic program and M a set of atoms from the Herbrand base. Define P_M to be the set of ground horn clauses formed by taking ground(P), in clausal form, and deleting:*

1. *each clause that has a negative literal $\neg l$ in its body, and $l \in M$.*
2. *all negative literals $\neg l$ in the body of clauses, where $l \notin M$.*

M is a stable model for P if M is the minimal model of P_M.

The intuition behind stable model semantics is that we take a logic program with negation and transform it into a ground logic program without negation. M represents the atoms that we believe to be true (which will become the minimal Herbrand model of the transformed program if it is stable). Consequently, if $\neg l$ is in a clause, but $l \in M$, then according to M, we believe $\neg l$ to be false and we can remove any clauses mentioning it. Negative literals $\neg l$ for which $l \notin M$, may be believed to be true, and thus can be removed from the body of clauses in P. If the minimal Herbrand model of the resultant program coincides with our original "beliefs" (M), then M is a stable model for the original program P.

The stable model semantics are generalized to deal with the abduction. The semantics for abductive inference is then achieved by associating a set of general stable models with an abductive framework and characterizing abductive explanation with respect to these models.

Definition 21 (Generalized Stable Model) *Let (P, A, IC) be a logic programming abductive framework, and $\triangle \subseteq atoms(A)$ be a set of abducibles. Then the set $M(\triangle)$ of ground atoms is a* **generalized stable model** *(GSM) for (P, A, IC) iff it is a stable model for the logic program $P \cup \triangle$, it is a model for the integrity constraints IC, and $\triangle = A \cap M(\triangle)$.*

An observation has an abductive explanation if it is true in at least one of the general stable models for the abductive framework.

Definition 22 (Abductive Explanation) *Given a logic programming abductive framework (P,A,IC), and observation q, a unit clause, $\triangle$ is an abductive explanation for q if there exists a generalized stable model $M(\triangle)$ in which q is true.*

Thus, the objective of abductive reasoning is to find a set of abducibles, $\triangle$ such that $M \models q$, $M \models IC$ and M is a stable model of $P \cup \triangle$.

3.3 Theorist Framework

Theorist (e.g., [57], [57], [52]) is by far the best-known and most expressive abductive reasoning framework. It is capable of performing abductive reasoning as well as default reasoning and prediction. Theorist is less syntactically restrictive than some of the other abductive frameworks proposed, particularly in its use of closed first-order formulae as explanations. Conventionally, abductive frameworks use distinguished literals, or conjunctions of distinguished literals as explanations. By enabling flexibility in the syntactic form of explanations, Theorist can use default rules as abductive explanations. This is particularly useful when developing applications where there is uncertainty in the relationships between elements of the domain. For example, in medical diagnosis, appendicitis generally causes pain in the lower right quadrant of the abdomen, but this is not always the case; sometimes the pain is not localized. Allowing *appendicitis* $\supset$ *low_rt_abdomen_pain* as a potential explanation rather than as an axiom in the background theory reflects the defeasible nature of the formula, and enables it to act with *appendicitis* as an abductive explanation for *low_rt_abdomen_pain*. Unfortunately, Theorist's expressiveness makes defining a semantics more difficult and no clear semantics has been defined to date. Also notable, the Theorist framework is implemented as a programming language that sits on top of Prolog, and may be used as a testbed or to develop applications.

Following [52], let $\mathcal{L}$ be a standard first-order language. A formula is a well-formed formula of a language. An instance of a formula refers to a substitution of terms in the language for free variables in the formula. The correspondence between the generic logic-based framework notation described in Section 2 and the one employed by Theorist is as follows: $\Sigma \rightarrow A$; $\mathcal{E} \rightarrow H$; $E \rightarrow D$; $O \rightarrow g$.

Definition 23 (Theorist Abductive Framework) *The Theorist abductive framework is a pair (A, H), where:*

- *a background theory A is a set of closed formulae,*
- *a distinguished set H is a set of (possibly open) formulae which are taken to be the "possible hypotheses" (the primitives which are used to compose explanations).*

Definition 24 (Scenario) *A scenario of (A, H) is a set D of ground instances of elements of H such that $D \cup A$ is consistent.*

A scenario defines a set of hypotheses that qualify to be considered as potential explanations, by virtue of the fact that they are consistent with our background theory.

Definition 25 (Abductive Explanation) *Given a closed formula g, D is an abductive explanation for g from (A, H) if D is a set of ground instance of elements of H such that $D \cup A$ is consistent and $D \cup A \models g$.*

Definition 26 (Extension) *An extension of (A, H) is the set of logical consequences of A together with a maximal (with respect to set inclusion) scenario of (A, H).*

Theorem 1 *There is an explanation of g from (A, H) iff g is in some extension of (A, H).*

As with the generic account, minimality, as defined in Section 2 is the preference criterion for selecting the "best" subset of abductive explanations.

3.3.1 Characterizing the Best Abductive Explanation

Poole and others have proposed several alternative comparators for defining the *best* abductive explanation within the Theorist framework. Some have been proposed for use in other implementations of abduction (e.g., [14]). They are outlined below.

1. Least Presumptive Abductive Explanation:

Definition 27 *Abductive explanation D_1 is less presumptive than abductive explanation $D_1^{'}$ iff $A \cup D_1^{'} \models D_1$.*

That is to say that abductive explanation D_1 assumes less than $D_1^{'}$. The least presumptive explanation is not the explanation of choice for all applications. For example, when abductive inference is applied to the task of diagnosis, we prefer the most specific (most presumptive) diagnosis, while in the case of an abductive learning application we may very well prefer the most general (least presumptive explanation).

2. Most Specific Abductive Explanation:

An alternative notion of best explanation is thus the most specific explanation.

