HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A:** Research Reports ISSN 0783–5396
No. **24**; August 1993 ISBN 951–22–1689–2

AUTOEPISTEMIC LOGIC AS A UNIFIED BASIS FOR NONMONOTONIC REASONING

Ilkka Niemelä

Digital Systems Laboratory
Department of Computer Science
Helsinki University of Technology
Otaniemi, FINLAND

Helsinki University of Technology
Department of Computer Science
Digital Systems Laboratory
Otaniemi, Otakaari 1
SF–02150 ESPOO, FINLAND

HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A:** Research Reports
No. **24**; August 1993

ISSN 0783–5396
ISBN 951–22–1689–2

Autoepistemic Logic as a Unified Basis for Nonmonotonic Reasoning

ILKKA NIEMELÄ
Lic. Tech., Senior Assistant

Abstract: Autoepistemic logic is investigated in a general setting where autoepistemic reasoning based on a given classical logic is studied. The possible sets of autoepistemic conclusions from a set of premises are defined in terms of expansions of the premises. First Moore style expansions defined by a fixed point equation are considered and a simple finitary characterization of expansions is developed. An alternative definition is proposed where autoepistemic reasoning is formalized as a sequence of introspection steps using an enumeration of sentences. The resulting enumeration-based expansions are a proper subclass of Moore style expansions. An even more tightly grounded subclass is captured by considering only L-hierarchic enumerations. Finitary characterizations of enumeration-based and L-hierarchic expansions are developed. It is shown that autoepistemic reasoning based on Moore style, enumeration-based, and L-hierarchic expansions is decidable if the monotonic consequence relation given by the underlying classical logic is decidable. In the propositional case, decision problems related to the three classes of expansions are shown to be complete problems with respect to the second level of the polynomial time hierarchy. Default logic, forms of circumscription, answer set semantics of logic programs with classical negation, nonmonotonic truth maintenance systems, and forms of abduction are shown to be special cases of autoepistemic logic based on L-hierarchic expansions. Hence, this variant of autoepistemic logic offers a promising unified basis for nonmonotonic reasoning. A decision method for autoepistemic reasoning in the general setting is developed. The method requires as a subroutine only a theorem prover for the underlying monotonic consequence relation.

Keywords: Nonmonotonic reasoning, autoepistemic logic, computational complexity, automated theorem proving.

Dissertation for the degree of Doctor of Technology to be presented with due permission for public examination and debate in Auditorium E at Helsinki University of Technology (Espoo, Finland) on the 17th of September, 1993, at 12 o'clock noon.

Printing: TKK Monistamo; Otaniemi 1993

Helsinki University of Technology
Department of Computer Science
Digital Systems Laboratory
Otaniemi, Otakaari 1
SF–02150 ESPOO, FINLAND

Phone: $\dfrac{90}{+358\text{–}0}$ 4511
Telex: 125 161 htkk sf
Telefax: +358–0–465 077
E-mail: lab@hutds.hut.fi

Foreword

The results in this thesis have been obtained over a number of years. The ideas that form the basis of this work were developed as early as 1987. As a first result, a decision procedure for autoepistemic logic was devised [103]. A simple finitary characterization of autoepistemic reasoning containing the underlying principles of the decision procedure was then presented and decidability and complexity results on autoepistemic reasoning obtained [105, 106, 107]. A new variant of autoepistemic logic was developed [108]. This variant, presented in Chapter 5, is a key element of this work and it is based on ideas elaborated when studying the computational complexity of McDermott and Doyle's nonmonotonic logic [104]. The whole approach was then generalized so that autoepistemic logic based on a given classical logic can be studied [110]. Chapters 2–6 contain essentially the same material as in [110]. The results in Chapter 7 have appeared in [111] and the results in Chapter 8 have been presented in [109].

This thesis work has been carried out in the Digital Systems Laboratory of Helsinki University of Technology. I would like to express my warmest gratitude to the supervisor of this thesis, Professor Leo Ojala, for arranging an excellent working environment and for his continuous support. I would also like to thank my colleagues in the Digital Systems Laboratory for creating a stimulating working atmosphere. In particularly, I am indebted to Tomi Janhunen and Jussi Rintanen for their help and support during the final stages of the work as well as for their comments and interesting discussions.

The one year period at the beginning of the thesis work when the author was employed as a research assistant of the Academy of Finland was crucial to starting on the thesis research. From June 1989 the author has been holding the post of senior assistant in the Digital Systems Laboratory. The thesis work was completed when the author was employed as a researcher of the Academy of Finland. The financial support received from the following foundations is gratefully acknowledged: Emil Aaltonen Foundation, Foundation for Financial Aid at Helsinki University of Technology, Foundation of Technology, Heikki and Hilma Honkanen Foundation, Alfred Kordelin Foundation, Jenny and Antti Wihuri Foundation as well as Oskar Öflund Foundation.

Last and most, I thank my wife Tuula, my daughter Eeva-Leena, and my son Matias for their support, encouragement, and patience.

