Handbook of Knowledge Representation
Edited by F. van Harmelen, V. Lifschitz and B. Porter

DOI: 10.1016/S1574-6526(07)03006-4

Chapter 6

Nonmonotonic Reasoning

Gerhard Brewka, Ilkka Niemelä,
Mirosław Truszczyński

6.1 Introduction

Classical logic is *monotonic* in the following sense: whenever a sentence A is a logical consequence of a set of sentences T, then A is also a consequence of an arbitrary superset of T. In other words, adding information never invalidates any conclusions.

Commonsense reasoning is different. We often draw plausible conclusions based on the assumption that the world in which we function and about which we reason is *normal* and *as expected*. This is far from being irrational. To the contrary, it is the best we can do in situations in which we have only incomplete information. However, as unexpected as it may be, it can happen that our normality assumptions turn out to be wrong. New information can show that the situation actually is abnormal in some respect. In this case we may have to revise our conclusions.

For example, let us assume that Professor Jones likes to have a good espresso after lunch in a campus cafe. You need to talk to her about a grant proposal. It is about 1:00 pm and, under normal circumstances, Professor Jones sticks to her daily routine. Thus, you draw a plausible conclusion that she is presently enjoying her favorite drink. You decide to go to the cafe and meet her there. As you get near the student center, where the cafe is located, you see people streaming out of the building. One of them tells you about the fire alarm that just went off. The new piece of information invalidates the normality assumption and so the conclusion about the present location of Professor Jones, too.

Such reasoning, where additional information may invalidate conclusions, is called *nonmonotonic*. It has been a focus of extensive studies by the knowledge representation community since the early eighties of the last century. This interest was fueled by several fundamental challenges facing knowledge representation such as modeling and reasoning about rules with exceptions or *defaults*, and solving the *frame* problem.

Rules with exceptions

Most rules we use in commonsense reasoning—like *university professors teach*, *birds fly*, *kids like ice-cream*, *Japanese cars are reliable*—have exceptions. The rules describe what is normally the case, but they do not necessarily hold without exception. This is obviously in contrast with universally quantified formulas in first order logic. The sentence

$$\forall x \left(\mathit{prof}(x) \supset \mathit{teaches}(x)\right)$$

simply excludes the possibility of non-teaching university professors and thus cannot be used to represent rules with exceptions. Of course, we can refine the sentence to

$$\forall x \left(\left(\mathit{prof}(x) \wedge \neg \mathit{abnormal}(x)\right) \supset \mathit{teaches}(x)\right).$$

However, to apply this rule, say to Professor Jones, we need to know whether Professor Jones is exceptional (for instance, professors who are department Chairs do not teach). Even if we assume that the unary predicate *abnormal*(.) can be defined precisely, which is rarely the case in practice as the list of possible exceptions is hard—if not impossible—to complete, we will most often lack information to derive that Professor Jones is not exceptional. We want to apply the rule even if all we know about Dr. Jones is that she is a professor at a university. If we later learn she is a department Chair—well, then we have to retract our former conclusion about her teaching classes. Such scenarios can only be handled with a nonmonotonic reasoning formalism.

The frame problem

To express effects of actions and reason about changes in the world they incur, one has to indicate under what circumstances a proposition whose truth value may vary, a *fluent*, holds. One of the most elegant formalisms to represent change in logic, *situation calculus* [89, 88, 112], uses situations corresponding to sequences of actions to achieve this. For instance, the fact that Fred is in the kitchen after walking there, starting in initial situation S_0, is represented as

$$\mathit{holds}\big(\mathit{in}(\mathit{Fred}, \mathit{Kitchen}), \mathit{do}\big(\mathit{walk}(\mathit{Fred}, \mathit{Kitchen}), S_0\big)\big).$$

The predicate *holds* allows us to state that a fluent, here $\mathit{in}(\mathit{Fred}, \mathit{Kitchen})$, holds in a particular situation. The expression $\mathit{walk}(\mathit{Fred}, \mathit{Kitchen})$ is an action, and the expression $\mathit{do}(\mathit{walk}(\mathit{Fred}, \mathit{Kitchen}), S_0)$ is the situation after Fred walked to the kitchen, while in situation S_0.

In situation calculus, effects of actions can easily be described. It is more problematic, however, to describe what does *not* change when an event occurs. For instance, the color of the kitchen, the position of chairs, and many other things remain unaffected by Fred walking to the kitchen. The frame problem asks how to represent the large amount of non-changes when reasoning about action.

One possibility is to use a persistence rule such as: *what holds in a situation typically holds in the situation after an action was performed, unless it contradicts the description of the effects of the action*. This rule is obviously nonmonotonic. Just adding such a persistence rule to an action theory is not nearly enough to solve problems arising in reasoning about action (see Chapters 16–19 in this volume). However, it is an important component of a solution, and so the frame problem has provided a major impetus to research of nonmonotonic reasoning.

About this chapter

Handling rules with exceptions and representing the frame problem are by no means the only applications that have been driving research in nonmonotonic reasoning. Belief revision, abstract nonmonotonic inference relations, reasoning with conditionals, semantics of logic programs with negation, and applications of nonmonotonic formalisms as database query languages and specification languages for search problems all provided motivation and new directions for research in nonmonotonic reasoning.

One of the first papers explicitly dealing with the issue of nonmonotonic reasoning was a paper by Erik Sandewall [115] written in 1972 at a time when it was sometimes argued that logic is irrelevant for AI since it is not capable of representing nonmonotonicity in the consequence relation. Sandewall argued that it is indeed possible, with a moderate modification of conventional (first order) logic, to accommodate this requirement. The basic idea in the 1972 paper is to allow rules of the form

$$A \text{ and Unless } B \Rightarrow C$$

where, informally, C can be inferred if A was inferred and B cannot be inferred. The 1972 paper discusses consequences of the proposed approach, and in particular it identifies that it leads to the possibility of multiple extensions. At about the same time Hewitt published his work on Planner [55], where he proposed using the *thnot* operator for referring to failed inference.

In this chapter we give a short introduction to the field. Given its present scope, we do not aim at a comprehensive survey. Instead, we will describe three of the major formalisms in more detail: default logic in Section 6.2, autoepistemic logic in Section 6.3, and circumscription in Section 6.4. We will then discuss connections between these formalisms. It is encouraging and esthetically satisfying that despite different origins and motivations, one can find common themes.

We chose default logic, autoepistemic logic, and circumscription for the more detailed presentation since they are prominent and typical representatives of two orthogonal approaches: fixed point logics and model preference logics. The former are based on a *fixed point operator* that is used to generate—possibly multiple—sets of acceptable beliefs (called extensions or expansions), taking into account certain consistency conditions. Nonmonotonicity in these approaches is achieved since what is consistent changes when new information is added. Model preference logics, on the other hand, are concerned with nonmonotonic inference relations rather than formation of belief sets. They select some *preferred* or *normal* models out of the set of all models and define nonmonotonic inference with respect to these preferred (normal) models only. Here nonmonotonicity arises since adding new information changes the set of preferred models: models that were not preferred before may become preferred once we learn new facts.

Preference logics and their generalizations are important not only as a broad framework for circumscription. They are also fundamental for studies of abstract nonmonotonic inference relations. In Section 6.5, we discuss this line of research in more detail and cover such related topics as reasoning about conditionals, rational closure, and system Z.

In the last section of the chapter, we discuss the relationship between the major approaches, and present an overview of some other research directions in nonmonotonic reasoning. By necessity we will be brief. For a more extensive treatment of non-monotonic reasoning we refer the reader to the books (in order of appearance) [43, 11, 78, 85, 25, 2, 16, 17, 80].

6.2 Default Logic

Default reasoning is common. It appears when we apply the Closed-World Assumption to derive negative information, when we use inference rules that admit exceptions (rules that hold under the *normality* assumption), and when we use frame axioms to reason about effects of actions. Ray Reiter, who provided one of the most robust formalizations of default reasoning, argued that understanding default reasoning is of foremost importance for knowledge representation and reasoning. According to Reiter defaults are meta-rules of the form "in the absence of any information to the contrary, assume ..." and default reasoning consists of applying them [111].

Usual inference rules sanction the derivation of a formula whenever some other formulas are derived. In contrast, Reiter's defaults require an additional consistency condition to hold. For instance, a default rule *normally, a university professor teaches* is represented in Reiter's default notation as

$$\frac{\mathit{prof}(x) : \mathit{teaches}(x)}{\mathit{teaches}(x)}.$$

It states that if $\mathit{prof}(J)$ is given or derived for a particular ground term J (which may represent Prof. Jones, for instance) and $\mathit{teaches}(J)$ is consistent (there is no information that $\neg\mathit{teaches}(J)$ holds), then $\mathit{teaches}(J)$ can be derived "by default". The key question of course is: consistent with what? Intuitively, $\mathit{teaches}(J)$ has to be consistent with the whole set of formulas which one can "reasonably" accept based on the available information. Reiter's far-reaching contribution is that he made this intuition formal. In his approach, depending on the choice of applied defaults, different sets of formulas may be taken as providing context for deciding consistency. Reiter calls these different sets *extensions*.

One can use extensions to define a *skeptical* inference relation (a formula is skeptically entailed by a default theory if it belongs to *all* of its extensions), or a *credulous* inference relation (a formula is credulously entailed by a default theory if it belongs to *at least one* of its extensions). In many applications such as diagnosis, planning and, more generally in all the situations where defaults model constraints, the extensions themselves are of interest as they represent different solutions to a problem (see Chapter 7 on Answer Sets in this volume).

6.2.1 Basic Definitions and Properties

In default logic, what we are certain about is represented by means of sentences of first-order logic (formulas without free variables). Defeasible inference rules which specify patterns of reasoning that normally hold are represented as defaults. Formally, a default d is an expression

$$\frac{A : B_1, \ldots, B_n}{C}, \tag{6.1}$$

where A, B_i, and C are formulas in first order logic. In this notation, A is the *prerequisite*, $B_1, \ldots, B_n$ are *consistency conditions* or *justifications*, and C is the *consequent*. We denote A, $\{B_1, \ldots, B_n\}$ and C by pre(d), just(d), and cons(d), respectively. To save space, we will also write a default (6.1) as $A : B_1, \ldots, B_n / C$.

Definition 6.1. *A default theory is a pair* (D, W), *where* W *is a set of sentences in first order logic and* D *is a set of defaults.*

A default is *closed* if its prerequisite, justifications, and consequent are sentences. Otherwise, it is *open*. A default theory is *closed* if all its defaults are closed; otherwise, it is *open*. A default theory determines its Herbrand universe. We will interpret open defaults as schemata representing all of their ground instances. Therefore, open default theories are just a shorthand notation for their closed counterparts and so, in this chapter, the term *default theory* always stands for a *closed* default theory.[1]

Before we define extensions of a default theory (D, W) formally, let us discuss properties we expect an extension E of (D, W) to satisfy.

1. Since W represents certain knowledge, we want W to be contained in E, that is, we require that $W \subseteq E$.

2. We want E to be deductively closed in the sense of classical logic, that is, we want $\mathrm{Cn}(E) = E$ to hold, where $\models$ is the classical logical consequence relation and $\mathrm{Cn}(E) = \{A \mid E \models A\}$ denotes the set of logical consequences of a set of formulas E.

3. We use defaults to expand our knowledge. Thus, E should be *closed* under defaults in D: whenever the prerequisite of a default $d \in D$ is in E and all its justifications are consistent with E, the consequent of the default must be in E.

These three requirements do not yet specify the right concept of an extension. We still need some condition of *groundedness* of extensions: each formula in an extension needs sufficient reason to be included in the extension. Minimality with respect to the requirements (1)–(3) does not do the job. Let $W = \emptyset$ and $D = \{\top : a/b\}$. Then $\mathrm{Cn}(\{\neg a\})$ is a minimal set satisfying the three properties, but the theory (D, W) gives no support for $\neg a$. Indeed $W = \emptyset$ and the only default in the theory cannot be used to derive anything else but b.

The problem is how to capture the inference-rule interpretation we ascribe to defaults. It is not a simple matter to adjust this as defaults have premises of two different types and this has to be taken into account. Reiter's proposal rests on an observation that given a set S of formulas to use when testing consistency of justifications, there is a *unique* least theory, say $\Gamma(S)$, containing W, closed under classical provability and also (in a certain sense determined by S) under defaults. Reiter argued that for a theory S to be grounded in (D, W), S must be precisely what (D, W) implies, given that S is used for testing the consistency of justifications, and used this property to define extensions [111].

[1] We note, however, that Reiter treats open defaults differently and uses a more complicated method to define extensions for them. A theory of open default theories was developed by [73]. Some problems with the existing treatments of open defaults are discussed in [5].

Definition 6.2 (Default logic extension). *Let (D, W) be a default theory. The operator Γ assigns to every set S of formulas the smallest set U of formulas such that*:

1. $W \subseteq U$,
2. $\mathrm{Cn}(U) = U$,
3. *if* $A : B_1, \ldots, B_n/C \in D$, $U \models A$, $S \not\models \neg B_i$, $1 \leqslant i \leqslant n$, *then* $C \in U$.

A set E of formulas is an extension of (D, W) if and only if $E = \Gamma(E)$, that is, E is a fixed point of Γ.

One can show that such a least set U exists so the operator Γ is well defined. It is also not difficult to see that extensions defined as fixed points of Γ satisfy the requirements (1)–(3).

In addition, the way the operator Γ is defined also guarantees that extensions are grounded in (D, W). Indeed, $\Gamma(S)$ can be characterized as the set of all formulas that can be derived from W by means of classical derivability and by using those defaults whose justifications are each consistent with S as additional *standard* inference rules (once every justification in a default d turns out to be consistent with S, the default d starts to function as the inference rule $\mathrm{pre}(d)/\mathrm{cons}(d)$, other defaults are ignored). This observation is behind a quasi-inductive characterization of extensions, also due to Reiter [111].

Theorem 6.1. *Let (D, W) be a default theory and S a set of formulas. Let*

$$E_0 = W, \quad \textit{and for } i \geqslant 0$$
$$E_{i+1} = \mathrm{Cn}(E_i) \cup \{C \mid A : B_1, \ldots, B_n/C \in D, E_i \models A, S \not\models \neg B_i,\ 1 \leqslant i \leqslant n\}.$$

Then $\Gamma(S) = \bigcup_{i=0}^{\infty} E_i$. Moreover, a set E of formulas is an extension of (D, W) if and only if $E = \bigcup_{i=0}^{\infty} E_i$.

The appearance of E in the definition of E_{i+1} is what renders this alternative definition of extensions non-constructive. It is, however, quite useful. Reiter [111] used Theorem 6.1 to show that every extension of a default theory (D, W) can be represented as the logical closure of W and the consequents of a subset of defaults from D.

Let E be a set of formulas. A default d is *generating* for E if $E \models \mathrm{pre}(d)$ and, for every $B \in \mathrm{just}(d)$, $E \not\models \neg B$. If D is a set of defaults, we write $GD(D, E)$ for the set of defaults in D that are generating for E.

Theorem 6.2 (Generating defaults). *Let E be an extension of a default theory (D, W). Then* $E = \mathrm{Cn}(W \cup \{\mathrm{cons}(d) \mid d \in GD(D, E)\})$.

This result is fundamental for algorithms to compute extensions. We will come back to this issue later. For now, we will restrict ourselves to a few examples. Let

$$D_1 = \{\mathit{prof}(x) : \mathit{teaches}(x)/\mathit{teaches}(x)\},$$

$$W_1 = \{prof(J)\}.$$

We recall that we interpret an open default as the set of its ground instantiations. Since there is only one constant (J) in the theory, the corresponding *closed* default theory is

$$D'_1 = \{prof(J) : teaches(J)/teaches(J)\},$$
$$W_1 = \{prof(J)\}.$$

By Theorem 6.2 an extension is the deductive closure of W and some of the available default consequents. Hence, there are only two candidates for an extension here, namely $S_1 = \text{Cn}(\{prof(J)\})$ and $S_2 = \text{Cn}(\{prof(J), teaches(J)\})$. We can now use Theorem 6.1 to compute $\Gamma(S_1)$. Clearly, $E_0 = \text{Cn}(W_1)$. Since $teaches(J)$ is consistent with S_1 and $E_0 \models prof(J)$, $E_1 = \text{Cn}(\{prof(J), teaches(J)\})$. Moreover, for every $i > 2$, $E_i = E_1$. Thus, $\Gamma(S_1) = \text{Cn}(\{prof(J), teaches(J)\})$. Since $teaches(J) \notin S_1$, $S_1 \neq \Gamma(S_1)$ and so, S_1 is not an extension of (D_1, W_1) (nor of (D'_1, W_1)). On the other hand, the same argument shows that $\Gamma(S_2) = \text{Cn}(\{prof(J), teaches(J)\})$. Thus, $S_2 = \Gamma(S_2)$, that is, S_2 is an extension of (D_1, W_1) (and also (D'_1, W_1)).