Definition 28 *Abductive explanation D_1 is more specific than abductive explanation $D_1^{'}$ iff $A \cup D_1 \models D_1^{'}$.*

3. Prioritized Abductive Explanation: Brewka [3] augmented the Theorist framework for default reasoning by defining priorities on defaults. Van Arragon [69] extended this definition of priorities to abductive explanations and implemented the extension in Theorist.

Let $H^i, 1 \leq i \leq n$ be a set of (possibly open) formulae representing the set of hypotheses of priority i.

Definition 29 (Potential Prioritized Scenario) *A potential prioritized scenario of $(A, H^1, \ldots, H^n)$ is a set $\{d \in D^i, i = 1, \ldots, n \mid D^i$ is a set of ground instances of elements of H^i and $A \cup D^1 \cup \ldots \cup D^n$ is consistent$\}$.*

Definition 30 (Prioritized Scenario) *A prioritized scenario of $(A, H^1, \ldots, H^n)$ is a potential prioritized scenario that violates no priority constraints.*

Definition 31 (Priority constraints) *Priority constraints are violated in a potential prioritized scenario $(A, H^1, \ldots, H^n)$ iff for any $D^i, 2 \leq i \leq n$, there is a prioritized scenario of $(A, H^1, \ldots, H^{n-1})$ containing $A \cup D^1 \cup \ldots \cup D^{i-1}$ that is inconsistent with D^i.*

Finally, we are able to define a prioritized abductive explanation.

Definition 32 (Prioritized Abductive Explanation) *Given a closed formula g, S is an abductive explanation of g from $(A, H^1, \ldots, H^n)$ if S is a prioritized scenario of $(A, H^1, \ldots, H^n)$, and $A \cup S \models g$.*

4 Computing Abductive Explanations

The previous section highlighted several specific frameworks for abductive reasoning. In this section we examine the computing machinery required to mechanically generate abductive explanations.

Most implementations of abductive inference are based on some form of resolution theorem proving, commonly used for deductive inference. Consequently, it is interesting to briefly contrast abduction with deduction. As observed by Pople [58], deduction determines *whether* a given formula is true, whereas abduction determines *why* a formula is true. Deduction returns yes/no, while abduction returns one or more conjectured formulae, each of which logically accounts for the original formula in question. In some sense, abduction subsumes deduction in that, in the process of determining *why* a formula is true, we must determine *whether* it could be true. Abduction differs from deduction in other ways. Derivation of an abductive explanation does not necessarily terminate the abductive inference process. We may wish to investigate other explanations for an observation. In contrast, once deduction has found a successful proof, it is finished.

The following steps are required to compute an abductive explanation for O from the abductive framework $(\Sigma, \mathcal{E})$:

1. Generate an explanation E for observation O from the background theory Σ.

2. Test the consistency of the explanation with respect to the background theory.

Discussion proceeds in reverse order. We first examine the problem of consistency checking. This is followed by a more lengthy overview of different techniques for generating the explanations themselves.

We assume that explanations are conjunctions of ground literals. The existence of variables in hypotheses is more problematic, because of the required reintroduction of quantifiers into generated explanations containing skolem constants. This in turn can lead to difficulty in consistency checking of hypotheses ([49], [14]). Several solutions have been proposed to deal with the so-called reverse skolemization problem. They will not be discussed here (see [27], [13]). Note also that in the Theorist framework, explanations are not restricted to conjunctions. In particular, Theorist sometimes employs default-rule-style explanations. These are implemented by naming the formulae with atomic sentences and simply using the atomic sentences in the implementation. Consequently, they are covered by our restriction.

4.1 Consistency Checking

$\Sigma \cup E$ is consistent iff $\Sigma \not\vdash \neg E$. The general problem of computing abductive explanations does not look promising at the outset because of this required consistency check. First-order logic is semi-decidable. (i.e., Given first-order proof theory and a closed formula, a proof will be found if the formula is valid, but the proof procedure may not terminate if the formula is not valid.) Consequently, there is no decision procedure for determining the consistency of first-order formulae in general. Fortunately, there are decidable first-order theories. In particular, first order Horn theories without function symbols are decidable. Similarly, some applications with finite domains may be rewritten as propositional theories, which are decidable. There are many examples in practice. If all else fails, consistency checking can be approximated. For example, if after a certain outlay of resources the formulae have not been proven to be inconsistent, then assume that they are consistent. It is up to the developer of an individual application to ensure that consistency checking is decidable either by syntactic restricitons on Σ or by using some reasonable approximation of consistency checking.

4.2 Generating Explanations

The problem of finding an explanation E such that $\Sigma \wedge E \vdash O$ may be computed several different ways. We describe four mechanisms for generating abductive explanations.

4.2.1 Proof-tree Completion

$\Sigma \wedge E \vdash O$ is equivalent to $\Sigma \wedge E \wedge \neg O \vdash \perp$. As such, the problem of generating an abductive explanation for O may be recast as finding a refutation proof for O which employs literals E drawn from $\mathcal{E}$ (e.g., [58], [14], [15], [50], [57]). Currently the most popular mechanism for computing abductive explanations, this technique is often referred to as *proof-tree completion.* The procedure is performed by converting Σ and $\neg O$ to clausal form and using linear resolution to attempt to derive $\perp$. Of course, unless O is trivially explainable, it can't be proven. Consequently, the proof either does not terminate, or if it does, it terminates in so-called *dead ends.* If these dead ends can resolve with elements drawn from $\mathcal{E}$ to complete the proof tree derivation, then the elements of $\mathcal{E}$ employed to complete the partial proof trees are the explanations for O. The Theorist implementation differs slightly in that the potential explanations $\mathcal{E}$, (H in Theorist terminology) are added to the axioms of Σ and rather than deriving dead ends, Theorist merely notes the elements of $\mathcal{E}$ (H) which were employed in deriving $\perp$.