Now let us consider a situation when Professor J is not a typical professor.

$$D_2 = D_1,$$
$$W_2 = \{prof(J), chair(J), \forall x.(chair(x) \supset \neg teaches(x))\}.$$

As before, there are two candidates for extensions, namely $S_1 = \text{Cn}(W_2)$ and $S_2 = \text{Cn}(W_2 \cup \{teaches(J)\})$. This time S_2 is inconsistent and one can compute, using Theorem 6.1, that $\Gamma(S_2) = \text{Cn}(W_2)$. Thus, S_2 is not a fixed point of Γ and so not an extension. On the other hand, $\Gamma(S_1) = \text{Cn}(W_2)$ and so S_1 is an extension of (D_2, W_2). Consequently, this default theory supports the inference that Professor J does not teach (as it should).

Finally, we will consider what happens if the universally quantified formula from W_2 is replaced by a corresponding default rule:

$$D_3 = \{prof(x) : teaches(x)/teaches(x), chair(x) : \neg teaches(x)/\neg teaches(x)\},$$
$$W_3 = \{prof(J), chair(J)\}.$$

The corresponding closed default theory has two defaults: $prof(J) : teaches(J)/teaches(J)$ and $chair(J) : \neg teaches(J)/\neg teaches(J)$. Thus, there are now four candidates for extensions:

$$\text{Cn}(\{prof(J), chair(J)\}),$$
$$\text{Cn}(\{prof(J), chair(J), teaches(J)\}),$$
$$\text{Cn}(\{prof(J), chair(J), \neg teaches(J)\}),$$
$$\text{Cn}(\{prof(J), chair(J), teaches(J), \neg teaches(J)\}).$$

In each case, one can compute the value of the operator Γ and check the condition for an extension. In this example, the second and third theories happen to be extensions. Since the theory offers no information whether Professor J is a typical professor or a typical chair (she cannot be both as this would lead to a contradiction), we get two extensions. In one of them Professor J is a typical professor and so teaches, in the other one she is a typical chair and so, does not teach.

Default theories can have an arbitrary number of extensions, including having no extensions at all. We have seen examples of default theories with one and two extensions above. A simple default theory without an extension is

$$\big(\{\top : \neg a/a\}, \emptyset\big).$$

If a deductively closed set of formulas S does not contain a, then S is not an extension since the default has not been applied even though $\neg a$ is consistent with S. In other words, $\Gamma(S)$ will contain a and thus $\Gamma(S) \neq S$. On the other hand, if S contains a, then $\Gamma(S)$ produces a set not containing a (more precisely: the set of all tautologies) since the default is inapplicable with respect to S. Again S is not an extension.

Theorem 6.2 has some immediate consequences.

Corollary 6.3. *Let (D, W) be a default theory.*

1. *If W is inconsistent, then (D, W) has a single extension, which consists of all formulas in the language.*
2. *If W is consistent and every default in D has at least one justification, then every extension of (D, W) is consistent.*

We noted that the minimality with respect to the requirements (1)–(3) we discussed prior to the formal definition of extensions does not guarantee groundedness. It turns out that the type of groundedness satisfied by extensions ensures their minimality and, consequently, implies that they form an antichain [111].

Theorem 6.4. *Let (D, W) be a default theory. If E is an extension of (D, W) and E' is a theory closed under classical consequence relation and defaults in D such that $E' \subseteq E$, then $E' = E$. In particular, if E and E' are extensions of (D, W) and $E \subseteq E'$, then $E = E'$.*

6.2.2 Computational Properties

The key reasoning problems for default logic are deciding *sceptical and credulous inference* and finding extensions. For first-order default logic these problems are not even semi-decidable [111]. This is different from classical first order logic which is semi-decidable. Hence, automated reasoning systems for first order default logic cannot provide a similar level of completeness as classical theorem provers: a formula can be a (nonmonotonic) consequence of a default theory but no algorithm is able to establish this. This can be compared to first order theorem proving where it can be guaranteed that for each valid formula a proof is eventually found.

Even in the propositional case extensions of a default theory are infinite sets of formulas. In order to handle them computationally we need characterizations in terms of formulas that appear in (D, W). We will now present two such characterizations which play an important role in clarifying computational properties of default logic and in developing algorithms for default reasoning.

We will write Mon(D) for the set of standard inference rules obtained by dropping justifications from defaults in D:

$$\mathrm{Mon}(D) = \left\{ \frac{\mathrm{pre}(d)}{\mathrm{cons}(d)} \;\middle|\; d \in D \right\}.$$

We define $\mathrm{Cn}_{\mathrm{Mon}(D)}(.)$ to be the consequence operator induced by the classical consequence relation and the rules in Mon(D). That is, if W is a set of sentences, $\mathrm{Cn}_{\mathrm{Mon}(D)}(W)$ is the closure of W with respect to classical logical consequence and the rules Mon(D) (the least set of formulas containing W and closed under the classical consequence relation and the rules in Mon(D)).

The first characterization of extensions is based on the observation that extensions can be described in terms of their generating defaults (Theorem 6.2). The details can be found in [85, 114, 5]. We will only state the main result. The idea is to project the requirements we impose on an extension to a set of its generating defaults. Thus, a set of generating defaults should be grounded in W, which means that for every default in this set the prerequisite should be derivable (in a certain specific sense) from W. Second, the set of generating defaults should contain *all* defaults that apply.

Theorem 6.5 (Extensions in terms of generating defaults). *A set E of formulas is an extension of a default theory (D, W) if and only if there is a set $D' \subseteq D$ such that* $E = \mathrm{Cn}(W \cup \{\mathrm{cons}(d) \mid d \in D'\})$ *and*

1. *for every* $d \in D'$, $\mathrm{pre}(d) \in \mathrm{Cn}_{\mathrm{Mon}(D')}(W)$,
2. *for all* $d \in D$: $d \in D'$ *if and only if* $\mathrm{pre}(d) \in \mathrm{Cn}(W \cup \{\mathrm{cons}(d) \mid d \in D'\})$ *and for all* $B \in \mathrm{just}(d)$, $\neg B \notin \mathrm{Cn}(W \cup \{\mathrm{cons}(d) \mid d \in D'\})$.

The second characterization was introduced in [98] and is focused on justifications. The idea is that default rules are inference rules guarded with consistency conditions given by the justifications. Hence, it is the set of justifications that determines the extension and the rest is just a monotonic derivation.

We denote by just(D) the set of all justifications in the set of defaults D. For a set S of formulas we define

$$\mathrm{Mon}(D, S) = \{\mathrm{pre}(d)/\mathrm{cons}(d) \mid d \in D,\ \mathrm{just}(d) \subseteq S\}$$

as the set of monotonic inference rules enabled by S. A set of justifications is called *full* with respect to the default theory if it consists of the justifications which are consistent with the consequences of the monotonic inference rules enabled by the set.

Definition 6.3 (Full sets). *For a default theory (D, W), a set of justifications* $S \subseteq \mathrm{just}(D)$ *is (D, W)-full if the following condition holds: for every* $B \in \mathrm{just}(D)$, $B \in S$ *if and only if* $\neg B \notin \mathrm{Cn}_{\mathrm{Mon}(D,S)}(W)$.

For each full set there is a corresponding extension and for each extension a full set that induces it.

Theorem 6.6 (Extensions in terms of full sets). *Let* (D, W) *a default theory.*

1. *If* $S \subseteq \text{just}(D)$ *is* (D, W)*-full, then* $\text{Cn}_{\text{Mon}(D,S)}(W)$ *is an extension of* (D, W)*.*
2. *If* E *is an extension of* (D, W)*, then* $S = \{B \in \text{just}(D) \mid \neg B \notin E\}$ *is* (D, W)*-full and* $E = \text{Cn}_{\text{Mon}(D,S)}(W)$*.*

Example 6.1. Consider the default theory (D_3, W_3), where

$$D_3 = \big\{\mathit{prof}(J) : \mathit{teaches}(J)/\mathit{teaches}(J),$$
$$\mathit{chair}(J) : \neg\mathit{teaches}(J)/\neg\mathit{teaches}(J)\big\},$$
$$W_3 = \big\{\mathit{prof}(J), \mathit{chair}(J)\big\}.$$

The possible (D_3, W_3)-full sets are the four subsets of $\{\mathit{teaches}(J), \neg\mathit{teaches}(J)\}$. It is easy to verify that from these only $\{\mathit{teaches}(J)\}$ and $\{\neg\mathit{teaches}(J)\}$ satisfy the fullness condition given in Definition 6.3. For instance, for $S = \{\neg\mathit{teaches}(J)\}$

$$\text{Mon}(D_3, S) = \left\{\frac{\mathit{chair}(J)}{\neg\mathit{teaches}(J)}\right\}$$

and $\text{Cn}_{\text{Mon}(D_3,S)}(W_3) = \text{Cn}(\{\mathit{prof}(J), \mathit{chair}(J), \neg\mathit{teaches}(J)\})$. As required we have $\neg\neg\mathit{teaches}(J) \notin \text{Cn}_{\text{Mon}(D_3,S)}(W_3)$ and $\neg\mathit{teaches}(J) \in \text{Cn}_{\text{Mon}(D_3,S)}(W_3)$.

The finitary characterization of extensions in Theorems 6.5 and 6.6 reveal important computational properties of default logic. A direct consequence is that propositional default reasoning is *decidable* and can be implemented in *polynomial space*. This is because the characterizations are based on classical reasoning which is decidable in polynomial space in the propositional case.

In order to contrast default logic more sharply to classical logic we consider a (hypothetical) setting where we have a highly efficient theorem prover for propositional logic and, hence, are able to decide classical consequences of a set of formulas W and inference rules R, that is $\text{Cn}_R(W)$, efficiently. Theorems 6.5 and 6.6 suggest that even in this setting constructing an extension of a propositional default theory involves a non-trivial search problem of finding a set of generating defaults or a full set. However, the characterizations imply an upper bound on the computational complexity of propositional default reasoning showing that it is on the second level of the polynomial hierarchy.[2] It turns out this is a tight upper bound as deciding extension existence and credulous inference are actually Σ_2^{P}-complete problems and sceptical inference is Π_2^{P}-complete [51, 127].

The completeness results imply that (propositional) default reasoning is strictly harder than classical (propositional) reasoning unless the polynomial hierarchy collapses which is regarded unlikely. This means that there are two orthogonal sources of complexity in default reasoning. One source originates from classical logic on top of which default logic is built. The other source is related to nonmonotonicity of default rules. These sources are independent of each other because even if we assume

[2] For an introduction to computational complexity theory and for basic definitions and results on polynomial hierarchy, see, for example, [46, 103].

that we are able to decide classical consequence in one computation step, deciding a propositional default reasoning problem remains on the difficulty level of an NP/co-NP-complete problem and no polynomial time algorithms are known even under this assumption. Hence, it is highly unlikely that general default logic can be implemented on top of a classical theorem prover with only a polynomial overhead.

In order to achieve tractable reasoning it is not enough to limit the syntactic form of allowed formulas because this affects only one source of complexity but also the way default rules interact needs to be restricted. This is nicely demonstrated by complexity results on restricted subclasses of default theories [60, 126, 8, 100]. An interesting question is to find suitable trade-offs between expressive power and computational complexity. For example, while general default logic has higher computational complexity, it enables very compact representation of knowledge which is exponentially more succinct than when using classical logic [50].

A number of decision methods for general (propositional) default reasoning have been developed. Methods based on the characterization of extensions in terms of generating defaults (Theorem 6.2) can be found, for example, in [85, 5, 114, 30], and in terms of full sets (Theorem 6.6) in [98]. There are approaches where default reasoning is reduced into another problem like a truth maintenance problem [59] or a constraint satisfaction problem [8]. An interesting approach to provide proof theory for default reasoning based on sequent calculus was proposed in [18, 19]. More details on automating default reasoning can be found also in [36].

Notice that for general default reasoning it seems infeasible to develop a fully goal-directed procedure, that is, a procedure which would examine only those parts of the default theory which are somehow syntactically relevant to a given query. This is because extensions are defined with a global condition on the whole theory requiring that each applicable default rule should be applied. There are theories with no extensions and in the worst case it is necessary to examine every default rule in order to guarantee the existence of an extension. For achieving a goal-directed decision method, one can consider a weaker notion of extensions or syntactically restricted subclasses of default theories such as normal defaults (see below) [117, 118].

6.2.3 Normal Default Theories

By restricting the form of defaults one obtains special classes of default theories. One of the most important of them is the class of *normal* default theories, where all defaults are of the form

$$\frac{A : B}{B}.$$

The distinguishing feature of normal default theories is that they are guaranteed to have extensions and extensions are determined by enumerations of the set of defaults. Let (D, W) be a normal default theory (as always, assumed to be closed) and let $D = \{d_1, d_2, \ldots\}$.

1. We define $E_0 = \mathrm{Cn}(W)$;

2. Let us assume E_i has been defined. We select the first default $d = A : B/B$ in the enumeration such that $E_i \models A$, $E_i \not\models B$ and $E_i \not\models \neg B$ and define $E_{i+1} = \mathrm{Cn}(E_i \cup \{B\})$. If no such default exists, we set $E_{i+1} = E_i$.

Theorem 6.7. *Let* (D, W) *be a normal default theory. Then, for every enumeration* $D = \{d_1, d_2, \ldots\}$, $E = \bigcup_{i=1}^{\infty} E_i$ *is an extension of* (D, W) *(where* E_i *are sets constructed above). Furthermore, for every extension* E *of* (D, W) *there is an enumeration, which yields sets* E_i *such that* $E = \bigcup_{i=1}^{\infty} E_i$.

Theorem 6.7 not only establishes the existence of extensions of normal default theories but it also allows us to derive several properties of extensions. We gather them in the following theorem.

Theorem 6.8. *Let* (D, W) *be a normal default theory. Then,*

1. *if* $W \cup \{\mathrm{cons}(d) \mid d \in D\}$ *is consistent, then* $\mathrm{Cn}(W \cup \{\mathrm{cons}(d) \mid d \in D\})$ *is a unique extension of* (D, W);
2. *if* E_1 *and* E_2 *are extensions of* (D, W) *and* $E_1 \neq E_2$, *then* $E_1 \cup E_2$ *is inconsistent*;
3. *if* E *is an extension of* (D, W), *then for every set* D' *of normal defaults, the normal default theory* $(D \cup D', W)$ *has an extension* E' *such that* $E \subseteq E'$.

The last property is often called the *semi-monotonicity* of normal default logic. It asserts that adding normal defaults to a normal default theory does not destroy existing extensions but only possibly augments them.

A default rule of the form

$$\frac{\top : B_1, \ldots, B_n}{C}$$

is called *prerequisite-free*. Default theories possessing only prerequisite-free normal defaults are called *supernormal*. They are closely related to a formalism for non-monotonic reasoning proposed by Poole [107] and so, are sometimes called *Poole defaults*. We will not discuss Poole's formalism here but only point out that the connection is provided by the following property of supernormal default theories.

Theorem 6.9. *Let* (D, W) *be a supernormal default theory such that* W *is consistent. Then,* E *is an extension of* (D, W) *if and only if* $E = \mathrm{Cn}(W \cup \{\mathrm{cons}(d) \mid d \in C\})$, *where* C *is a maximal subset of* D *such that* $W \cup \{\mathrm{cons}(d) \mid d \in C\}$ *is consistent. In particular, if* E *is an extension of* (D, W), *then for every* $d \in D$, $\mathrm{cons}(d) \in E$ *or* $\neg\,\mathrm{cons}(d) \in E$.

Normal defaults are sufficient for many applications (cf. our discussion of CWA below). However, to represent more complex default reasoning involving interactions among defaults, non-normal defaults are necessary.