4.2.2 Direct Proof Method

There are several ways of computing abductive explanations using a direct proof method. The term "direct proof method" is often used to refer to the task of *consequence-finding* – finding the consequences of a theory. In the case of abductive inference, $\Sigma \wedge E \vdash O$ can be recast as $\Sigma \vdash \neg E \vee O$ and so we can retrieve the abductive explanations of O by computing the logical consequences of Σ. Similarly, $\Sigma \wedge E \vdash O$ can be recast as $\Sigma \wedge \neg O \vdash \neg E$ (assuming $\Sigma \wedge \neg O$ is consistent). In this case, we can acquire the abductive explanations for O from the logical consequences of $\Sigma \wedge \neg O$. Unfortunately, while resolution is refutation complete (complete for proof-finding), it is not deductively complete and so does not find all the logical consequences of a theory.

Fortunately, in the case of abduction, we are only interested in a subset of the logical consequences of our theories. Specifically, we want the minimal [2] clauses of the form $\neg E \vee O$ and $\neg E$ respectively. Recent advances have been made in developing complete consequence-finding theorem provers for first-order and propositional theories. In particular, Inoue [34] has developed a complete resolution procedure for consequence-finding, generalized to finding only interesting clauses having certain properties. A set of so-called *characteristic clauses* can be defined to specify both a set of distinguished literals from which the characteristic clauses must be drawn and any other conditions to be satisfied. In our case, the characteristic clauses would be of the form $\neg E \vee O$ and $\neg E$ respectively. The augmentation of the theorem prover with a skip rule allows it to focus on generating only the characteristic clauses, rather than generating all minimal logical consequences and further pruning to retrieve the desired subset of clauses.

Finger's RESIDUE system ([28], [34]) used in the implementation of Genesereth's well-known Design Automated Reasoning Tool (DART) is also a first-order consequence-finding procedure. It was employed in the DART system to generate potential diagnosis candidates by direct proof from $\Sigma \wedge \neg O$. The $\neg E$ which were entailed were referred to as the *residues* of the proof procedure. RESIDUE does not focus search as extensively as Inoue's system.

When dealing with propositional theories, the task of finding the minimal logical consequences of a theory is by definition equivalent to computing the prime implicates of that theory.

Definition 33 (Prime implicates) *C is a prime implicate for Σ iff $\Sigma \models C$, and for no proper subset $C^{'}$ of C does $\Sigma \models C^{'}$.*

Much of the formal work on clause management systems ([64], [37]) and on consistency-based and abductive diagnosis has been cast in terms of prime implicates [18]. Any such formalizations can then be realized in the Assumption-based Truth Maintenance System (ATMS) [17].

The ATMS is arguably one of the best-known AI programs. It is frequently used for applications of diagnosis, reasoning in multiple contexts and in our case, abductive reasoning. At the core of the ATMS is the computation of the prime implicates of a propositional Horn theory, Σ [64]. Thus, the ATMS contains a propositional consequence-finding procedure for Σ, restricted to propositional Horn theories. It does a great deal more than this though. The ATMS holds a distinguished set of literals called *assumptions*. For our purposes, these assumptions can be thought of as equivalent to our distinguished set of abducibles $\mathcal{E}$. Given a unit clause query O, the ATMS returns minimal support sets for that query, drawn from the set of assumptions. These minimal support sets are exactly our minimal abductive explanations. We can formally characterize the computation of the ATMS as follows.

[2] We use the term minimal as it was used in Definition 6.

Definition 34 *Given a set of propositional Horn clauses Σ, a set of assumptions $\mathcal{E}$, and propositional symbol O, called the query, the ATMS procedure returns E, a conjunction of literals drawn from $\mathcal{E}$, for all E such that $\neg E \vee O$ is a prime implicate of Σ.*

E is called the minimal support for O with respect to Σ.

Thus, the ATMS is truly a Horn clause abductive reasoning system. A drawback of the ATMS algorithm is that it can only explain observations which are unit clauses. Kean et al. [37] have developed a more extensive procedure for determining prime implicates which enables a generalized clause management system to provide minimal support sets for conjunctions of clauses. Finally, because of its formal characterization, the ATMS task has been used in the complexity analysis of abductive inference.

4.2.3 Deduction on Closed Causal Theories

The final mechanism for generating abductive explanations is restricted to the causal abductive framework described in Section 2. By computing the Clark completion Σ^* of the causal theory Σ, we can compute the abductive explanations deductively. Recall, $\Sigma^* \wedge O \vdash \bigvee E_i$, where each E_i is an individual abductive explanation for O. From this description and the previous discussion of direct proof method, we might conceivably use one of the computation mechanisms described above, tailored to compute the disjunction of explanations, $\bigvee E_i$.

An alternative is to use a model generation theorem prover ([43], [21]). Since the abductive explanations for O are the abducible atoms in the minimal models of $\Sigma^* \wedge O$, we can generate the set of minimal Herbrand models using a model generator and then retrieve the abducible atoms as our abductive explanations.

5 The Complexity of Abduction

The task of computing an abductive explanations in the general case is NP-hard ([66], [65]). Even with a Horn theory, the task of generating one explanation drawn from a set of abducibles $\mathcal{E}$ is NP-hard. Fortunately, tractable subclasses have been defined by placing syntactic restrictions on the expressiveness of Σ or by limiting the notion of abductive reasoning [41].

Contributions towards defining the complexity of abductive reasoning began with investigation of the complexity of truth maintenance systems and in particular, the ATMS [60]. From Definition 34, we know that the ATMS performs propositional Horn clause abduction. As a consequence, some researchers have based their complexity analysis of abduction on the ATMS task. The term *assumption* in the complexity results to follow is derived from the ATMS and denotes the distinguished set of literals $\mathcal{E}$ from which explanations are composed.

From the outset, it was clear that given a set of assumptions $\mathcal{E}$ of size n, there could be an exponential number of explanations which would take exponential time to list. de Kleer [17] argued that in practice, most problems had only a few explanations, but Provan [60] countered by noting that even if a problem didn't have an exponential number of final solutions, it could still have an exponential number of partial solutions. Subsequent research into the complexity of the ATMS demonstrated that the source of complexity is much more deeply rooted than the problem of an exponential number of partial or final solutions.

Selman and Levesque ([66], [65]) analyzed the complexity of abductive reasoning for propositional theories by further examing the ATMS task. In their work, they distinguished between looking for abductive explanations which are drawn from an assumption set $\mathcal{E}$ and abductive explanations which are simply comprised of literals of the language. We refer to the former as assumption-based explanations and the latter as non-assumption-based explanations. In their analysis, Selman and Levesque assume a small number of explanations and focus on the underlying task. The highlights of their results follow.

Proposition 5 *If Σ is a conjunction of arbitrary clauses, the problem of finding any explanation is NP-hard.*

This is because an explanation must be consistent, and so an explanation procedure could be used to test the satisfiability of a set of clauses, which we know to be NP-complete for arbitrary clauses. Fortunately, when Σ is restricted to Horn clauses, a non-assumption-based explanation can be computed efficiently.

Proposition 6 *Given a set of Horn clauses Σ and a letter q, a non-trivial explanation for q can be computed in time $O(kn)$, where k is the number of propositional letters and n is the number of occurrences of literals in Σ.*

This positive result relies on finding clauses in Σ which are explicitly of the form $\neg q_1 \vee \neg q_2 \vee \ldots \vee \neg q_n \vee q$. $q_1 \wedge q_2 \wedge \ldots \wedge q_n$ explains q. It can be minimized in linear time to ensure it doesn't minimize to .

In an effort to extend the analysis to assumption-based explanations, the following negative result is shown. Even for Horn clause theories, it is difficult to find an abductive explanation containing a particular letter.

Proposition 7 *Given a set of Horn clauses Σ and a letters p and q, the problem of generating an explanation for q that contains p is NP-hard.*

This negative result extends to the generation of assumption-based explanations.

Proposition 8 *Given a set of Horn clauses Σ, a set of assumptions A, and a query letter q, finding an assumption-based explanation for q is NP- hard.*

From these results, we conclude that even for Horn clause theories, the task of generating one of the assumption-based explanations discussed throughout this paper is NP-hard. Proposition 6 shows that finding certain non-assumption-based explanations is easy, in particular, explanations that are explicitly represented in Σ as $\neg E \vee O$. Explanations E retrieved from clauses of this form correspond to the the explanations found by limiting abductive reasoning to reasoning with explicit belief, as discussed in Section 2 [41]. Finding an abductive explanation when abductive reasoning is limited to explicit belief is thus a tractable abductive reasoning task.

Though not explicitly investigated by Selman and Levesque, we conjecture that these negative results do not hold for the case when Σ is a positive Horn theory. In particular,

Proposition 9 *Given a set of positive Horn clauses Σ, a set of assumptions A, and a query letter q, a non-trivial explanation for q, drawn from A, can be computed in polynomial time.*

The justification of this proposition is as follows. Selman and Levesque prove Proposition 8 by reduction to the path with forbidden pairs. Since there are no forbidden pairs in a positive Horn theory, we do not have the same difficulty with the consistency of our explanations.

Finally, Bylander et al. [4] also analyzed the complexity of abduction, using a more simplistic definition of abductive explanation. Although their negative results are comparable to those identified above, they further investigated the effect of plausibility and also how the inter-relationship between potential explanations affected the task of generating abductive explanations. In so doing, they identified two new classes of tractable abduction problems relevant to us here.

The following simplified definition of abduction was used in the analysis.

Definition 35 (Abductive Framework) *An abductive framework is a quadruple (D_{all},H_{all},e,pl) where:*

- *D_{all} is a finite set of all the data to be explained,*
- *H_{all} is a finite set of all the individual hypotheses from which abductive explanations are composed,*
- *e is a map from subsets of H_{all} to subsets of D_{all} (H explains e(H)),*
- *pl is a map from subsets of H_{all} to a partially ordered set (H has plausibility pl(H)).*

Definition 36 (Abductive Explanation) *Given an abductive framework (D_{all}, H_{all}, e, pl), $H \subseteq H_{all}$ is an abductive explanation for D_{all} if $e(H) = D_{all}$ and no subset $H^{'}$ of H_{all} exists such that $e(H^{'}) = D_{all}$. e is said to explain D_{all} if $e(H) = D_{all}$, regardless of whether it is minimal.*

The relation e is a simplified representation of what is contained in Σ and is assumed to be simple to compute. Since we are only focussing on logic-based abduction, we will not review the complexity results related to the use of plausibility.

Two tractable abduction problems follow. The intuition behind their tractability is to put constraints on the interaction between potential explanations which in turn constrains search. Although fairly stringent, these restrictions are reasonable in certain domains.

Proposition 10 (Single Fault Assumption) *Given an abductive framework (D_{all}, H_{all}, e, pl), where abductive explanations are restricted to single-element explanations, finding an abductive explanation for D_{all} takes polynomial time (if one exists).*

Proposition 11 (Independent Hypothesis Assumption) *Given an abductive framework (D_{all}, H_{all}, e, pl), where data explained by the set H is the union of the data that each of the individual members of H explains, then finding an abductive explanation for D_{all} takes polynomial time (if one exists).*

This assumption of independent hypotheses reduces the framework to the PCT framework of Section 2.