6.2.4 Closed-World Assumption and Normal Defaults

The *Closed-World Assumption* (or *CWA*, for short) was introduced by Reiter in [110] in an effort to formalize ways databases handle negative information. It is a defeasible inference rule based on the assumption that a set W of sentences designed to represent an application domain determines *all* ground atomic facts that hold in it (*closed-world* *assumption*). Taking this assumption literally, the CWA rule infers the *negation* of every ground atom not implied by W. Formally, for a set W of sentences we define

$$CWA(W) = W \cup \{\neg a \mid a \text{ is a ground atom and } W \not\models a\}.$$

To illustrate the idea, we will consider a simple example. Let GA be the set of all ground atoms in the language and let $W \subseteq GA$. It is easy to see that

$$CWA(W) = W \cup \{\neg a \mid a \in GA \setminus W\}.$$

In other words, CWA derives the negation of every ground atom not in W. This is precisely what happens when databases are queried. If a fact is not in the database (for instance, there is no information about a direct flight from Chicago to Dallas at 5:00 pm on Delta), the database infers that this fact is false and responds correspondingly (there is *no* direct flight from Chicago to Dallas at 5:00 pm on Delta).

We note that the database may contain errors (there may in fact be a flight from Chicago to Dallas at 5:00 pm on Delta). Once the database is fixed (a new ground atom is included that asserts the existence of the flight), the derivation sanctioned previously by the CWA rule, will not longer be made. It is a classic example of defeasible reasoning!

In the example above, CWA worked precisely as it should, and resulted in a consistent theory. In many cases, however, the CWA rule is too strong. It derives too many facts and yields an inconsistent theory. For instance, if $W = \{a \vee b\}$, where a, b are two ground atoms, then

$$W \not\models a \quad \text{and} \quad W \not\models b.$$

Thus, $CWA(W) = \{a \vee b, \neg a, \neg b\}$ is inconsistent. The question whether $CWA(W)$ is consistent is an important one. We note a necessary and sufficient condition given in [85].

Theorem 6.10. *Let W be a set of sentences. Then $CWA(W)$ is consistent if and only if W has a least Herbrand model.*

If W is a set of ground atoms (the case discussed above) or, more generally, a consistent Horn theory, then W has a least Herbrand model. Thus, we obtain the following corollary due to Reiter [110].

Corollary 6.11. *If W is a consistent Horn theory, then $CWA(W)$ is consistent.*

The main result of this section shows that CWA can be expressed by means of *supernormal* defaults under the semantics of extensions. For a ground atom a we define a supernormal default

$$cwa(a) = \frac{\top : \neg a}{\neg a}$$

and we set

$$D_{CWA} = \big\{cwa(a) \mid a \in GA\big\}.$$

We have the following result [85].

Theorem 6.12. *Let W be a set of sentences.*

1. *If $CWA(W)$ is consistent, then* $\mathrm{Cn}(CWA(W))$ *is the unique extension of the default theory (D_{CWA}, W).*

2. *If (D_{CWA}, W) has a unique consistent extension, then $CWA(W)$ is consistent and* $\mathrm{Cn}(CWA(W))$ *is this unique extension of (D_{CWA}, W).*

6.2.5 Variants of Default Logic

A number of modifications of Reiter's default logic have been proposed in the literature which handle several examples differently. We present some of them briefly here.

To guarantee existence of extensions, Lukaszewicz [77] has defined a default logic based on a two-place fixed point operator. The first argument contains the believed formulas, the second is used to keep track of justifications of applied defaults. A default is only applied if its consequent does not contradict the justification of any other applied default. Then, E is an extension if and only if there is a set S_E such that (E, S_E) is a fixed point. Lukaszewicz showed that, in his logic, both existence of extensions and semi-monotony are satisfied.

In [22], a cumulative version of default logic is presented. The basic elements of this logic are so-called *assertions* of the form (p, Q), in which p is a formula, and Q the set of consistency conditions needed to derive p. A default can only be applied in an extension if its justifications are jointly consistent with the extension and with all justifications of other applied defaults. The logic is called cumulative as the inference relation it determines satisfies the property of *Cumulativity* [79], now more commonly called *Cautious Monotony* (cf. Section 6.5).

Joint consistency is also enforced in variants of default logic called *constrained default logics*, which have been proposed independently by [116] and [31] (see also [32]). The major difference between cumulative default logic and these two variants is that the latter work with standard formulas and construct an additional single set containing all consistency conditions of applied defaults, whereas cumulative default logic keeps track of this information in the assertions.

A number of researchers have investigated default theories with preferences among the defaults, e.g., [85, 6, 23, 113, 26]. For a comparison of some of these approaches the reader is referred to [119]. Finally, [23] contains an approach in which reasoning not only with, but also about priorities is possible. In this approach, the preference information is represented in the logical language and can thus be derived and reasoned upon dynamically. This makes it possible to describe conflict resolution strategies declaratively and has interesting applications, for instance, in legal reasoning.

6.3 Autoepistemic Logic

In this section, we discuss autoepistemic logic, one of the most studied and influential nonmonotonic logics. It was proposed by Moore in [92, 93] in a reaction to an earlier modal nonmonotonic logic of McDermott and Doyle [91]. Historically, autoepistemic logic played a major role in the development of nonmonotonic logics of belief. Moreover, intuitions underlying autoepistemic logic and studied in [47] motivated the concept of a stable model of a logic program [49][3] as discussed in detail in the next chapter of the Handbook.

6.3.1 Preliminaries, Intuitions and Basic Results

Autoepistemic logic was introduced to provide a formal account of a way in which an *ideally rational* agent forms *belief* sets given some initial assumptions. It is a formalism in a modal language. In our discussion we assume implicitly a fixed set At of propositional variables. We denote by $\mathcal{L}_K$ the modal language generated from At by means of boolean connectives and a (unary) modal operator K. The role of the modal operator K is to mark formulas as "believed". That is, intuitively, formulas KA stand for "A is believed". We refer to subsets of $\mathcal{L}_K$ as *modal theories*. We call formulas in $\mathcal{L}_K$ that do not contain occurrences of K *modal-free* or *propositional*. We denote the language consisting of all modal-free formulas by $\mathcal{L}$.

Let us consider a situation in which we have a rule that Professor Jones, being a university professor, normally teaches. To capture this rule in modal logic, we might say that if we do not believe that Dr. Jones does not teach (that is, if it is possible that she does), then Dr. Jones does teach. We might represent this rule by a modal formula.[4]

$$K\,prof_J \wedge \neg K \neg teaches_J \supset teaches_J. \tag{6.2}$$

Knowing only $prof_J$ (Dr. Jones is a professor) a rational agent should build a belief set containing $teaches_J$. The problem is to define the semantics of autoepistemic logic so that indeed it is so.

We see here a similarity with default logic, where the same rule is formalized by a default

$$prof(J) : teaches(J)/teaches(J) \tag{6.3}$$

(cf. Section 6.2.1). In default logic, given $W = \{prof(J)\}$, the conclusion $teaches(J)$ is supported as $(\{prof(J) : teaches(J)/teaches(J)\}, W)$ has exactly one extension and it does contain $teaches(J)$.

The correspondence between the formula (6.2) and the default (6.3) is intuitive and compelling. The key question is whether formally the autoepistemic logic interpretation of (6.2) is the same as the default logic interpretation of (6.3). We will return to this question later.

Before we proceed to present the semantics of autoepistemic logic, we will make a few comments on (classical) modal logics—formal systems of reasoning with modal formulas. This is a rich area and any overview that would do it justice is beyond the

[3] We note however, that default logic also played a role in the development of the stable-model semantics [13] and, in fact, the default-logic connection of stable models ultimately turned out to be more direct [82, 15, 14].

[4] To avoid problems with the treatment of quantifiers, we restrict our attention to the propositional case. Consequently, we have to list "normality" rules explicitly for each object in the domain rather than use schemata (formulas with variables) to represent concisely families of propositional rules, as it is possible in default logic. The "normality" rule in our example concerns Professor Jones only. If there were more professors in our domain, we would need rules of this type for each of them.

scope of this chapter. For a good introduction, we refer to [28, 57]. Here we only mention that many important modal logics are defined by a selection of modal axioms such K, T, D, 4, 5, etc. For instance, the axioms K, T, 4 and 5 yield the well-known modal logic S5. The consequence operator for a modal logic $\mathcal{S}$, say $\mathrm{Cn}_{\mathcal{S}}$, is defined syntactically in terms of the corresponding provability relation.[5]

For the reader familiar with *Kripke models* [28, 57], we note that the consequence operator $\mathrm{Cn}_{\mathcal{S}}$ can often be described in terms of a class of *Kripke models*, say $\mathcal{C}$: $A \in \mathrm{Cn}_{\mathcal{S}}(E)$ if and only if for every Kripke model $M \in \mathcal{C}$ such that $M \models_K E$, $M \models_K A$, where $\models_K$ stands for the relation of satisfiability of a formula or a set of formulas in a Kripke model. For instance, the consequence operator in the modal logic S5 is characterized by *universal* Kripke models. This characterization played a fundamental role in the development of autoepistemic logic. To make our chapter self-contained, rather than introducing Kripke models formally, we will use a different but equivalent characterization of the consequence operator in S5 in terms of *possible-world structures*, which we introduce formally later in the text.

After this brief digression we now come back to autoepistemic logic. What is an *ideally rational agent* or, more precisely, which modal theories could be taken as belief sets of such agents? Stalnaker [125] argued that to be a belief set of an ideally rational agent a modal theory $E \subseteq \mathcal{L}_K$ must satisfy three closure properties.

First, E must be closed under the propositional consequence operator. We will denote this operator by Cn.[6] Thus, the first property postulated by Stalnaker can be stated concisely as follows:

(B1) $\mathrm{Cn}(E) \subseteq E$.

We note that modal logics offer consequence operators which are stronger than the operator Cn. One might argue that closure under one of these operators might be a more appropriate for the condition (B1). We will return to this issue in a moment.

Next, Stalnaker postulated that theories modeling belief sets of ideally rational agents must be closed under *positive introspection*: if an agent believes in A, then the agent believes she believes A. Formally, we will require that a belief set E satisfies:

(B2) if $A \in E$, then $KA \in E$.

Finally, Stalnaker postulated that theories modeling belief sets of ideally rational agents must also be closed under *negative introspection*: if an agent does not believe A, then the agent believes she does not believe A. This property is formally captured by the condition:

(B3) if $A \notin E$, then $\neg KA \in E$.

Stalnaker's postulates have become commonly accepted as the defining properties of belief sets of an ideally rational agent. Thus, we refer to modal theories satisfying conditions (B1)–(B3) simply as *belief sets*. The original term used by Stalnaker was a *stable* theory. We choose a different notation since in nonmonotonic reasoning the term

[5] Proofs in a modal logic use as premises given assumptions (if any), instances of propositional tautologies in the language $\mathcal{L}_K$, and instances of modal axioms of the logic. As inference rules, they use modus ponens and the *necessitation rule*, which allows one to conclude KA once A has been derived.

[6] When applying the propositional consequence operator to *modal* theories, as we do here, we treat formulas KA as propositional variables.

stable is now most commonly associated with a class of models of logic programs, and there are fundamental differences between the two notions.

Belief sets have a rich theory [85]. We cite here only two results that we use later in the chapter. The first result shows that in the context of the conditions (B2) and (B3) the choice of the consequence operator for the condition (B1) becomes essentially immaterial. Namely, it implies that no matter what consequence relation we choose for (B1), as long as it contains the propositional consequence relation and is contained in the consequence relation for S5, we obtain the same notion of a belief set.

Proposition 6.13. *If $E \subseteq \mathcal{L}_K$ is a belief set, then E is closed under the consequence relation in the modal logic* S5.

The second result shows that belief sets are determined by their modal-free formulas. This property leads to a representation result for belief sets.

Proposition 6.14. *Let $T \subseteq \mathcal{L}$ be closed under propositional consequence. Then $E = \mathrm{Cn}_{S5}(T \cup \{\neg KA \mid A \in \mathcal{L} \setminus T\})$ is a belief set and $E \cap \mathcal{L} = T$. Moreover, if E is a belief set then $T = E \cap \mathcal{L}$ is closed under propositional consequence and $E = \mathrm{Cn}_{S5}(T \cup \{\neg KA \mid A \in \mathcal{L} \setminus T\})$.*

Modal nonmonotonic logics are meant to provide formal means to study mechanisms by which an agent forms belief sets starting with a set T of initial assumptions. These belief sets must contain T but may also satisfy some additional properties. A precise mapping assigning to a set of modal formulas a family of belief sets is what determines a modal nonmonotonic logic.

An obvious possibility is to associate with a set $T \subseteq \mathcal{L}_K$ *all* belief sets E such that $T \subseteq E$. This choice, however, results in a formalism which is *monotone*. Namely, if $T \subseteq T'$, then every belief set for T' is a belief set for T. Consequently, the set of "safe" beliefs—beliefs that belong to every belief set associated with T—grows monotonically as T gets larger. In fact, this set of safe beliefs based on T coincides with the set of consequences of T in the logic S5. As we aim to capture nonmonotonic reasoning, this choice is not of interest to us here.

Another possibility is to employ a minimization principle. Minimizing entire belief sets is of little interest as belief sets are incomparable with respect to inclusion and so, each of them is inclusion-minimal. Thus, this form of minimization does not eliminate any of the belief sets containing T, and so, it is equivalent to the approach discussed above.

A more interesting direction is to apply the minimization principle to modal-free fragments of belief sets (cf. Proposition 6.14, which implies that there is a one-to-one correspondence between belief sets and sets of modal-free formulas closed under propositional consequence). The resulting logic is in fact nonmonotonic and it received some attention [54].

The principle put forth by Moore when defining the autoepistemic logic can be viewed as yet another form of minimization. The conditions (B1)–(B3) imply that every belief set E containing T satisfies the inclusion

$$\mathrm{Cn}\big(T \cup \{KA \mid A \in E\} \cup \{\neg KA \mid A \notin E\}\big) \subseteq E.$$

Belief sets, for which the inclusion is proper, contain beliefs that do not follow from initial assumptions and from the results of "introspection" and so, are undesirable. Hence, Moore [93] proposed to associate with T only those belief sets E, which satisfy the *equality*:

$$\mathrm{Cn}\big(T \cup \{KA \mid A \in E\} \cup \{\neg KA \mid A \notin E\}\big) = E. \tag{6.4}$$

In fact, when a theory satisfies (6.4), we no longer need to assume that it is a belief set—(6.4) implies that it is.

Proposition 6.15. *For every* $T \subseteq \mathcal{L}_K$, *if* $E \subseteq \mathcal{L}_K$ *satisfies* (6.4), *then* E *satisfies* (B1)–(B3), *that is*, *it is a belief set*.

Moore called belief sets defined by (6.4) *stable expansions* of T. We will refer to them simply as *expansions* of T, dropping the term *stable* due to the same reason as before. We formalize our discussion in the following definition.

Definition 6.4. *Let* T *be a modal theory. A modal theory* E *is an expansion of* T *if* E *satisfies the identity* (6.4).

Belief sets have an elegant semantic characterization in terms of possible-world structures. Let $\mathcal{I}$ be the set of all 2-valued interpretations (truth assignments) of At. *Possible-world structures* are subsets of $\mathcal{I}$. Intuitively, a possible-world structure collects all interpretations that *might* be describing the actual world and leaves out those that definitely do not.

A possible-world structure is essentially a Kripke model with a total accessibility relation [28, 57]. The difference is that the universe of a Kripke model is required to be nonempty, which guarantees that the *theory* of the model (the set of all formulas true in the model) is consistent. Some modal theories consistent with respect to the propositional consequence relation determine inconsistent sets of beliefs. Allowing possible-world structures to be empty is a way to capture such situations and differentiate them from those situations, in which a modal theory determines no belief sets at all.

Possible-world structures interpret modal formulas, that is, assign to them truth values.