6 Applications of Abduction

At first glance, abduction does not have much to recommend it. It is intractable for all but the most trivial problems and for many first order theories, the consistency check makes it undecidable. Yet, abduction has been used for a number of different applications as we will see listed below. In its favour, abduction has a logical characterization; several accessible systems in which an application can be implemented (e.g., Theorist, ATMS, Prolog); and some well-defined complexity results, including the definition of some tractable, though fairly trivial abduction problems.

Intuitively, abduction captures the notion of hypothetical reasoning and as such provides us with a means of reasoning hypothetically with incomplete information. The task may be described as follows:

Given a theory Σ, describing some state of affairs, and an an observation O, conjecture some hypotheses which account for the observation in the context of the theory. The hypotheses may be drawn from a predetermined set of potential hypotheses, $\mathcal{E}$.

Before enumerating specific problems that fall within this task description, we examine issues relevant to the application of abduction.

6.1 Issues

There are several important issues that impact the characterization and implementation of a problem using abduction.

The Suitability of Abduction: Given an observation, the expectation that a theory Σ can generate an abductive explanation for it, assumes that the encoding of the domain in Σ has anticipated all possible observations. In the case of logic-based abduction, either E explains an observation, or it does not. For example, in a medical diagnosis setting, if a patient is observed to have swollen glands and be female ($swollen\_glands \wedge female$), $mumps$ is not an abductive explanation, because although $\Sigma \cup mumps \models swollen\_glands$, it does not explain the observation that the patient is female, and thus $mumps$ is rejected as an explanation. $female$ was not anticipated as a symptom by the axiom writer.

This problem can be addressed in several ways. Identifying suitable observations a priori is the simplest solution, but limiting. The causal abductive framework employs this tactic [9]. Defining a notion of partial explanation is also a solution, but introduces another set of problems with respect to differentiating preferred partial explanations. Introducing "dummy causes" ([50], [53], [19]) for unexplainable observations would prevent partial explanations from being rejected as abductive explanations, but would suffer some of the same problems as partial explanations. Probabilistic accounts of abduction overcome the problem of explaining all observations by simply selecting the most likely explanation given the observations [47]. This of course depends on the availability of probabilty distributions and may require a compromise in the expressiveness of the logic.

Another alternative is to substitute consistency-based reasoning for abduction ([18], [12]). As discussed in Section 3, a consistency-based hypothesis accounts for an observation if it is consistent with the observation conjoined to the theory. The set of abductive explanations is a subset of the set of consistency-based hypotheses. Thus, consistency-based reasoning is less exacting. It would not have rejected $mumps$ in the example above. Unfortunately, it also can result in a far greater number of hypotheses. A compromise is to employ consistency-based reasoning with an abductive bias [44] so that hypotheses must at least be consistent with the observation, but are preferred if they explain the observation.

Representing the Space of Hypotheses: In applications with a potentially large number of explanations, simply enumerating all explanations can be expensive. In ([63], [18]) proposals were put forth for characterizing the space of abductive and consistency-based hypotheses in terms of the set minimal or kernel hypotheses. They depend on syntactic restrictions placed on the axioms of Σ.

Knowledge Representation: Axiomatizing the domain is a nontrivial task. In addition to accurately representing the relationship between elements of the domain, the resultant theory must be decidable or the consistency check modified. If closure axioms can be added to the domain, this will simplify computation. The problem is likely to be intractable, so if the problem space is large and time is of concern then some satisfactory notion of limited reasoning should be defined.

Preference: As previously mentioned, logic-based abduction has the potential to produce a large number of minimal explanations. Many applications lend themselves to further problem-specific or domain-specific preference criteria. For example, in the case of diagnosis, we may want the most-specific explanations, or the most-likely explanation. In the case of language understanding, the most coherent explanation [46] may be desirable; whereas, in the case of object recognition, the single-explanation assumption may be invoked. Selecting and incorporating domain-specific preference criteria is an important issue.

6.2 Specific Applications

Abduction provides a characterization for many human reasoning tasks. Examples of application domains which have been characterized using abductive inference are provided below. The list is not exhaustive but serves to identify some of the specific application areas for the interested reader. Although not included in the list, given the relationship of the ATMS to abduction, much of what passes as qualitative reasoning is likely to encompass some form of abductive inference.

Diagnosis is by far the most prevalent application of abduction. In general, given a theory describing how a system malfunctions and some observation of erroneous behavior, abduction conjectures malfunctioning components which explain the observation. Preference is generally given to the minimal explanation and where relevant, to the most specific explanation. Popular application domains are medical diagnosis (e.g., [52], [68]) and logic circuit diagnosis (e.g., [15], [54]).

Abduction has been used both to characterize diagnosis (e.g., [18], [12], [54], [59]) and as a tool for developing diagnosis systems (e.g. [68]). In the diagnosis research community, the most popular procedure for diagnostic reasoning is the General Diagnostic Engine (GDE) [20] which employs the ATMS to compute consistency-based hypotheses. In general, abduction is used with a theory of faulty behavior in order to explain observed faulty behavior of a system. Consistency-based reasoning is used with a theory of correct behavior, to find components of a system which must be malfunctioning given the observation of incorrect behavior. This rule of thumb is not always adhered to though ([18], [44]). In the diagnostic monitoring of a process, abduction can be used with a theory of correct behavior and an observation of correct behavior to generate an explanation that a component must be behaving correctly, given the observation of correct behavior. In this way, components can be systematically observed and exonerated.