Definition 6.5. *Let* $Q \subseteq \mathcal{I}$ *be a possible-world structure and* $I \in \mathcal{I}$ *a two-valued interpretation. We define the truth function* $\mathcal{H}_{Q,I}$ *inductively as follows*:

1. $\mathcal{H}_{Q,I}(p) = I(p)$, *if* p *is an atom.*
2. $\mathcal{H}_{Q,I}(A_1 \wedge A_2) =$ **true** *if* $\mathcal{H}_{Q,I}(A_1) =$ **true** *and* $\mathcal{H}_{Q,I}(A_2) =$ **true**. *Otherwise*, $\mathcal{H}_{Q,I}(A_1 \wedge A_2) =$ **false**.
3. *Other boolean connectives are treated similarly.*
4. $\mathcal{H}_{Q,I}(KA) =$ **true**, *if for every interpretation* $J \in Q$, $\mathcal{H}_{Q,J}(A) =$ **true**. *Otherwise*, $\mathcal{H}_{Q,I}(KA) =$ **false**.

It follows directly from the definition that for every formula $A \in \mathcal{L}_K$, the truth value $\mathcal{H}_{Q,I}(KA)$ does not depend on I. It is fully determined by the possible-world structure Q and we will denote it by $\mathcal{H}_Q(KA)$, dropping I from the notation. Since Q determines the truth value of every modal atom, every modal formula A is either *believed* ($\mathcal{H}_Q(KA) = \mathbf{true}$) or *not believed* in Q ($\mathcal{H}_Q(KA) = \mathbf{false}$). In other words, the *epistemic* status of every modal formula is well defined in every possible-world structure.

The *theory* of a possible-world structure Q is the set of all modal formulas that are *believed* in Q. We denote it by $Th(Q)$. Thus, formally,

$$Th(Q) = \{A \mid \mathcal{H}_Q(KA) = \mathbf{true}\}.$$

We now present a characterization of belief sets in terms of possible-world structures, which we promised earlier.

Theorem 6.16. *A set of modal formulas $E \subseteq \mathcal{L}_K$ is a belief set if and only if there is a possible-world structure $Q \subseteq \mathcal{I}$ such that $E = Th(Q)$.*

Expansions of a modal theory can also be characterized in terms of possible-world structures. The underlying intuitions arise from considering a way to revise possible-world structures, given a set T of initial assumptions. The characterization is also due to Moore. Namely, for every modal theory T, Moore [92] defined an operator D_T on $\mathcal{P}(\mathcal{I})$ (the space of all possible-world structures) by setting

$$D_T(Q) = \{I \mid \mathcal{H}_{Q,I}(A) = \mathbf{true}, \text{ for every } A \in T\}.$$

The operator D_T specifies a process to revise belief sets encoded by the corresponding possible-world structures. Given a modal theory $T \subseteq \mathcal{L}_K$, the operator D_T revises a possible-world structure Q with a possible-world structure $D_T(Q)$. This revised structure consists of all interpretations that are *acceptable* given the current structure Q and the constraints on belief sets encoded by T. Specifically, the revision consists precisely of those interpretations that make all formulas in T true with respect to Q.

Fixed points of the operator D_T are of particular interest. They represent "stable" possible-world structures (and so, belief sets)—they cannot be revised any further. This property is behind the role they play in the autoepistemic logic.

Theorem 6.17. *Let $T \subseteq \mathcal{L}_K$. A set of modal formulas $E \subseteq \mathcal{L}_K$ is an expansion of T if and only if there is a possible-world structure $Q \subseteq \mathcal{I}$ such that $Q = D_T(Q)$ and $E = Th(Q)$.*

This theorem implies a systematic procedure for constructing expansions of *finite* modal theories (or, to be more precise, possible-world structures that determine expansions). Let us continue our "Professor Jones" example and let us look at a theory

$$T = \{prof_J, Kprof_J \wedge \neg K\neg teaches_J \supset teaches_J\}.$$

There are two propositional variables in our language and, consequently, four propositional interpretations:

$I_1 = \emptyset$ (neither $prof_J$ nor $teaches_J$ is true),

$I_2 = \{prof_J\}$,

$I_3 = \{teaches_J\}$,

$I_4 = \{prof_J, teaches_J\}$.

There are 16 possible-world structures one can build of these four interpretations. Only one of them, though, $Q = \{prof_J, teaches_J\}$, satisfies $D_T(Q) = Q$ and so, generates an expansion of T. We skip the details of verifying it, as the process is long and tedious, and we present a more efficient method in the next section. We note however, that for the basic "Professor Jones" example autoepistemic logic gives the same conclusions as default logic.

We close this section by noting that autoepistemic logic can also be obtained as a special case of a general fixed point schema to define modal nonmonotonic logics proposed by McDermott [90]. In this schema, we assume that an agent uses some modal logic $\mathcal{S}$ (extending propositional logic) to capture her basic means of inference. We then say that a modal theory $E \subseteq \mathcal{L}_K$ is an *$\mathcal{S}$-expansion* of a modal theory T if

$$E = \mathrm{Cn}_{\mathcal{S}}\big(T \cup \{\neg KA \mid A \notin E\}\big). \tag{6.5}$$

In this equation, $\mathrm{Cn}_{\mathcal{S}}$ represents the consequence relation in the modal logic $\mathcal{S}$. If E satisfies (6.5), then E is closed under the propositional consequence relation. Moreover, E is closed under the necessitation rule and so, E is closed under positive introspection. Finally, since $\{\neg KA \mid A \notin E\} \subseteq E$, E is closed under negative introspection. It follows that solutions to (6.5) are belief sets containing T. They can be taken as models of belief sets of agents reasoning by means of modal logic $\mathcal{S}$ and justifying what they believe on the basis of initial assumptions in T and *assumptions* about what *not* to believe (negative introspection). By choosing different monotone logics $\mathcal{S}$, we obtain from this schema different classes of $\mathcal{S}$-expansions of T.

If we disregard inconsistent expansions, autoepistemic logic can be viewed as a special instance of this schema, with $\mathcal{S} = \mathrm{KD45}$, the modal logic determined by the axioms K, D, 4 and 5 [57, 85]. Namely, we have the following result.

Theorem 6.18. *Let $T \subseteq \mathcal{L}_K$. If $E \subseteq \mathcal{L}_K$ is consistent, then E is an expansion of T if and only if E is a* KD45*-expansion of T, that is,*

$$E = \mathrm{Cn}_{\mathrm{KD45}}\big(T \cup \{\neg KA \mid A \notin E\}\big).$$

6.3.2 Computational Properties

The key reasoning problems for autoepistemic logic are deciding *skeptical inference* (whether a formula is in all expansions), *credulous inference* (whether a formula is in some expansion), and finding expansions. Like default logic, first order autoepistemic logic is not semi-decidable even when quantifying into the scope of the modal operator is not allowed [94]. If quantifying-in is allowed, the reasoning problems are highly undecidable [63].

In order to clarify the computational properties of propositional autoepistemic logic we present a finitary characterization of expansions based on *full sets* [94, 95]. A full set is constructed from the KA and $\neg KA$ subformulas of the premises and it serves as the characterizing kernel of an expansion. An overview of other approaches to characterizing expansions can be found in [95].

The characterization is based on the set of all subformulas of the form KA in a set of premises T. We denote this set by $\mathrm{Sf}_K(T)$. We stress that in the characterization only the classical consequence relation (Cn) is used where KA formulas are treated as propositional variables and no modal consequence relation is needed. To simplify the notation, for a set T of formulas we will write $\neg T$ as a shorthand for $\{\neg F \mid F \in T\}$.

Definition 6.6 (Full sets). *For a set of formulas T, a set $S \subseteq \mathrm{Sf}_K(T) \cup \neg\,\mathrm{Sf}_K(T)$ is T-full if and only if the following two conditions hold for every $KA \in \mathrm{Sf}_K(T)$:*

- *$A \in \mathrm{Cn}(T \cup S)$ if and only if $KA \in S$.*
- *$A \notin \mathrm{Cn}(T \cup S)$ if and only if $\neg KA \in S$.*

In fact, for a T-full set S, the classical consequences of $T \cup S$ provide the modal-free part of an expansion. As explained in Proposition 6.14 this uniquely determines the expansion. Here we give an alternative way of constructing an expansion from a full set presented in [95] which is more suitable for automation. For this we employ a restricted notion of subformulas: $\mathrm{Sf}^p_K(F)$ is the set of *primary* subformulas of F, i.e., all subformulas of the form KA of F which are not in the scope of another K operator in F. For example, if p and q are atomic, $\mathrm{Sf}^p_K(K(\neg Kp \rightarrow q) \wedge K\neg q) = \{K(\neg Kp \rightarrow q), K\neg q\}$. The construction uses a simple consequence relation $\models_K$ which is given recursively on top of the classical consequence relation Cn. It turns out that this consequence relation corresponds exactly to membership in an expansion when given its characterizing full set.

Definition 6.7 (K-consequence). *Given a set of formulas T and a formula F,*

$$T \models_K F \quad \textit{if and only if} \quad F \in \mathrm{Cn}(T \cup \mathrm{SB}_T(F))$$

where $\mathrm{SB}_T(F) = \{KA \in \mathrm{Sf}^p_K(F) \mid T \models_K A\} \cup \{\neg KA \in \neg\,\mathrm{Sf}^p_K(F) \mid T \not\models_K A\}$.

For an expansion E of T, there is a corresponding T-full set

$$\{KF \in E \mid KF \in \mathrm{Sf}_K(T)\} \cup \{\neg KF \in E \mid KF \in \mathrm{Sf}_K(T)\}$$

and for a T-full set S,

$$\{F \mid T \cup S \models_K F\}$$

is an expansion of T. In fact it can be shown [95] that there is a one-to-one correspondence between full sets and expansions.

Theorem 6.19 (Expansions in terms of full sets). *Let T be a set of autoepistemic formulas. Then a function SE_T defined as*

$$SE_T(S) = \{F \mid T \cup S \models_K F\}$$

gives a bijective mapping from the set of T-full sets to the set of expansions of T and for a T-full set S, $SE_T(S)$ is the unique expansion E of T such that $S \subseteq \{KF \mid F \in E\} \cup \{\neg KF \mid F \notin E\}$.

Example 6.2. Consider our "Professor Jones" example and a set of formulas

$$T = \{prof_J, Kprof_J \wedge \neg K\neg teaches_J \supset teaches_J\}.$$

Now $\mathrm{Sf}_K(T) = \{Kprof_J, K\neg teaches_J\}$ and there are four possible full sets:

$$\{\neg Kprof_J, \neg K\neg teaches_J\}, \quad \{Kprof_J, \neg K\neg teaches_J\},$$
$$\{\neg Kprof_J, K\neg teaches_J\}, \quad \{Kprof_J, K\neg teaches_J\}.$$

It is easy to verify that only $S_1 = \{Kprof_J, \neg K\neg teaches_J\}$ satisfies the conditions in Definition 6.6, that is, $prof \in \mathrm{Cn}(T \cup S_1)$ and $\neg teaches_J \notin \mathrm{Cn}(T \cup S_1)$. Hence, T has exactly one expansion $SE_T(S_1)$ which contains, for instance, $KKprof_J$ and $\neg K\neg K teaches_J$ as $T \cup S_1 \models_K KKprof_J$ and $T \cup S_1 \models_K \neg K\neg K teaches_J$ hold.

Example 6.3. Consider a set of formulas

$$T' = \{Kp \supset p\}.$$

Now $\mathrm{Sf}_K(T') = \{Kp\}$ and there are two possible full sets: $\{\neg Kp\}$ and $\{Kp\}$ which are both full. For instance, $p \in \mathrm{Cn}(T' \cup \{Kp\})$. Hence, T' has exactly two expansions $SE_{T'}(\{\neg Kp\})$ and $SE_{T'}(\{Kp\})$.

The finitary characterization of expansions in Theorem 6.19 implies that propositional autoepistemic reasoning is *decidable* and can be implemented in *polynomial space*. This is because the conditions on a full set and on membership of an arbitrary autoepistemic formula in an expansion induced by a full set are based on the classical propositional consequence relation which is decidable in polynomial space.

Similar to default logic, deciding whether an expansion exists and credulous inference are Σ_2^{P}-complete problems and sceptical inference is Π_2^{P}-complete for autoepistemic logic as well as for many other modal nonmonotonic logics [51, 94, 95, 121]. This implies that modal nonmonotonic reasoning is strictly harder than classical reasoning (unless the polynomial hierarchy collapses) and achieving tractability requires substantial restrictions on how modal operators can interact [83, 84]. For more information on automating autoepistemic reasoning, see for instance [97, 36].

6.4 Circumscription

6.4.1 Motivation

Circumscription was introduced by John McCarthy [86, 87]. Many of its formal aspects were worked out by Vladimir Lifschitz who also wrote an excellent overview [74]. We follow here the notation and terminology used in this overview article.

The idea underlying circumscription can be explained using the teaching professors example discussed in the introduction. There we considered using the following first order formula to express *professors normally teach*:

$$\forall x\big(prof(x) \wedge \neg abnormal(x) \supset teaches(x)\big).$$

The problem with this formula is the following: in order to apply it to Professor Jones, we need to prove that Jones is not abnormal. In many cases we simply do not have enough information to do this. Intuitively, we do not expect objects to be abnormal—unless we have explicit information that tells us they indeed are abnormal. Let us assume there is no reason to believe Jones is abnormal. We implicitly assume—in McCarthy's words: jump to the conclusion—$\neg$*abnormal*(*Jones*) and use it to conclude *teaches*(*Jones*).

What we would like to have is a mechanism which models this form of jumping to conclusions. Note that what is at work here is a minimization of the extent of the predicate *abnormal*: we want as few objects as possible—given the available information—to satisfy this predicate. How can this be achieved?

The answer provided by circumscription has a syntactical and a corresponding semantical side. From the syntactical point of view, circumscription is a transformation (more precisely, a family of transformations) of logical formulas. Given a sentence A representing the given information, circumscription produces a logically stronger sentence A^*. The formulas which follow from A using circumscription are simply the formulas classically entailed by A^*. In our example, A contains the given information about professors, their teaching duties, and Jones. In addition to this information, A^* also expresses that the extent of *abnormal* is minimal. Note that in order to express minimality of a predicate one has to quantify over predicates. For this reason A^* will be a second order formula.

Semantically, circumscription gives up the classical point of view that all models of a sentence A have to be regarded as equal possibilities. In our example, different models of A may have different extents for the predicate *abnormal* (the set of objects belonging to the interpretation of *abnormal*) even if the domain of the models is the same. It is natural to consider models with fewer abnormal objects—in the sense of set inclusion—as more plausible than those with more abnormal objects. This induces a preference relation on the set of all models. The idea now is to restrict the definition of entailment to the most preferred models only: a formula f is preferentially entailed by A if and only if f is true in all maximally preferred models of A.

We will see that this elegant model theoretic construction captures exactly the syntactic transformation described above.

6.4.2 Defining Circumscription

For the definition of circumscription some abbreviations are useful. Let P and Q be two predicate symbols of the same arity n:

$$\begin{aligned} P = Q &\quad \text{abbreviates} \quad \forall x_1 \cdots x_n((P(x_1,\ldots,x_n) \equiv Q(x_1,\ldots,x_n)),\\ P \leqslant Q &\quad \text{abbreviates} \quad \forall x_1 \cdots x_n((P(x_1,\ldots,x_n) \supset Q(x_1,\ldots,x_n)),\\ P < Q &\quad \text{abbreviates} \quad (P \leqslant Q) \wedge \neg(P = Q). \end{aligned}$$

The formulas express: P and Q have the same extent, the extent of P is a subset of the extent of Q, and the extent of P is a proper subset of the extent of Q, respectively.

Definition 6.8. *Let $A(P)$ be a sentence containing a predicate symbol P. Let p be a predicate variable of the same arity as P. The circumscription of P in $A(P)$, which will be denoted by CIRC$[A(P); P]$, is the second order sentence*

$$A(P) \wedge \neg\exists p\big[A(p) \wedge p < P\big].$$

Table 6.1. Examples of circumscribing P

$A(P)$	$CIRC[A(P); P]$
$P(a)$	$\forall x(P(x) \equiv x = a)$
$P(a) \wedge P(b)$	$\forall x(P(x) \equiv (x = a \vee x = b))$
$P(a) \vee P(b)$	$\forall x(P(x) \equiv (x = a)) \vee \forall x(P(x) \equiv (x = b))$
$\neg P(a)$	$\forall x \neg P(x)$
$\forall x(Q(x) \supset P(x))$	$\forall x(Q(x) \equiv P(x))$

By $A(p)$ we denote here the result of uniformly substituting predicate constant P in $A(P)$ by variable p. Intuitively, the second order formula $\neg\exists p[A(p) \wedge p < P]$ says: it is not possible to find a predicate p such that

1. p satisfies what is said in $A(P)$ about P, and
2. the extent of p is a proper subset of the extent of P.

In other words: the extent of P is minimal subject to the condition $A(P)$.

Table 6.1 presents some simple formulas $A(P)$ together with the result of circumscribing P in $A(P)$. The examples are taken from [74].

Although it gives desired results in simple cases, this form of circumscription is not yet powerful enough for most applications. It allows us to minimize the extent of a predicate, but only if this does not change the interpretation of any other symbol in the language. In the Professor Jones example, for instance, minimizing the predicate *abnormal* alone is not sufficient to conclude *teaches*(*Jones*). To obtain this conclusion, we have to make sure that the extent of *teaches* is allowed to change during the minimization of *abnormal*. This can be achieved with the following more general definition:

Definition 6.9. *Let $A(P, Z_1, \ldots, Z_m)$ be a sentence containing the predicate constant P and predicate/function constants Z_i. Let $p, z_1, \ldots, z_m$ be predicate/function variables of the same type and arity as $P, Z_1, \ldots, Z_m$. The circumscription of P in $A(P, Z_1, \ldots, Z_m)$ with varied $Z_1, \ldots, Z_m$, denoted CIRC$[A(P, Z_1, \ldots, Z_m); P; Z_1, \ldots, Z_m]$, is the second order sentence*

$$A(P, Z_1, \ldots, Z_m) \wedge \neg\exists p z_1 \ldots z_m \big[A(p, z_1, \ldots, z_m) \wedge p < P\big].$$

A further generalization where several predicates can be minimized in parallel is also very useful. Whenever we want to represent several default rules, we need different abnormality predicates ab_1, ab_2 etc., since being abnormal with respect to one default is not necessarily related to being abnormal with respect to another default.

We first need to generalize the abbreviations $P = Q$, $P \leqslant Q$ and $P < Q$ to the case where P and Q are sequences of predicate symbols. Let $P = P_1, \ldots, P_n$ and $Q = Q_1, \ldots, Q_n$, respectively:

$$P = Q \quad \text{abbreviates} \quad P_1 = Q_1 \wedge \cdots \wedge P_n = Q_n,$$
$$P \leqslant Q \quad \text{abbreviates} \quad P_1 \leqslant Q_1 \wedge \cdots \wedge P_n \leqslant Q_n,$$
$$P < Q \quad \text{abbreviates} \quad P \leqslant Q \wedge \neg(P = Q).$$

Here is the generalized definition:

Definition 6.10. *Let $P = P_1, \dots, P_k$ be a sequence of predicate constants, $Z = Z_1, \dots, Z_m$ a sequence of predicate/function constants. Furthermore, let $A(P, Z)$ be a sentence containing the predicate constants P_i and predicate/function constants Z_j. Let $p = p_1, \dots, p_k$ and $z = z_1, \dots, z_m$ be predicate/function variables of the same type and arity as $P_1, \dots, P_k$, respectively, $Z_1, \dots, Z_m$. The (parallel) circumscription of P in $A(P, Z)$ with varied Z, denoted CIRC$[A(P, Z); P; Z]$, is the second order sentence*

$$A(P, Z) \wedge \neg\exists pz\big[A(p, z) \wedge p < P\big].$$

Predicate and function constants which are neither minimized nor varied, i.e., neither in P nor in Z, are called fixed.

6.4.3 Semantics

Circumscription allows us to minimize the extent of predicates. This can be elegantly described in terms of a preference relation on the models of the circumscribed sentence A. Intuitively, we prefer a model M_1 over a model M_2 whenever the extent of the minimized predicate(s) P is smaller in M_1 than in M_2. Of course, M_1 can only be preferred over M_2 if the two models are comparable: they must have the same universe, and they have to agree on the fixed constants.

In the following, for a structure M we use $|M|$ to denote the universe of M and $M[\![C]\!]$ to denote the interpretation of the (individual/function/predicate) constant C in M.

Definition 6.11. *Let M_1 and M_2 be structures, P a sequence of predicate constants, Z a sequence of predicate/function constants. M_1 is at least as $P; Z$-preferred as M_2, denoted $M_1 \leqslant^{P;Z} M_2$, whenever the following conditions hold:*

1. $|M_1| = |M_2|$,
2. $M_1[\![C]\!] = M_2[\![C]\!]$ *for every constant C which is neither in P nor in Z,*
3. $M_1[\![P_i]\!] \subseteq M_2[\![P_i]\!]$ *for every predicate constant P_i in P.*

The relation $\leqslant^{P;Z}$ is obviously transitive and reflexive. We say a structure M is $\leqslant^{P;Z}$-minimal within a set of structures $\mathcal{M}$ whenever there is no structure $M' \in \mathcal{M}$ such that $M' <^{P;Z} M$. Here $<^{P;Z}$ is the strict order induced by $\leqslant^{P;Z}$: $M' <^{P;Z} M$ if and only if $M' \leqslant^{P;Z} M$ and not $M \leqslant^{P;Z} M'$.

The following proposition shows that the $P; Z$-minimal models of A capture exactly the circumscription of P in A with varied Z:

Proposition 6.20. *M is a model of CIRC$[A; P; Z]$ if and only if M is $\leqslant^{P;Z}$-minimal among the models of A.*

It should be pointed out that circumscription may lead to inconsistency, even if the circumscribed sentence A is consistent. This happens whenever we can find a better model for each model, implying that there is an infinite chain of more and more preferred models. A discussion of conditions under which consistency of circumscription is guaranteed can be found in [74]. For instance, it is known that $\mathit{CIRC}[A; P; Z]$ is consistent whenever A is universal (of the form $\forall x A$ where x is a tuple of object variables and A is quantifier-free) and Z does not contain function symbols.

6.4.4 Computational Properties

In circumscription the key computational problem is that of sceptical inference, i.e., determining whether a formula is true in all minimal models. However, general first order circumscription is highly uncomputable [120]. This is not surprising as circumscription transforms a first order sentence into a second order formula and it is well known that second order logic is not even semi-decidable. This means that in order to compute circumscription we cannot just use our favorite second order prover—such a prover simply cannot exist. We can only hope to find computational methods for certain special cases of first order formulas.

We first discuss techniques for computing circumscriptive inference in the first order case and then present a finitary characterization of minimal models which illustrates computational properties of circumscription.

Methods for computing circumscription can be roughly categorized as follows:

- *guess and verify*: the idea is to guess right instances of second order variables to prove conjectures about circumscription. Of course, this is a method requiring adequate user interaction, not a full mechanization,
- *translation to first order logic*: this method is based on results depending on syntactic restrictions and transformation rules,
- *specialized proof procedures*: these can be modified first order proof procedures or procedures for restricted second order theories.

As an illustration of the guess and verify method consider the Jones example again. Abbreviating *abnormal* with *ab* we have

$$A(\mathit{ab}, \mathit{teaches}) = \mathit{prof}(J) \wedge \forall x(\mathit{prof}(x) \wedge \neg \mathit{ab}(x) \supset \mathit{teaches}(x)).$$

We are interested in $\mathit{CIRC}[A(\mathit{ab}, \mathit{teaches}); \mathit{ab}; \mathit{teaches}]$ which is

$$A(\mathit{ab}, \mathit{teaches}) \wedge \neg \exists p z \big[A(p, z) \wedge p < \mathit{ab}\big].$$

By simple equivalence transformations and by spelling out the abbreviation $p < \mathit{ab}$ we obtain

$$A(\mathit{ab}, \mathit{teaches}) \wedge \forall p z \big[A(p, z) \wedge \forall x(p(x) \supset \mathit{ab}(x)) \supset \forall x(\mathit{ab}(x) \supset p(x))\big].$$

If we substitute the right predicate expressions for the now universally quantified predicate variables p and z, we can indeed prove $\mathit{teaches}(J)$. By a predicate expression we mean an expression of the form $\lambda x_1, \ldots, x_n.F$ where F is a first order formula. Applying this predicate expression to n terms $t_1, \ldots, t_n$ yields the formula obtained by substituting all variables x_i in F uniformly by t_i.

In our example we guess that no object is *ab*, that is we substitute for p the expression $\lambda x.\mathit{false}$. Similarly, we guess that professors are the teaching objects, i.e., we substitute for z the expression $\lambda x.\mathit{prof}(x)$. The resulting first order formula (after simple equivalence transformations) is

$$A(ab, \mathit{teaches}) \wedge$$
$$\big[\mathit{prof}(J) \wedge \forall x(\mathit{prof}(x) \supset \mathit{prof}(x)) \wedge$$
$$\forall x(\mathit{false} \supset ab(x)) \supset \forall x(ab(x) \supset \mathit{false})\big].$$

It is easy to verify that the first order formula obtained with these substitutions indeed implies *teaches*(J). In cases where derivations are more difficult one can, of course, use a standard first order theorem prover to verify conjectures after substituting predicate expressions.

For the second method, the translation of circumscription to first order logic, a number of helpful results are known. We cannot go into much detail here and refer the reader to [74] for an excellent overview. As an example of the kind of results used we present two useful propositions.

Let $A(P)$ be a formula and P a predicate symbol occurring in A. A formula A, without any occurrence of $\supset$ and $\equiv$, is *positive/negative* in P if all occurrences of P in $A(P)$ are positive/negative. (We recall that the occurrence of a predicate symbol P in a formula $A(P)$ without occurrences of $\supset$ and $\equiv$ is positive if the number of its occurrences in the range of the negation operator is positive. Otherwise, it is negative.)

Proposition 6.21. *Let $B(P)$ be a formula without any occurrences of $\supset$ and $\equiv$. If $B(P)$ is negative in P, then $CIRC[A(P) \wedge B(P);\, P]$ is equivalent to $CIRC[A(P);\, P] \wedge B(P)$.*

Proposition 6.22. *Let $A(P, Z)$ be a formula without any occurrences of $\supset$ and $\equiv$. If $A(P, Z)$ is positive in P, then $CIRC[A(P, Z);\, P;\, Z]$ is equivalent to*

$$A(P, Z) \wedge \neg\exists xz\big[P(x) \wedge A(\lambda y.(P(y) \wedge x \neq y), z)\big].$$

Here x and y stand for n-tuples of distinct object variables, where n is the arity of predicate symbol P. As a corollary of these propositions we obtain that $CIRC[A(P) \wedge B(P);\, P]$ is equivalent to a first order formula whenever $A(P)$ is positive and $B(P)$ negative in P (assuming $A(P)$ and $B(P)$ do not contain $\supset$ and $\equiv$).

Apart from translations to first order logic, translations to logic programming have also been investigated [48].

Several specialized theorem proving methods and systems have been developed for restricted classes of formulas. Among these we want to mention Przymusinski's MILO-resolution [109], Baker and Ginsberg's argument based circumscriptive prover [7], the tableaux based method developed by Niemelä [99], and two algorithms based on second order quantifier elimination: the SCAN algorithm [45, 102] and the DLS algorithm [37].

We now turn to the question how minimal models, the key notion in circumscription, can be characterized in order to shed light on computational properties of circumscription and its relationship to classical logic. We present a characterization of minimal models where the minimality of a model can be determined independently of other models using a test for classical consequence. We consider here parallel predicate circumscription in the clausal case and with respect to Herbrand interpretations and a characterization proposed in [99]. A similar characterization but for the propositional case has been used in [41] in the study of the computational complexity of propositional circumscription.

Definition 6.12 (Grounded models). *Let T be a set of clauses and let P and R be sets of predicates. A Herbrand interpretation M is said to be* grounded *in $\langle T, P, R\rangle$ if and only if for all ground atoms $p(\vec{t})$ such that $p \in P$, $M \models p(\vec{t})$ implies $p(\vec{t}) \in \mathrm{Cn}(T \cup \mathrm{N}^{\langle P,R\rangle}(M))$ where*

$$\mathrm{N}^{\langle P,R\rangle}(M) = \{\neg q(\vec{t}) \mid q(\vec{t}) \textit{ is a ground atom}, q \in P \cup R, M \not\models q(\vec{t})\} \cup \{q(\vec{t}) \mid q(\vec{t}) \textit{ is a ground atom}, q \in R, M \models q(\vec{t})\}.$$

Theorem 6.23 (Minimal models). *Let T be a set of clauses and let P and Z be the sets of minimized and varied predicates, respectively. A Herbrand interpretation M is a $\leqslant^{P;Z}$-minimal model of T if and only if M is a model of T and grounded in $\langle T, P, R\rangle$ where R is the set of predicates in T that are in neither P nor Z.*

Example 6.4. Let $T = \{p(x) \vee \neg q(x)\}$ and let the underlying language have only one ground term a. Then the Herbrand base is $\{p(a), q(a)\}$. Consider the sets of minimized predicates $P = \{p\}$ and varied predicates $Z = \emptyset$. Then the set of fixed predicates $R = \{q\}$. Now the Herbrand interpretation $M = \{p(a), q(a)\}$, which is a model of T, is grounded in $\langle T, P, R\rangle$ because $\mathrm{N}^{\langle P,R\rangle}(M) = \{q(a)\}$ and $p(a) \in \mathrm{Cn}(T \cup \mathrm{N}^{\langle P,R\rangle}(M))$ holds. Hence, M is a minimal model of T. If $Z = \{q\}$, then $R = \emptyset$ and M is not grounded in $\langle T, P, R\rangle$ because $\mathrm{N}^{\langle P,R\rangle}(M) = \emptyset$ and $p(a) \notin \mathrm{Cn}(T \cup \mathrm{N}^{\langle P,R\rangle}(M))$. Thus, if p is minimized but q is varied, M is not a minimal model of T.

Theorem 6.23 implies that circumscriptive inference is decidable in polynomial space in the propositional case. Like for default logic, it is strictly harder than classical propositional reasoning unless the polynomial hierarchy collapses as it is Π_2^{P}-complete [40, 41]. For tractability considerable restrictions are needed [27].

6.4.5 Variants

Several variants of circumscription formalizing different kinds of minimization have been developed. For instance, pointwise circumscription [71] allows us to minimize the value of a predicate for each argument tuple separately, rather than minimizing the extension of the predicate. This makes it possible to specify very flexible minimization policies. Autocircumscription [105] combines minimization with introspection.

We will focus here on prioritized circumscription [70]. In many applications some defaults are more important than others. In inheritance hierarchies, for instance, a default representing more specific information is intuitively expected to "win" over a conflicting default: if birds normally fly, penguins normally do not, then one would expect to conclude that a penguin does not fly, although it is a bird. This can be modeled by minimizing some abnormality predicates with higher priority than others.

Prioritized circumscription splits the sequence P of minimized predicates into disjoint segments $P^1, \dots, P^k$. Predicates in P^1 are minimized with highest priority, followed by those in P^2, etc. Semantically, this amounts to a lexicographic comparison of models. We first compare two models M_1 and M_2 with respect to $\leqslant^{P^1,Z}$, where Z are the varied symbols. If the models are incomparable, or if one of the models is strictly preferred ($<^{P^1,Z}$ holds), then the relationship between the models is established and we are done. If $M_1 =^{P^1,Z} M_2$, we go on with $\leqslant^{P^2,Z}$, etc.

The prioritized circumscription of $P^1, \dots, P^k$ in A with varied Z is denoted

$$CIRC[A;\, P^1 > \cdots > P^k;\, Z].$$

We omit its original definition and rather present a characterization based on a result in [70] which shows that prioritized circumscription can be reduced to a sequence of parallel circumscriptions:

Proposition 6.24. *$CIRC[A;\, P^1 > \cdots > P^k;\, Z]$ is equivalent to the conjunction of circumscriptions*

$$\bigwedge_{i=1}^{k} CIRC[A;\, P^i;\, P^{i+1}, \dots, P^k, Z].$$

6.5 Nonmonotonic Inference Relations

Having discussed three specific nonmonotonic formalisms in considerable detail, we will now move on to an orthogonal theme in nonmonotonic reasoning research: an abstract study of inference relations associated with nonmonotonic (defeasible) reasoning. Circumscription fits in this theme quite well—it can be viewed as an example of a preferential model approach, yielding a preferential inference relation. However, as we mention again at the end of this chapter, it is not so for default and autoepistemic logics. In fact, casting these two and other fixed point logics in terms of the semantic approach to nonmonotonic inference we are about to present is one of major problems of nonmonotonic reasoning research.