Model-based vision is another research area to which abduction is relevant. For example, given a theory of how features in a scene relate to features in an image, abduction can be used to conjecture scene objects which explain features in the image. Examples in the literature include ([14], [54], [31]).

Recently, researchers have used abduction as a tool to characterize the task of plan recognition (e.g., [2], [7], [32]). Given a theory describing how actions relate to goals and an observation of agent action, agent goals can be conjectured to account for the actions. Natural language is a particularly challenging domain, both for abductive plan recognition and for abductive natural language understanding (e.g., [5], [33], [6], [16]).

Finally, as a demonstration of the diversity of abductive inference, some other problems that have been characterized abductively are: user modeling (e.g., [69], [57]), temporal reasoning [67], planning [23] and database updates [35].

7 Abduction and Nonmonotonic Reasoning

Like much of commonsense reasoning, abductive inference is nonmonotonic; however, the relationship between abduction and other forms of nonmonotonic reasoning has not been fully investigated. In Section 3, we saw how abductive inference in a causal framework related to predicate completion. We also saw that by using abducibles to represent default rules, the Theorist framework could perform default reasoning. In this section we briefly highlight the work of several researchers who have drawn correspondences between abductive inference and other forms of nonmonotonic reasoning.

7.1 Abduction and Default Reasoning

Both Poole ([50], [54]) and Selman and Levesque [66] independently noted the strong relationship between abduction and default logic.

Poole showed that default logic could be viewed as theory formation in the same way that abduction is characterized as theory formation. Given a default theory (D, W) with default rules restricted to normal defaults of the form $: \delta/\delta$, $: \delta/\delta$ in Reiter's default logic [62] corresponds exactly to $\delta \in H$ in the Theorist framework.

More generally, Selman and Levesque demonstrated that there was a computational core, namely the Support Selection Task which is shared between abduction and default logic. Recall from Section 4 that the ATMS computes minimal support sets for O with respect to Σ and that these are identical to the minimal abductive explanations for O, given $(\Sigma, \mathcal{E})$. The Support Selection Task as defined below is equivalent to our definition of abductive explanation (Definition 3), for the restricted abductive framework, and thus is equivalent to the ATMS computation, without the minimality requirement.

Definition 37 (Support Selection Task) *Given a set of Horn clauses W, a set of letters $A \subseteq P$, and a letter q, find a set of units clauses α, called a support set, such that the following conditions hold:*

- $W \cup \alpha \models q$
- $W \cup \alpha$ *is consistent, and*
- α *contains only letters from A.*

The task of finding a support set is also NP-hard.

Proposition 12 *Given a Horn theory W, a set of letters $A \subseteq P$, and a letter q, finding a support set for q is NP-hard.*

It is confirmed that the inherent complexity of ATMS-style abductive inference does not depend on the minimization of the explanations, and that the Support Selection Task is at the core of its complexity.

The Support Selection Task can be related to goal-directed default reasoning.

Definition 38 (Goal-directed Default Reasoning) *The task of goal-directed default reasoning is defined as follows: Given an acyclic Horn theory W, a set of elementary defaults D, and a letter q, find an extension of (D, W) that contains q.*

To demonstrate the correspondence to default logic, consider the following result from [63].

Proposition 13 *Let W be a Horn theory, q be a letter, $A \subseteq P$ be a set of letters, and let $D = \{: p/p \mid p \in A\}$. Then, $Th(W \cup \alpha)$ is an extension of (D, W) that contains q iff α is a maximal support set of q.*

Every extension of (D, W) is just $Th(\Sigma \cup \alpha)$. Thus, the Support Selection Task is also at the computational core of goal-directed default logic and as with abduction, the source of complexity. The fact that goal-directed default logic requires maximal support does not impact the complexity.

7.2 Abduction and Negation by Failure

The relationship between abduction and negation by failure (NAF) has also been examined by recasting NAF as an abduction problem [25]. Just as Poole represented the consequent of certain normal defaults as abducibles, negative literals in a logic program can be represented as abducibles.

This is performed in the context of the logic programming abductive framework of Section 3. A logic program P is transformed into a corresponding ground logic program without negation in the logic programming abductive framework (P^*, A, IC). The transformation is as follows:

1. Replace every occurrence of a negative condition $\neg q(x)$ by a new predicate $q^*(x)$. Add q^* to the set of abducibles, A.

2. Define IC as the set of all denial $\leftarrow q(x), q^*(x)$ for all $q^* \in A$

Inference in the abductive framework (P^*, A, IC) produces the same results as NAF in the original program P.

From this brief discussion, there seems to be potential for other nonmonotonic reasoning problems to be reformulated in terms of abductive inference.

8 Future Work

We have examined a wide range of issues relating to abduction: characterizations of abductive inference, frameworks for logic-based abduction, the computation of abductive explanations, the complexity of abduction, applications of abduction, and finally the relationship between abduction and other forms of nonmonotonic reasoning. Although a great deal of research has been done in the area of abductive inference, a number of research problems remain – some big, some small. In what follows we briefly outline a few potential areas of future research.

- **Testing**
  Peirce's original conception of abduction was as inference to a best probationary hypothesis which would be confirmed by scientific experiment. In a similar spirit, we can define the notion of testing for hypothetical reasoning, to assist in identifying a best abductive explanation. In its simplest form, a test consists of a query to the user to ascertain the truth or falsity of a particular predicate which would in turn discriminate a set of (abductive) hypotheses. In more complex domains, such as medical diagnosis, the notion of a test is more involved. A test can be viewed as a knowledge-producing action whose execution may be preconditioned on the performance of other actions which have effects in the world. Of course the design and selection of tests is nontrivial. They should be achievable, provide maximal discriminatory power, and not result in any undesirable side effects in the world; this all conditioned on the notion that any of the (abductive) hypotheses might be true. Generating tests can itself be

viewed as an abductive reasoning task and thus is intractable in the general case. The subject of testing for hypothetical reasoning is rich with research problems which have applications in many diverse areas including diagnosis, active vision and databases.

- **Semantics**
  The semantics of abductive inference, like so many nonmonotonic reasoning systems is unclear. The demonstrated correspondence to theory completion, stable model semantics, as well as the relationship between abduction and other forms of nonmonotonic reasoning, indicates that there is probably much more to be said on this subject.

- **Relationships**
  On a related note, many areas of AI can be reformulated as abductive inference. It is worth investigating the relationships between abduction and other areas of AI, and in turn bringing to bear the computational machinery and complexity results. In particular, the relationship between abduction and nonmonotonic reasoning, abduction and belief revision, and abduction and induction appear worthy of further investigation.

- **Preference**
  The issue of preference was interleaved into discussion of specific abductive reasoning frameworks and some application domains, but was perhaps not given the prominence it deserves. Although it is unlikely that there is an infallible domain independent explanation comparator, there is still work required in examining the representation and computation of preference criteria for abductive inference. In particular, two preference criteria briefly discussed here deserve further attention. The first is coherence. Long cited in the philosophy and cognitive science literature, coherence has been investigated by AI researchers with respect to belief revision and natural language understanding among others. It should play an important role in the selection and persistence of explanations, but as yet, little formal work has been done in this area. The other preference criterion deserving of more research is probability. The incorporation of some notion of probability into logic-based abductive inference is desirable to add discriminating power to the expressiveness of logic.

- **Applications of Abduction**
  Section 6 enumerated some of the many domains in which abductive inference can be applied. Certainly there are other problems to which abduction is relevant. More principled work is needed on the characterization, implementation and analysis of domain specific preference criteria. Additionally, research is needed into how to limit reasoning or trade-off expressiveness to achieve tractability in specific application domains.

- **General Computational Paradigm**
  From the discussion in Section 7, it seems possible that many nonmonotonic reasoning tasks could be reformulated in terms of abductive inference. The relationship between abduction and deduction, as realized through resolution theorem proving indicates that perhaps some general computing paradigm could be developed which would encompasses much of AI problem solving.

References

[1] D. Allemang, M. Tanner, T. Bylander, and J. Josephson. On the computational complexity of hypothesis assembly. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, 1987.

[2] Douglas E. Appelt and Martha E. Pollack. Weighted abduction for plan ascription. Technical Note 491, SRI International, Menlo Park, CA, May 1990.

[3] G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.

[4] T. Bylander, D. Allemang, M. Tanner, and J. Josephson. The computational complexity of abduction. *Artificial Intelligence*, 49:25–60, 1991.

[5] E. Charniak. Motivation analysis, abductive unification and nonmonotonic equality. *Artificial Intelligence*, 34:275–295, 1988.

[6] E. Charniak and D. McDermott. *Introduction to Artificial Intelligence.* Addison-Wesley Publishing Company, 1985.

[7] Eugene Charniak and Robert Goldman. A probabilistic model of plan recognition. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, volume 1, pages 160–165, 1991.

[8] K.L. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 292–322. Plenum Press, New York, 1978.

[9] L. Console, D. Theseider Dupre, and P. Torasso. Abductive reasoning through direct deduction from completed domain models. In Z. Ras, editor, *Methodologies for Intelligent Systems 4*, pages 175–182. North Holland, 1989.

[10] L. Console, D. Theseider Dupre, and P. Torasso. A completion semantics for object-level abduction. In *Proc. AAAI Symposium in Automated Abduction*, 1990.

[11] L. Console, D. Theseider Dupre, and P. Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1, 1991.

[12] L. Console, D. Theseider Dupre, and P. Torasso. Spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 1, 1991.

[13] P. Cox and T. Pietrzykowski. A complete nonredundant algorithm for reversed skolemization. *theoretical computer science*, 28:239–261, 1984.

[14] P. Cox and T Pietrzykowski. Causes for events: their computation and applications. In *Eighth International Conference on Automated Deduction*, pages 608–621, 1986.

[15] P. Cox and T Pietrzykowski. General diagnosis by abductive inference. In *Proc. IEEE Symposium on Logic Programming*, pages 183–189, 1987.

[16] V. Dasigi. parsing=parsimonious covering? In *Proceedings of the Twelth International Joint Conference on Artificial Intelligence*, pages 1031–1036, 1991.

[17] J. de Kleer. An assumption-based tms. *Artificial Intelligence*, 28:127–162, 1986.

[18] J. de Kleer, A.K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56:197–222, 1992.

[19] J. de Kleer and B. Williams. Diagnosis with behavioural modes. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1324–1330, 1989.

[20] J. de Kleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

[21] M. Denecker and D. de Schreye. On the duality of abduction and model generation. In *Proceedings of the Fifth Generation Computer Systems Conference (FGCS'92)*, pages 650–657, 1992.

[22] U. Eco and T. Sebeok. *The Sign of Three: Dupin, Holmes, Peirce.* Indiana University Press, Bloomington, IN, 1983.

[23] K. Eshghi. Abductive planning with event calculus. In *Proc. Fifth International Logic Programming Conference*, 1988.

[24] K. Eshghi and R. Kowalski. Abduction as deduction. Technical report, Department of Computing, Imperial College, London, 1988.

[25] K. Eshghi and R. Kowalski. Abduction compared with negation by failure. In *Proc. Sixth International Logic Programming Conference*, 1989.

[26] C. Hartshorn et al., editor. *Collected Papers of Charles Sanders Peirce.* Harvard University Press, 1931.