Given what we know about the world, when could a formula B reasonably be concluded from a formula A? One "safe" answer is provided by the classical concept of entailment. Let T be a set of first order logic sentences (an agent's knowledge about the world). The agent *classically* infers a formula B if B holds in *every* model of T in which A holds.

However, the agent's knowledge of the world is typically incomplete, and so, inference relations based on formalisms of defeasible reasoning are of significant interest, too. Under circumscription, the agent might infer B from A if B holds in every *minimal* model of T, in which A holds, $A \mathrel{|\!\sim}_{T,circ} B$. In default logic, assuming the knowledge of the world is given in terms of a set D of defaults, the agent might infer B from A, $A \mathrel{|\!\sim}_D B$, if B is in *every* extension of the default theory $(D, \{A\})$.

These examples suggest that inference can be modeled as a binary relation on $\mathcal{L}$. The question we deal with in this section is: which binary relations on $\mathcal{L}$ are inference relations and what are their properties?

In what follows, we restrict ourselves to the case when $\mathcal{L}$ consists of formulas of propositional logic. We use the infix notation for binary relations and write $A \mathrel{|\!\sim} B$ to denote that B follows from A, under a concept of inference modeled by a binary relation $\mathrel{|\!\sim}$ on $\mathcal{L}$.

6.5.1 Semantic Specification of Inference Relations

Every propositional theory T determines a set of its *models*, $Mod(T)$, consisting of propositional interpretations satisfying T. These interpretations can be regarded as complete specifications of worlds consistent with T or, in other words, possible given T.

An agent whose knowledge is described by T might reside in any of these worlds. Such an agent may decide to infer $B \in \mathcal{L}$ from $A \in \mathcal{L}$, written $A \vdash_T B$, if in *every* world in which A holds, B holds, as well. This approach sanctions only the most conservative inferences. They will hold no matter what additional information about the world an agent may acquire. Inference relations of the form $\vdash_T$ are important. They underlie classical propositional logic and are directly related to the logical entailment relation $\models$. Indeed, we have that $A \vdash_T B$ if and only if $T, A \models B$.

The class of inference relations of the form $\vdash_T$ has a characterization in terms of abstract properties of binary relations on $\mathcal{L}$. The list gives some examples of properties of binary relations relevant for the notion of inference.

Monotony	if $A \supset B$ is a tautology and $B \mathrel{|\!\sim} C$, then $A \mathrel{|\!\sim} C$,
Right Weakening	if $A \supset B$ is a tautology and $C \mathrel{|\!\sim} A$, then $C \mathrel{|\!\sim} B$,
Reflexivity	$A \mathrel{|\!\sim} A$,
And	if $A \mathrel{|\!\sim} B$ and $A \mathrel{|\!\sim} C$, then $A \mathrel{|\!\sim} B \wedge C$,
Or	if $A \mathrel{|\!\sim} C$ and $B \mathrel{|\!\sim} C$, then $A \vee B \mathrel{|\!\sim} C$.

It turns out that these properties provide an alternative (albeit non-constructive) specification of the class of relations of the form $\vdash_T$. Namely, we have the following theorem [64].

Theorem 6.25. *A binary relation on $\mathcal{L}$ is of the form $\vdash_T$ if and only if it satisfies the five properties given above.*

Due to the property of *Monotony*, inference relations $\vdash_T$ do not give rise to defeasible arguments. To model defeasible arguments we need less conservative inference relations. To this end, one may relax the requirement that B must hold in *every* world in which A holds. In commonsense reasoning, humans often differentiate between possible worlds, regarding some of them as more typical or normal than others. When making inferences they often consider only those worlds that are most typical given the knowledge they have. Thus, they might infer B from A if B holds in every most typical world in which A holds (and not in each such world).

Preferential models [64] provide a framework for this general approach. The key idea is to use a *strict* partial order,[7] called a *preference relation*, to compare worlds

[7] A binary relation that is irreflexive and transitive.

with respect to their "typicality", with more typical worlds preferred to less typical ones. Given a strict partial order $\prec$ on a set W, an element $w \in W$ is $\prec$-minimal if there is no element $w' \in W$ such that $w' \prec w$.

In the following definition, we use again the term a *possible-world structure*. This time, however, we use it to denote a slightly broader class of objects than sets of interpretations.

Definition 6.13. *A general possible-world structure is a tuple* $\langle W, v \rangle$, *where* W *is a set of worlds and* v *is a function mapping worlds to interpretations.*[8] *If* A *is a formula, we define*

$$W(A) = \{w \in W \mid v(w) \models A\}.$$

A preferential model is a tuple $\mathcal{W} = \langle W, v, \prec \rangle$, *where* $\langle W, v \rangle$ *is a general possible-world structure and* $\prec$ *is a strict partial order on* W *satisfying the following smoothness condition*: *for every sentence* A *and for every* $w \in W(A)$, w *is* $\prec$*-minimal in* $W(A)$ *or there is* $w' \in W(A)$ *such that* $w' \prec w$ *and* w' *is a* $\prec$*-minimal element of* $W(A)$.

The set $W(A)$ gathers worlds in which A holds. Minimal elements in $W(A)$ can be viewed as most typical states where A holds. The smoothness condition guarantees that for every world $w \in W(A)$ which is not most typical itself, there is a most typical state in $W(A)$ that is preferred to w.

Preferential models formalize the intuition of reasoning on the basis of most preferred (typical) models only and allow us to specify the corresponding concept of inference.

Definition 6.14. *If* $\mathcal{W}$ *is a preferential model* (*with the ordering* $\prec$), *then the inference relation determined by* $\mathcal{W}$, $\mid\!\sim_{\mathcal{W}}$, *is defined as follows*: *for* $A, B \in \mathcal{L}$, $A \mid\!\sim_{\mathcal{W}} B$ *if* B *holds in every* $\prec$*-minimal world in* $W(A)$.

We call inference relations of the form $\mid\!\sim_{\mathcal{W}}$, where $\mathcal{W}$ is a preferential model, *preferential*. In general, they do not satisfy the property of *Monotony*.

Propositional circumscription is an example of this general method of defining inference relations. Let $\mathcal{I}$ stand for the set of all interpretations of $\mathcal{L}$. Furthermore, let P and Z be two disjoint sets of propositional variables in the language. We note that the relation $<^{P;Z}$ satisfies the smoothness condition. Thus, $\langle \mathcal{I}, v, <^{P;Z} \rangle$, where v is the identity function, is a preferential model. Moreover, it defines the same inference relation as does circumscription.

Shoham's preference logic [123] is another specialization of the preferential model approach. As in circumscription, the set of worlds consists of all interpretations of $\mathcal{L}$ but an arbitrary strict partial order satisfying the smoothness condition[9] can be used.

Preference logics are very close to preferential models. However, allowing multiple worlds with the same interpretation (in other words, using general possible-world

[8] Typically, W is assumed to be nonempty. This assumption is not necessary for our considerations here and so we do not adopt it.

[9] In the original paper by Shoham, a stronger condition of well-foundedness was used.

structures rather than possible-world structures) is essential. The resulting class of inference relations is larger (we refer to [25] for an example).

Can preferential relations be characterized by means of meta properties? The answer is yes but we need two more properties of binary relations $\mid\!\sim$ on $\mathcal{L}$:

Left Logical Equivalence if A and B are logically equivalent and $A \mid\!\sim C$, then $B \mid\!\sim C$

Cautious Monotony if $A \mid\!\sim B$ and $A \mid\!\sim C$, then $A \wedge B \mid\!\sim C$

We have the following theorem [64].

Theorem 6.26. *A binary relation* $\mid\!\sim$ *on* $\mathcal{L}$ *is a preferential inference relation if and only if it satisfies Left Logical Equivalence*, *Right Weakening*, *Reflexivity*, *And*, *Or and Cautious Monotony.*

We note that many other properties of binary relations were considered in an effort to formalize the concept of nonmonotonic inference. Gabbay [44] asked about the weakest set of conditions a binary relation should satisfy in order to be a nonmonotonic inference relation. The result of his studies as well as of Makinson [79] was the notion of a cumulative inference relation. A semantic characterization of cumulative relations exists but there are disputes whether cumulative relations are indeed the right ones. Thus, we do not discuss cumulative inference relations here.

Narrowing the class of orders in preferential models yields subclasses of preferential relations. One of these subclasses is especially important for nonmonotonic reasoning. A strict partial order $\prec$ on a set P is *ranked* if there is a function l from P to ordinals such that for every $x, y \in P$, $x \prec y$ if and only if $l(x) < l(y)$.

Definition 6.15. *A preferential model* $\langle \mathcal{W}, v, \prec \rangle$ *is ranked if* $\prec$ *is ranked.*

We will call inference relations defined by ranked models *rational*. It is easy to verify that rational inference relations satisfy the property of *Rational Monotony*:

Rational Monotony if $A \wedge B \not\mid\!\sim C$ and $A \not\mid\!\sim \neg B$, then $A \not\mid\!\sim C$.

The converse is true, as well. We have the following theorem [68].

Theorem 6.27. *An inference relation is rational if and only if it is preferential and satisfies Rational Monotony.*

6.5.2 Default Conditionals

Default conditionals are meant to model defeasible statements such as *university professors normally give lectures*. Formally, a *default conditional* is a syntactic expression $A \mid\!\sim B$, with an intuitive reading "if A then *normally* B". We denote the operator constructing default conditionals with the same symbol $\mid\!\sim$ we used earlier for inference relations. While it might be confusing, there are good reasons to do so and they will become apparent as we proceed. It is important, however, to keep in mind that in one case, $\mid\!\sim$ stands for a constructor of syntactic (language) expressions, and in the other it stands for a binary (inference) relation.

Given a set K of default conditionals, when is a default conditional $A \mathrel{|\!\sim} B$ a consequence of K? When is a formula A a consequence of K? Somewhat disappointingly no single commonly accepted answer has emerged. We will now review one of the approaches proposed that received significant attention. It is based on the notion of a *rational closure* developed in [67, 68] and closely related to the system Z [104].

Let K be a set of default conditionals. The set of all default conditionals implied by K should be closed under some rules of inference for conditionals. For instance, we might require that if A and B are logically equivalent and $A \mathrel{|\!\sim} C$ belongs to a closure of K, $B \mathrel{|\!\sim} C$ belongs to the closure of K, as well. This rule is nothing else but *Left Logical Equivalence*, except that now we view expressions $A \mathrel{|\!\sim} B$ as default conditionals and not as elements of an inference relation. In fact, modulo this correspondence (a conditional $A \mathrel{|\!\sim} B$ versus an element $A \mathrel{|\!\sim} B$ of an binary relation), several other rules we discussed in the previous section could be argued as possible candidates to use when defining a closure of K.

Based on this observation, we postulate that a closure of K should be a set of conditionals that corresponds to an inference relation. The question is, which inference relation extending K should one adopt as *the* closure of K. If one is given a preferential model whose inference relation extends K, this inference relation might be considered as the closure of K. This is not a satisfactory solution as, typically, all we have is K and we would like to determine the closure on the basis of K only. Another answer might be the intersection of all preferential relations extending K. The resulting relation does not in general satisfy *Rational monotony*, a property that arguably all *bona fide* nonmonotonic inference relations should satisfy. Ranked models determine inference relations that are preferential and, moreover, satisfy *Rational Monotony*. However, the intersection of all rational extensions of K coincides with the intersection of all preferential extensions and so, this approach collapses to the previous one.

If the closure of K is not the intersection of all rational extensions, perhaps it is a specific rational extension, if there is a natural way to define one. We will focus on this possibility now. Lehmann and Magidor [68] introduce a partial ordering on rational extensions of a set of conditional closures of K. In the case when this order has a least element, they call this element the *rational closure* of K. They say that $A \mathrel{|\!\sim} B$ is a rational consequence of K if $A \mathrel{|\!\sim} B$ belongs to the rational closure of K. They say that A is a rational consequence of K if the conditional $\mathbf{true} \mathrel{|\!\sim} A$ is in the rational closure of K.

There are sets of conditionals that do not have the rational closure. However, [68] show that in many cases, including the case when K is finite, the rational closure exists. Rather than discuss the ordering of rational extensions that underlies the definition of a rational closure, we will now discuss an approach which characterizes it in many cases when it exists.

A formula A is exceptional for K, if $\mathbf{true} \mathrel{|\!\sim} \neg A$ belongs to the preferential extension of K, that is, if $\neg A$ is true in every minimal world of every preferential model of K. A default conditional is exceptional for K, if its antecedent is exceptional for K. By $E(K)$ we denote the set of all default conditionals in K that are exceptional for K.

Given K, we define a sequence of subsets of K as follows: $C_0 = K$. If $\tau = \eta + 1$ is a successor ordinal, we define $C_\tau = E(C_\eta)$. If τ is a limit ordinal, we define $C_\tau = \bigcup_{\eta<\tau} C_\eta$.

The rank $r(A)$ of a formula A is the least ordinal τ such that A is not exceptional for C_τ. If for every ordinal τ, A is exceptional for C_τ, A has no rank.

A formula A is *inconsistent* with K if for every preferential model of K and every world w in the model, $w \models \neg A$.

A set of conditionals K is *admissible* if all formulas that have no rank are inconsistent for K. Admissible sets of default conditionals include all finite sets.

Theorem 6.28. *If K is admissible, then its rational closure $\bar{K}$ exists. A default conditional $A \mathrel{|\!\sim} B \in \bar{K}$ if and only if $A \wedge \neg B$ has no rank, or if A and $A \wedge \neg B$ have ranks and $r(A) < r(A \wedge \neg B)$.*

6.5.3 Discussion

Properties of inference relations can reveal differences between nonmonotonic formalisms. Earlier in this section, we showed how circumscription or default logic can be used to specify inference relations. The relation determined by circumscription is a special case of a preferential inference relation and so, satisfies all properties of preferential relations. The situation is different for the inference relation defined by a set of defaults. Let us recall that B can be inferred from A with respect to a set D of defaults, $A \mathrel{|\!\sim}_D B$, if B is in every extension of the default theory $(D, \{A\})$.

The inference relation $\mathrel{|\!\sim}_D$, where D consists of normal defaults, in general does not satisfy the properties *Or* and *Cautious Monotony*. For instance, let $D = \{A : C/C, B : C/C\}$. Then we have $A \mathrel{|\!\sim}_D C$ and $B \mathrel{|\!\sim}_D C$, but not $A \vee B \vdash_D C$. The reason, intuitively, is that none of the defaults can be applied if only the disjunction of prerequisites is given.

An example for the violation of cumulativity due to Makinson [79] is given by $D = \{\top : A/A, A \vee B : \neg A/\neg A\}$. We have $\top \mathrel{|\!\sim}_D A$ and thus $\top \mathrel{|\!\sim}_D A \vee B$, but not $A \vee B \vdash_D A$. The reason is that the default theory $(D, \{A \vee B\})$ has a second extension containing $\neg A$.

Contrary to normal defaults, supernormal defaults satisfy both *Cautious Monotony* and *Or* [35], as they happen to be preferential.

Finally, we conclude this section with a major unresolved problem of nonmonotonic reasoning. Nonmonotonicity can be achieved through fixed point constructions and this approach leads to such formalisms as default and autoepistemic logics. On the other hand, interesting nonmonotonic inference relations can be defined in terms of preferential models. What is missing is a clear link between the two approaches. An open question is: can nonmonotonic inference relations defined by default logic (or other fixed point system) be characterized in semantic terms along the lines of preferential models?

6.6 Further Issues and Conclusion

In this section we discuss the relationship between the major approaches we presented earlier. We first relate default logic and autoepistemic logic (Section 6.6.1), then default logic and circumscription (Section 6.6.2). Finally, we give pointers to some other approaches which we could not present in more detail in this chapter (Section 6.6.3).