[27] G. Ferguson. Identity and skolem functions in resolution-based hypothetical reasoning. Technical report, M.Sc. Thesis, Department of Computing Science, University of Alberta, 1989.

[28] J. Finger and M. Genesereth. Residue: a deductive approach to design synthesis. Technical Report HPP-85-5, Department of Computer Science, Stanford University, 1985.

[29] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proc. Fifth International Logic Programming Conference*, pages 1070–1080, 1988.

[30] M.R. Genesereth. The use of design descriptions in automated diagnosis. *Artificial Intelligence*, 24:411–436, 1984.

[31] M. Gruninger. *to be determined.* PhD thesis, Department of Computer Science, University of Toronto, to appear.

[32] Nicolas Helft and Kurt Konolige. Plan recognition as abduction and relevance. Unpublished MS, January 1991.

[33] J. Hobbs and M. Stickel. Interpretation as abduction. In *Proc. Association for Computational Linguistics*, 1988.

[34] K. Inoue. Consequence-finding based on ordered linear resolution. In *Proceedings of the Twelth International Joint Conference on Artificial Intelligence*, pages 158–164, 1991.

[35] A. C. Kakas and P. Mancarella. Database updates through abduction. In *16th International Conference on Very Large Data Bases*, pages 650–661, 1990.

[36] A. C. Kakas and P. Mancarella. Generalized stable models: a semantics for abduction. In *European Conference on Artificial Intelligence ECAI90*, 1990.

[37] A. Kean and G. Tsiknis. Assumption-based reasoning and clause management systems. *Computational Intelligence*, 8(1):1–24, 1992.

[38] K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53:255–272, 1992.

[39] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of the National Conference on Artificial Intelligence*, pages 198 – 202, 1984.

[40] H. L. Levesque. Foundations of a functional approach to knowledge representation. *Artificial Intelligence*, 23:155–212, 1984.

[41] H. L. Levesque. A knowledge-level account of abduction. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1061–1067, 1989.

[42] J.W. Lloyd. *Foundations of Logic Programming.* Springer Verlag, second edition, 1987.

[43] R. Manthey and F. Bry. Satchmo: a theorem prover implemented in prolog. In *Ninth International Conference on Automated Deduction*, 1988.

[44] S. McIlraith. Further contributions to characterizing diagnosis. *Annals of Mathematics and Artificial Intelligence*, to appear.

[45] A. Newell. The knowledge level. *Artificial Intelligence*, 18:87–127, 1982.

[46] H. Ng and R. Mooney. On the role of coherence in abductive explanation. In *Proceedings of the National Conference on Artificial Intelligence*, pages 337–342, 1990.

[47] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., 1988.

[48] Y. Peng and J. Reggia. *Abductive Inference Models for Diagnostic Problem-solving.* Springer Verlag, 1990.

[49] D. Poole. Variables in hypotheses. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 905–908, 1987.

[50] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36(1):27–47, 1988.

[51] D. Poole. Representing knowledge for logic-based diagnosis. In *Proceedings of the Fifth Generation Computer Systems Conference (FGCS'88)*, pages 1282–1290, 1988.

[52] D. Poole. Explanation and prediction: an architecture for default and abductive reasoning. *Computational Intelligence*, 5:97–110, 1989.

[53] D. Poole. Normality and faults in logic-based diagnosis. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1304–1310, 1989.

[54] D. Poole. A methodology for using a default and abductive reasoning system. *International Journal of Intelligent Systems*, 5:521 –548, 1990.

[55] D. Poole. Representing diagnostic knowledge for probabilistic horn abduction. In *Proceedings of the Twelth International Joint Conference on Artificial Intelligence*, pages 1129–1135, 1991.

[56] D. Poole. Logic programming, abduction and probability. In *Proceedings of the Fifth Generation Computer Systems Conference (FGCS'92)*, pages 530–538, 1992.

[57] D. Poole, R.G. Goebel, and R. Aleliunas. Theorist: a logical reasoning system for defaults and diagnosis. In N. Cercone and G. McCalla, editors, *The Knowledge Frontier: Essays in the Representation of Knowledge*, pages 331–352. Springer Verlag, 1987.

[58] H. Pople. On the mechanization of abductive logic. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pages 147–152, 1973.

[59] C. Preist and K Eshghi. Consistency-based and abductive diagnoses as generalised stable models. In *Proceedings of the Fifth Generation Computer Systems Conference (FGCS'92)*, page to appear, 1992.

[60] G. Provan. Efficiency of multiple-context tmss in scene interpretation. In *Proceedings of the National Conference on Artificial Intelligence*, pages 173–178, 1987.

[61] J. Reggia. Diagnostic expert systems based on set-covering model. *International Journal of Man Machine Studies*, 19:437–460, 1983.

[62] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

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

[64] R. Reiter and J. de Kleer. Foundations for assumption-based truth maintenance systems: Preliminary report. In *Proceedings of the National Conference on Artificial Intelligence*, pages 183–188, 1987.

[65] B. Selman. *Tractable Default Reasoning.* PhD thesis, Department of Computer Science, University of Toronto, 1990.

[66] B. Selman and H.J. Levesque. Abductive and default reasoning: a computational core. In *Proceedings of the National Conference on Artificial Intelligence*, pages 343–348, 1990.

[67] M. Shanahan. Prediction is deduction but explanation is abduction. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1055–1060, 1989.

[68] J. Smith. Red: A red cell antibody identification expert module. *Journal of Medical Systems*, 9:121–138, 1985.

[69] P. van Arragon. Nested default reasoning for user modeling. Technical Report CS-90-25, Department of Computer Science, University of Waterloo, 1990.