6.6.1 Relating Default and Autoepistemic Logics

A basic pattern of nonmonotonic reasoning is: "*in the absence of any information contradicting B, infer B*". Normal defaults are designed specifically with this reasoning pattern in mind: it is modeled by the normal default $\frac{:B}{B}$. McDermott and Doyle [91] suggested that in modal nonmonotonic systems this reasoning pattern should be represented by the modal formula $\neg K\neg B \supset B$ (or using a common abbreviation M for $\neg K\neg$, which can be read as "consistent" or "possible": $MB \supset B$). Even though the modal nonmonotonic logic of [91] was found to have counterintuitive properties and was abandoned as a knowledge representation formalism, the connection between a default $\frac{:B}{B}$ and a modal formula $MB \supset B$ was an intriguing one and prompted extensive investigations. Since autoepistemic logic emerged in the mid 1980s as the modal nonmonotonic logic of choice, these investigations focused on relating default and autoepistemic logics.

Building on the suggestion of McDermott and Doyle, Konolige [61] proposed to encode an arbitrary default

$$d = \frac{A : B_1, \ldots, B_k}{C}$$

with a modal formula

$$T(d) = KA \wedge \neg K\neg B_1 \wedge \cdots \wedge \neg K\neg B_k \supset C,$$

and to translate a default theory $\Delta = (D, W)$ into a modal theory $T(\Delta) = W \cup \{T(d) \mid d \in D\}$.

The translation seems to capture correctly the intuitive reading of a default: if A is known and all B_i are possible (none is contradicted or inconsistent) then infer C. There is a problem, though. Let us consider a default theory $\Delta = (\{d\}, \emptyset)$, where

$$d = \frac{A : B}{A}.$$

Konolige's translation represents Δ as a modal theory

$$T(\Delta) = \{KA \wedge \neg K\neg B \supset A\}.$$

Using methods we presented earlier in this chapter one can verify that Δ has exactly one extension, $\mathrm{Cn}(\emptyset)$, while $T(\Delta)$ has *two* expansions, $\mathrm{Cn}_{S5}(\emptyset)$ and $\mathrm{Cn}_{S5}(\{A\})$. It follows that Konolige's translation does not yield a connection between the two logics that would establish a one-to-one correspondence between extensions and expansions. Still several interesting properties hold.

First, as shown in [81], for prerequisite-free default theories, Konolige's translation does work! We have the following result.

Theorem 6.29. *Let Δ be a default theory such that each of its defaults is prerequisite-free. Then, a propositional theory E is an extension of Δ if and only if the belief set determined by E (cf. Proposition 6.14) is an expansion of $T(\Delta)$. Conversely, a modal theory E' is an expansion of $T(\Delta)$ if and only if the modal-free part of E', $E' \cap \mathcal{L}$, is an extension of Δ.*

Second, under Konolige's translation, extensions are mapped to expansions (although, as our example above shows—the converse fails in general).

Theorem 6.30. *Let Δ be a default theory. If a propositional theory E is an extension of Δ, then* $\mathrm{Cn}_{\mathrm{S5}}(E)$ *is an expansion of* $T(\Delta)$.

Despite providing evidence that the two logics are related, ultimately, Konolige's translation does not properly match extensions with expansions. The reason boils down to a fundamental difference between extensions and expansions. Both extensions and expansions consist only of formulas that are justified ("grounded") in default and modal theories, respectively. However, expansions allow for self-justifications while extensions do not. The difference is well illustrated by the example we used before. The belief set determined by $\{A\}$ (cf. Proposition 6.14) is an expansion of the theory $\{KA \wedge \neg K\neg B \supset A\}$. In this expansion, A is justified through the formula $KA \wedge \neg K\neg B \supset A$ by means of a *circular* argument relying on believing in A (since there is no information contradicting B, the second premise needed for the argument, $\neg K\neg B$, holds). Such self-justifications are not sanctioned by extensions: in order to apply the default $\frac{A:B}{A}$ we must first *independently* derive A. Indeed, one can verify that the theory $\mathrm{Cn}(\{A\})$ is not an extension of $(\{\frac{A:B}{A}\}, \emptyset)$.

This discussion implies that extensions and expansions capture different types of nonmonotonic reasoning. As some research suggests default logic is about the modality of "knowing" (no self-supporting arguments) and autoepistemic logic is about the modality of "believing" (self-supporting arguments allowed) [75, 122].

Two natural questions arise. Is there a default logic counterpart of expansions, and is there an autoepistemic logic counterpart of extensions? The answer in each case is positive. Denecker et al. [34] developed a uniform treatment of default and autoepistemic logics exploiting some basic operators on possible-world structures that can be associated with default and modal theories. This algebraic approach (developed earlier in more abstract terms in [33]) endows each logic with both expansions and extensions in such a way that they are perfectly aligned under Konolige's translation. Moreover, extensions of default theories and expansions of modal theories defined by the algebraic approach of [34] coincide with the original notions defined by Reiter and Moore, respectively, while expansions of default theories and extensions of modal theories defined in [34] fill in the gaps to complete the picture.

A full discussion of the relation between default and autoepistemic logic is beyond the scope of this chapter and we refer to [34] for details. Similarly, we only briefly note other work attempting to explain the relationship between the two logics. Most efforts took as the starting point the observation that to capture a default logic within a modal system, a different modal nonmonotonic logic or a different translation must be used. Konolige related default logic to a *version* of autoepistemic logic based on the notion of a *strongly grounded expansion* [61]. Marek and Truszczyński [82] proposed an alternative translation and represented extensions as expansions in a certain modal nonmonotonic logic constructed following McDermott [90]. Truszczyński [128] found that the Gödel translation of intuitionistic logic to modal logic S4 could be used to translate the default logic into a nonmonotonic modal logic S4 (in fact, he showed that several modal nonmonotonic logics could be used in place of nonmonotonic S4).

Gottlob [52] returned to the original problem of relating default and autoepistemic logics with their original semantics. He described a mapping translating default theories into modal ones so that extensions correspond precisely to expansions. This translation is not modular. The autoepistemic representation of a default theory depends on the whole theory and cannot be obtained as the union of independent translations of individual defaults. Thus, the approach of Gottlob does not provide an autoepistemic reading of an individual default. In fact, in the same paper Gottlob proved that a modular translation from default logic with the semantics of extensions to autoepistemic logic with the semantics of expansions does not exist. In conclusion, there is *no* modal interpretation of a default under which *extensions* would correspond to *expansions*.

6.6.2 Relating Default Logic and Circumscription

The relationships between default logic and circumscription as well as between autoepistemic logic and circumscription have been investigated by a number of researchers [42, 43, 58, 72, 62]. Imielinski [58] points out that even normal default rules with prerequisites cannot be translated modularly into circumscription. This argument applies also to autoepistemic logic and thus circumscription cannot modularly capture autoepistemic reasoning [96].

On the other hand, circumscription is closely related to prerequisite-free normal defaults. For example, it is possible to capture minimal models of a set of formulas using such rules. The idea is easy to explain in the propositional case. Consider a set of formulas T and sets P and Z of minimized and varied atoms (0-ary predicates), respectively, and let R be the set of fixed atoms (those not in P or Z). Now $\leqslant^{P;Z}$-minimal models of T can be captured by the default theory $(\mathrm{MIN}(P) \cup \mathrm{FIX}(R), T)$ where the set of defaults consists of

$$\mathrm{MIN}(P) = \left\{ \frac{\top : \neg A}{\neg A} \;\middle|\; A \in P \right\},$$

$$\mathrm{FIX}(R) = \left\{ \frac{\top : \neg A}{\neg A} \;\middle|\; A \in R \right\} \cup \left\{ \frac{\top : A}{A} \;\middle|\; A \in R \right\}.$$

Now a formula F is true in every $\leqslant^{P;Z}$-minimal model of T if and only if F is in every extension of the default theory $(\mathrm{MIN}(P) \cup \mathrm{FIX}(R), T)$. The idea here is that defaults $\mathrm{MIN}(P)$ minimize atoms in P and defaults $\mathrm{FIX}(R)$ fix atoms in R by minimizing each atom and its complement.

The same approach can be used for autoepistemic logic as prerequisite-free default theories can be translated to autoepistemic logic as explained in Section 6.6.1. However, capturing first order circumscription is non-trivial and the results depend on the treatment of open defaults (or quantification into the scope of K operators in the case of autoepistemic logic). For example, Etherington [42] reports results on capturing circumscription using default logic in the first order case but without any fixed predicates and with a finite, fixed domain. Konolige [62] shows how to encode circumscription in the case of non-finite domains using a variant of autoepistemic logic which allows quantification into the scope of K operators.

6.6.3 Further Approaches

Several other formalizations of nonmonotonic reasoning have been proposed in the literature. Here we give a few references to those we consider most relevant but could not handle in more detail.

- Possibilistic logics [38] assign degrees of necessity and possibility to sentences. These degrees express the extent to which these sentences are believed to be necessarily or possibly true, respectively. One of the main advantages of this approach is that it leads to a notion of graded inconsistency which allows non-trivial deductions to be performed from inconsistent possibilistic knowledge bases. The resulting consequence relation is nonmonotonic and default rules can be conveniently represented in this approach [10].

- Defeasible logic, as proposed by Nute [101] and further developed by Antoniou and colleagues [4, 3], is an approach to nonmonotonic reasoning based on strict and defeasible rules as well as defeaters. The latter specify exceptions to defeasible rules. A preference relation among defeasible rules is used to break ties whenever possible. An advantage of defeasible logic is its low complexity: inferences can be computed very efficiently. On the other hand, some arguably intuitive conclusions are not captured. The relationship between defeasible logic and prioritized logic programs under well-founded semantics is discussed in [24].

- Inheritance networks are directed graphs whose nodes represent propositions and a directed (possibly negated) link between two nodes A and B stands for "As *are normally (not)* Bs" (some types of networks also distinguish between strict and defeasible links). The main goal of approaches in this area is to capture the idea that more specific information should win in case of a conflict. Several notions of specificity have been formalized, and corresponding notions of inference were developed. Reasoning based on inheritance networks is nonmonotonic since new, possibly more specific links can lead to the retraction of former conclusions. [56] gives a good overview.

- Several authors have proposed approaches based on ranked knowledge bases, that is, sets of classical formulas together with a total preorder on the formulas [21, 9]. The preorder represents preferences reflecting the willingness to stick to a formula in case of conflict: if two formulas A and B lead to inconsistency, then the strictly less preferred formula is given up. If they are equally preferred, then different preferred maximal consistent subsets (preferred subtheories in the terminology of [21]) of the formulas will be generated. There are different ways to define the preferred subtheories. Brewka [21] uses a criterion based on set inclusion, Benferhat and colleagues [9] investigate a cardinality based approach.

- When considering knowledge-based agents it is natural to assume that the agent's beliefs are exactly those beliefs which follow from the assumption that the knowledge base is *all* that is believed. Levesque was the first to capture this notion in his logic of *only-knowing* [69]. The main advantage of this approach is that beliefs can be analyzed in terms of a modal logic without requiring additional meta-logical notions like fixpoints and the like. The logic uses two modal operators, K for belief and O for only knowing. Levesque showed that his logic captures autoepistemic logic. In [65] the approach was generalized to capture default logic as well. [66] presents a sound and complete axiomatization for the propositional case. Multi-agent only knowing is explored in [53].

- Formal argument systems (see, for instance, [76, 124, 106, 39, 20, 129, 1, 130, 12]) model the way agents reason on the basis of arguments. In some approaches arguments have internal structure, in others they remain abstract entities whose structure is not analyzed further. In each case a defeat relation among arguments plays a central role in determining acceptable arguments and acceptable beliefs. The approaches are too numerous to be discussed here in more detail. We refer the reader to the excellent overview articles [29] and [108].

With the above references to further work we conclude this overview chapter on formalizations of general nonmonotonic reasoning. As we said in the introduction, our aim was not to give a comprehensive overview of all the work that has been done in the area. We decided to focus on the most influential approaches, thus providing the necessary background for several of the other chapters of this Handbook. Indeed, the reader will notice that the topic of this chapter pops up again at various places in this book—with a different, more specialized focus. Examples are the chapters on Answer Sets (Chapter 7), Model-based Problem Solving (Chapter 10), and the various approaches to reasoning about action and causality (Chapters 16–19).

Acknowledgements

We would like to thank Vladimir Lifschitz and Hudson Turner for helpful comments and suggestions.

Bibliography

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Ann. Math. Artif. Intell.*, 34(1–3):197–215, 2002.

[2] G. Antoniou. *Nonmonotonic Reasoning*. 1st edition. MIT Press, Cambridge, 1997.

[3] G. Antoniou, D. Billington, G. Governatori, and M.J. Maher. Representation results for defeasible logic. *ACM Trans. Comput. Log.*, 2(2):255–287, 2001.

[4] G. Antoniou, D. Billington, G. Governatori, M.J. Maher, and A. Rock. A family of defeasible reasoning logics and its implementation. In W. Horn, editor. *ECAI*, pages 459–463. IOS Press, 2000.

[5] F. Baader and B. Hollunder. Embedding defaults into terminological knowledge representation formalisms. *J. Automat. Reason.*, 14:149–180, 1995.

[6] F. Baader and B. Hollunder. Priorities on defaults with prerequisites, and their application in treating specificity in terminological default logic. *J. Automat. Reason.*, 15(1):41–68, 1995.

[7] A.B. Baker and M.L. Ginsberg. A theorem prover for prioritized circumscription. In N.S. Sridharan, editor. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 463–467. Morgan Kaufmann, San Mateo, CA, 1989.

[8] R. Ben-Eliyahu and R. Dechter. Default logic, propositional logic and constraints. In *Proceedings of the 9th National Conference on Artificial Intelligence*, pages 370–385. MIT Press, July 1991.

[9] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax-based entailment. In *IJCAI*, pages 640–647, 1993.

[10] S. Benferhat, D. Dubois, and H. Prade. Representing default rules in possibilistic logic. In B. Nebel, C. Rich, and W. Swartout, editors. *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 673–684. Morgan Kaufmann, San Mateo, CA, 1992.

[11] P. Besnard. *An Introduction to Default Logic*. Springer-Verlag, New York, 1989.

[12] P. Besnard and A. Hunter. Practical first-order argumentation. In *Proceedings Twentieth National Conference on Artificial Intelligence, AAAI*, pages 590–595, 2005.

[13] N. Bidoit and C. Froidevaux. Minimalism subsumes default logic and circumscription. In *Proceedings of IEEE Symposium on Logic in Computer Science, LICS-87*, pages 89–97. IEEE Press, 1987.

[14] N. Bidoit and C. Froidevaux. General logical databases and programs: default logic semantics and stratification. *Inform. and Comput.*, 91(1):15–54, 1991.

[15] N. Bidoit and C. Froidevaux. Negation by default and unstratifiable logic programs. *Theoret. Comput. Sci. (Part B)*, 78(1):85–112, 1991.

[16] A. Bochman. *A Logical Theory of Nonmonotonic Inference and Belief Change*. 1st edition. Springer, Berlin, 2001.

[17] A. Bochman. *Explanatory Nonmonotonic Reasoning*. 1st edition. *Advances in Logic*, vol. 4. World Scientific, 2005.

[18] P.A. Bonatti. Sequent calculi for default and autoepistemic logics. In *Proceedings of the Fifth Workshop on Theorem Proving with Analytic Tableaux and Related Methods*, Terrasini, Italy, May 1996. *Lecture Notes in Artificial Intelligence*, vol. 1071, pages 127–142. Springer-Verlag, 1996.

[19] P.A. Bonatti and N. Olivetti. A sequent calculus for skeptical default logic. In *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, Pont-à-Mousson, France, May 1997. *Lecture Notes in Artificial Intelligence*, vol. 1227, pages 107–121. Springer-Verlag, 1997.

[20] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[21] G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *IJCAI* pages 1043–1048, 1989.

[22] G. Brewka. Cumulative default logic: In defense of nonmonotonic inference rules. *Artificial Intelligence*, 50(2):183–205, 1991.

[23] G. Brewka. Reasoning about priorities in default logic. In *AAAI*, pages 940–945, 1994.

[24] G. Brewka. On the relationship between defeasible logic and well-founded semantics. *Lecture Notes in Computer Science*, 2173:121–132, 2001.

[25] G. Brewka, J. Dix, and K. Konolige. *Nonmonotonic Reasoning: An Overview*. 1st edition. CSLI Publications, Stanford, 1997.

[26] G. Brewka and T. Eiter. Preferred answer sets for extended logic programs. *Artificial Intelligence*, 109(1–2):297–356, 1999.

[27] M. Cadoli and M. Lenzerini. The complexity of propositional closed world reasoning and circumscription. *J. Comput. System Sci.*, 48(2):255–310, 1994.

[28] B.F. Chellas. *Modal Logic. An Introduction*. Cambridge University Press, Cambridge–New York, 1980.

[29] C.I. Chesñevar, A.G. Maguitman, and R.P. Loui. Logical models of argument. *ACM Comput. Surv.*, 32(4):337–383, 2000.

[30] P. Cholewiński, V.W. Marek, A. Mikitiuk, and M. Truszczyński. Computing with default logic. *Artificial Intelligence*, 112:105–146, 1999.

[31] J.P. Delgrande and W.K. Jackson. Default logic revisited. In *KR*, pages 118–127, 1991.

[32] J.P. Delgrande, T. Schaub, and W.K. Jackson. Alternative approaches to default logic. *Artificial Intelligence*, 70(1–2):167–237, 1994.

[33] M. Denecker, V.W. Marek, and M. Truszczyński. Approximations, stable operators, well-founded fixpoints and applications in nonmonotonic reasoning. In J. Minker, editor. *Logic-Based Artificial Intelligence*, pages 127–144. Kluwer Academic Publishers, 2000.

[34] M. Denecker, V.W. Marek, and M. Truszczyński. Uniform semantic treatment of default and autoepistemic logics. *Artificial Intelligence*, 143:79–122, 2003.

[35] J. Dix. Default theories of Poole-type and a method for constructing cumulative versions of default logic. In *ECAI*, pages 289–293, 1992.

[36] J. Dix, U. Furbach, and I. Niemelä. Nonmonotonic reasoning: Towards efficient calculi and implementations. In A. Voronkov and A. Robinson, editors. *Handbook of Automated Reasoning*, vol. II, pages 1241–1354. Elsevier Science, Amsterdam, 2001 (Chapter 19).

[37] P. Doherty, W. Lukaszewicz, and A. Szalas. Computing circumscription revisited: A reduction algorithm. *J. Automat. Reason.*, 18(3):297–336, 1997.

[38] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors. *Nonmonotonic Reasoning and Uncertain Reasoning*, *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, pages 439–513. Oxford University Press, Oxford, 1994.

[39] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[40] T. Eiter and G. Gottlob. Propositional circumscription and extended closed world reasoning are Π_2^P-complete. *Theoret. Comput. Sci.*, 144(2):231–245, 1993; Addendum: *Theoret. Comput. Sci.*, 118:315, 1993.

[41] T. Eiter and G. Gottlob. On the computational cost of disjunctive logic programming: Propositional case. *Ann. Math. Artificial Intelligence*, 15(3–4):289–323, 1995.

[42] D.W. Etherington. Relating default logic and circumscription. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, Milan, Italy, August 1987, pages 489–494. Morgan Kaufmann Publishers, 1987.

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

[44] D.M. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In *Proceedings of the NATO Advanced Study Institute on Logics and Models of Concurrent Systems*, pages 439–457. Springer, 1989.
[45] D.M. Gabbay and H.J. Ohlbach. Quantifier elimination in second-order predicate logic. In B. Nebel, C. Rich, and W. Swartout, editors. *Principles of Knowledge Representation and Reasoning (KR92)*, pages 425–435. Morgan Kaufmann, 1992.
[46] M.R. Garey and D.S. Johnson. *Computers and Intractability*. W.H. Freeman and Company, San Francisco, 1979.
[47] M. Gelfond. On stratified autoepistemic theories. In *Proceedings of AAAI-87*, pages 207–211. Morgan Kaufmann, 1987.
[48] M. Gelfond and V. Lifschitz. Compiling circumscriptive theories into logic programs. In *Proceedings of the 7th National Conference on Artificial Intelligence, AAAI-88*, pages 455–449, 1988.
[49] M. Gelfond and V. Lifschitz. The stable semantics for logic programs. In *Proceedings of the 5th International Conference on Logic Programming*, pages 1070–1080. MIT Press, 1988.
[50] G. Gogic, C. Papadimitriou, B. Selman, and H. Kautz. The comparative linguistics of knowledge representation. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Canada, August 1995, pages 862–869. Morgan Kaufmann Publishers, 1995.
[51] G. Gottlob. Complexity results for nonmonotonic logics. *J. Logic Comput.*, 2(3):397–425, 1992.
[52] G. Gottlob. Translating default logic into standard autoepistemic logic. *J. ACM*, 42(4):711–740, 1995.
[53] J.Y. Halpern and G. Lakemeyer. Multi-agent only knowing. *J. Logic Comput.*, 11(1):41–70, 2001.
[54] J.Y. Halpern and Y. Moses. Towards a theory of knowledge and ignorance (preliminary report). In K. Apt, editor. *Logics and Models of Concurrent Systems (La Colle-sur-Loup, 1984)*, *NATO ASI Series F: Computer and Systems Sciences*, vol. 13, pages 459–476. Springer, 1985.
[55] C. Hewitt. Description and theoretical analysis (using schemata) of planner: A language for proving theorems and manipulating models in a robot. PhD thesis, MIT, 1971.
[56] J.F. Horty. Some direct theories of nonmonotonic inheritance. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors. *Nonmonotonic Reasoning and Uncertain Reasoning*, *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, pages 111–187. Oxford University Press, Oxford, 1994.
[57] G.E. Hughes and M.J. Cresswell. *A Companion to Modal Logic*. Methuen and Co., Ltd., London, 1984.
[58] T. Imielinski. Results on translating defaults to circumscription. *Artificial Intelligence*, 32:131–146, 1987.
[59] U. Junker and K. Konolige. Computing the extensions of autoepistemic and default logics with a truth maintenance system. In *Proceedings of the 8th National Conference on Artificial Intelligence*, Boston, MA, USA, July 1990, pages 278–283. MIT Press, 1990.
[60] H. Kautz and B. Selman. Hard problems for simple default logics. *Artificial Intelligence*, 49:243–279, 1991.

[61] K. Konolige. On the relation between default and autoepistemic logic. *Artificial Intelligence*, 35(3):343–382, 1988.

[62] K. Konolige. On the relation between autoepistemic logic and circumscription. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Detroit, MI, USA, August 1989, pages 1213–1218. Morgan Kaufmann Publishers, 1989.

[63] K. Konolige. Quantification in autoepistemic logic. Technical Note 510, SRI International, Menlo Park, CA, USA, September 1991.

[64] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[65] G. Lakemeyer and H.J. Levesque. Only-knowing: Taking it beyond autoepistemic reasoning. In *Proceedings Twentieth National Conference on Artificial Intelligence, AAAI*, pages 633–638, 2005.

[66] G. Lakemeyer and H.J. Levesque. Towards an axiom system for default logic. In *Proc. AAAI-06*. AAAI Press, 2006.

[67] D. Lehmann. What does a conditional knowledge base entail? In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning, KR-89*, pages 212–222. Morgan Kaufmann, 1989.

[68] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[69] H.J. Levesque. All I know: A study in autoepistemic logic. *Artificial Intelligence*, 42(2–3):263–309, 1990.

[70] V. Lifschitz. Computing circumscription. In *IJCAI*, pages 121–127, 1985.

[71] V. Lifschitz. Pointwise circumscription: Preliminary report. In *AAAI*, pages 406–410, 1986.

[72] V. Lifschitz. Between circumscription and autoepistemic logic. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 235–244. Morgan Kaufmann Publishers, Toronto, Canada, May 1989.

[73] V. Lifschitz. On open defaults. In J. Lloyd, editor. *Proceedings of the Symposium on Computational Logic*, pages 80–95. Springer-Verlag, Berlin, 1990.

[74] V. Lifschitz. Circumscription. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors. *Nonmonotonic Reasoning and Uncertain Reasoning*, *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, pages 298–352. Oxford University Press, 1994.

[75] V. Lifschitz. Minimal belief and negation as failure. *Artificial Intelligence*, 70(1–2):53–72, 1994.

[76] F. Lin and Y. Shoham. Argument systems: A uniform basis for nonmonotonic reasoning. In *KR*, pages 245–255, 1989.

[77] W. Lukaszewicz. Chronological minimization of abnormality: Simple theories of action. In *ECAI*, pages 574–576, 1988.

[78] W. Lukaszewicz. *Non-Monotonic Reasoning: Formalization of Commonsense Reasoning, Ellis Horwood Series in Artificial Intelligence*. Chichester, England, 1990.

[79] D. Makinson. General theory of cumulative inference. In *Proceedings of the 2nd International Workshop on Non-Monotonic Reasoning, NMR-88*, *Lecture Notes in Computer Science*, vol. 346, pages 1–18. Springer, 1989.

[80] D. Makinson. *Bridges from Classical to Nonmonotonic Logic. King's College Publications: Texts in Computing*, London, 2005.

[81] V.W. Marek and M. Truszczyński. Relating autoepistemic and default logics. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, ON, 1989, pages 276–288. Morgan Kaufmann, San Mateo, CA, 1989.

[82] V.W. Marek and M. Truszczyński. Stable semantics for logic programs and default theories. In E. Lusk and R. Overbeek, editors. *Proceedings of the North American Conference on Logic Programming*, pages 243–256. MIT Press, 1989.

[83] V.W. Marek and M. Truszczyński. Autoepistemic logic. *J. ACM*, 38:588–619, 1991.

[84] V.W. Marek and M. Truszczyński. Computing intersection of autoepistemic expansions. In *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 37–50, MIT Press, Washington, DC, USA, July 1991.

[85] V.W. Marek and M. Truszczyński. *Nonmonotonic Logics: Context-Dependent Reasoning*. 1st edition. Springer, Berlin, 1993.

[86] J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1–2):27–39, 1980.

[87] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28(1):89–116, 1986.

[88] J. McCarthy. *Formalization of Common Sense*. Papers by John McCarthy edited by V. Lifschitz. Ablex, 1990.

[89] J. McCarthy and P.J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors. *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969 (reprinted in [88]).

[90] D. McDermott. Nonmonotonic logic II: Nonmonotonic modal theories. *J. ACM*, 29(1):33–57, 1982.

[91] D. McDermott and J. Doyle. Nonmonotonic logic I. *Artificial Intelligence*, 13(1–2):41–72, 1980.

[92] R.C. Moore. Possible-world semantics for autoepistemic logic. In *Proceedings of the Workshop on Non-Monotonic Reasoning*, pages 344–354, 1984. Reprinted in: M. Ginsberg, editor, *Readings on Nonmonotonic Reasoning*, pages 137–142. Morgan Kaufmann, 1990.

[93] R.C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1):75–94, 1985.

[94] I. Niemelä. Towards automatic autoepistemic reasoning. In *Proceedings of the European Workshop on Logics in Artificial Intelligence—JELIA'90*, pages 428–443, Amsterdam, The Netherlands, September 1990. Springer-Verlag.

[95] I. Niemelä. On the decidability and complexity of autoepistemic reasoning. *Fundamenta Informaticae*, 17(1,2):117–155, 1992.

[96] I. Niemelä. Autoepistemic logic as a unified basis for nonmonotonic reasoning. Doctoral dissertation. Research report A24, Helsinki University of Technology, Digital Systems Laboratory, Espoo, Finland, August 1993.

[97] I. Niemelä. A decision method for nonmonotonic reasoning based on autoepistemic reasoning. *J. Automat. Reason.*, 14:3–42, 1995.

[98] I. Niemelä. Towards efficient default reasoning. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 312–318, Montreal, Canada, August 1995. Morgan Kaufmann Publishers.

[99] I. Niemelä. Implementing circumscription using a tableau method. In W. Wahlster, editor, *Proceedings of the European Conference on Artificial Intelligence*, pages 80–84, Budapest, Hungary, August 1996. J. Wiley.

[100] I. Niemelä and J. Rintanen. On the impact of stratification on the complexity of nonmonotonic reasoning. *Journal of Applied Non-Classical Logics*, 4(2):141–179, 1994.

[101] D. Nute. Defeasible logic. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors. *Nonmonotonic Reasoning and Uncertain Reasoning*, *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, pages 353–395. Oxford University Press, Oxford, 1994.

[102] H.J. Ohlbach. SCAN—elimination of predicate quantifiers. In M.A. McRobbie and J.K. Slaney, editors. *Automated Deduction: CADE-13*, *Notes in Artificial Intelligence Lecture*, vol. 1104, pages 161–165. Springer, 1996.

[103] C. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[104] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge, TARK-90*, pages 121–135. Morgan Kaufmann, 1990.

[105] D. Perlis. Autocircumscription. *Artificial Intelligence*, 36(2):223–236, 1988.

[106] J.L. Pollock. Justification and defeat. *Artificial Intelligence*, 67(2):377–407, 1994.

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

[108] H. Prakken and G. Vreeswijk. Logical systems for defeasible argumentation. In D. Gabbay and F. Guenthner, editors. *Handbook of Philosophical Logic*, vol. 4, 2nd edition, pages 219–318. Kluwer Academic Publishers, Dordrecht, 2002.

[109] T.C. Przymusinski. An algorithm to compute circumscription. *Artificial Intelligence*, 38(1):49–73, 1989.

[110] R. Reiter. On closed world data bases. In H. Gallaire and J. Minker, editors. *Logic and Data Bases*, pages 55–76. Plenum Press, 1978.

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

[112] R. Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. 1st edition. MIT Press, Cambridge, 2001.

[113] J. Rintanen. On specificity in default logic. In *IJCAI*, pages 1474–1479, 1995.

[114] V. Risch and C. Schwind. Tableau-based characterization and theorem proving for default logic. *J. Automat. Reason.*, 13:223–242, 1994.

[115] E. Sandewall. An approach to the frame problem, and its implementation. In B. Meltzer and D. Michie, editors. *Machine Intelligence*, vol. 7, pages 195–204. Edinburgh University Press, 1972.

[116] T. Schaub. On commitment and cumulativity in default logics. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, ECSQARU-91*, *Lecture Notes in Computer Science*, vol. 548, pages 305–309. Springer, 1991.

[117] T. Schaub. A new methodology for query-answering in default logics via structure-oriented theorem proving. *J. Automat. Reason.*, 15(1):95–165, 1995.

[118] T. Schaub and S. Brüning. Prolog technology for default reasoning. In *Proceedings of the European Conference on Artificial Intelligence*, pages 105–109, Budapest, Hungary, August 1996. J. Wiley.

[119] T. Schaub and K. Wang. A comparative study of logic programs with preference. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence, IJCAI-01*, pages 597–602, 2001.

[120] J.S. Schlipf. How uncomputable is general circumscription? In *Proceedings of the Symposium on Logic in Computer Science*, pages 92–95, Cambridge, USA, June 1986. IEEE Computer Society Press.

[121] G.F. Schwarz and M. Truszczyński. Nonmonotonic reasoning is sometimes simpler. In *Proceedings of the 3rd K. Gödel Colloquium on Computational Logic and Proof Theory*, pages 313–324, Brno, Czech Republic, August 1993. Springer.

[122] G.F. Schwarz and M. Truszczyński. Minimal knowledge problem: a new approach. *Artificial Intelligence*, 67(1):113–141, 1994.

[123] Y. Shoham. A semantical approach to nonmonotic logics. In *Proceedings of the Symposium on Logic in Computer Science, LICS-87*, pages 275–279. IEEE Computer Society, 1987.

[124] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2–3):125–157, 1992.

[125] R. Stalnaker. A note on nonmonotonic modal logic. *Artificial Intelligence*, 64(2):183–196, 1993.

[126] J. Stillman. It's not my default: The complexity of membership problems in restricted propositional default logics. In *Proceedings of the 9th National Conference on Artificial Intelligence*, pages 571–578, Boston, MA, USA, July 1990. MIT Press.

[127] J. Stillman. The complexity of propositional default logics. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 794–800, San Jose, CA, USA, July 1992. MIT Press.

[128] M. Truszczyński. Modal interpretations of default logic. In *Proceedings of IJCAI-91*, pages 393–398. Morgan Kaufmann, 1991.

[129] G. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90˙1–2:225–279, 1997.

[130] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M.P. Singh, and M. Wooldridge, editors. *AAMAS*, pages 560–567. ACM, 2005.