Espoo, August 1993

Ilkka Niemelä

Contents

Chapter 1

Introduction

Knowledge representation and reasoning have been among the central themes in artificial intelligence research since the very beginning of the field. In one of the first papers on knowledge representation, McCarthy [90] proposed classical first-order predicate logic as a foundation for knowledge representation. However, quite soon extensions to first-order logic were proposed which violated one of the characteristic features of logical reasoning: the monotonicity property.

> A logical calculus is said to be *monotonic* if its derivability relation $\vdash$ has the following property: if $\Sigma \vdash \phi$ and $\Sigma \subseteq \Sigma'$, then $\Sigma' \vdash p$.

1.1 Nonmonotonic Reasoning in Knowledge-Based Systems

Reasoning not satisfying the monotonicity property is said to be *nonmonotonic*. The operator *thnot* in PLANNER [45] was one of the first such nonmonotonic extensions of first-order logic. A similar operator *Unless* was suggested by Sandewall [125] in an attempt to solve the frame problem [94].

Example 1.1 The idea of nonmonotonic reasoning is illustrated by an example which is taken from Reiter [122]. A database of commercially available flights is given:

flight(Helsinki,New-York,Finnair)
flight(Helsinki,Stockholm,SAS)
⋮

If the database system is asked whether there is a flight from Oslo to Torino, the answer is "yes" if there is a piece of information in the database to that effect and otherwise the answer is "no". The "no" answers are derived using the *closed world assumption* [121] which informally states that all positive facts are given in the database and thus a positive fact not in the database is assumed not to hold. Of course, the conclusion that there is not a flight from Oslo to Torino is retracted if the database is extended by a fact about such a flight. Thus reasoning using the closed world assumption does not satisfy the monotonicity property. ■

Human commonsense reasoning and reasoning in artificial intelligence programs typically violates the monotonicity principle. People "jump" to conclusions which they may later retract when given more information. The conclusion that the local cafeteria is open at 2 p.m. on Friday is a typical example of a conclusion that would be retracted when given more information, e.g. that the cafeteria is under repair. Artificial intelligence programs must also be able to draw such conclusions. Even in the simple database example there is such an enormous amount of negative information that it cannot be recorded explicitly in practice. The closed world assumption provides a more effective method for representing negative information. Only the positive information needs to be represented explicitly and conclusions about the negative information are drawn on the basis of the closed world assumption.

Similar patterns of reasoning have been found useful in many contexts. In addition to databases where the *closed world assumption* [121] is used, logic programming where the *negation as failure* rule [15] is applied to implement a form of negation, inheritance hierarchies where properties are *inherited by default* [97], reasoning about action where the *frame problem* [94] and the *qualification problem* [91] have to be solved as well as *diagnosis* [123] are all examples of areas where reasoning patterns not satisfying the monotonicity property are used. As the list illustrates, nonmonotonic reasoning is one of the most fundamental phenomena in knowledge representation and reasoning. However, the exact nature of nonmonotonic reasoning is somewhat difficult to capture. The following example demonstrates the difficulties.

Example 1.2 Consider the following pieces of knowledge.

- A quaker who is not known to be a hawk is a pacifist.
- A republican who is not known to be a pacifist is a hawk.

We use a PROLOG system to represent the knowledge and interpret "not known" as PROLOG's negation (as failure). For simplicity we consider a given person so that no variables are needed. If we assume that the person is both a quaker and a republican, the resulting set of PROLOG rules is given below.

```
pacifist:- quaker, not(hawk).
hawk:- republican, not(pacifist).
quaker.
republican.
```
(1.1)

Given a query `pacifist`, a standard PROLOG system loops forever. But what is the right answer to the query? Usually in a knowledge representation system, answers are defined by giving a procedure which computes the answers by examining the knowledge base containing some symbolic representation of the pieces of knowledge. Assume that an answer to the query `pacifist` is provided by such a system. The answer is a function of the method of breaking the recursion in the two rules above and thus it is tightly coupled with a particular implementation of the procedure computing the answers. This kind of a situation makes concepts like negation as failure employed in knowledge representation difficult to understand and use. Maintenance and updating of such a system becomes tedious. Comparison of different systems is seriously hindered. Correctness of such a system is harder to assess. Areas of optimization are limited. A declarative way of characterizing correct answers would improve the situation considerably. By such a characterization we mean a way of defining correct answers without any reference to a particular procedure. But as the reasoning patterns defy one of the fundamental properties of classical logical reasoning, such a characterization is not straightforward to devise. In the case of logic programs this has resulted in a large number of competing proposals. For example, Przymusinska and Przymusinski [114] and Bidoit [7] offer surveys of the various approaches. ■

An obvious approach to characterizing nonmonotonic reasoning declaratively is to devise a suitable logical calculus that captures the correct reasoning patterns. It seems that classical logical calculi are insufficient for the purpose because virtually all of them satisfy the monotonicity property. Minsky [97] puts forward the monotonicity of logical calculi as one of the reasons why the logical approaches to capture commonsense reasoning will not work. However, the emphasis is here on the fact that classical monotonic logic is insufficient to solve the problems *effectively*. For example, the negative facts given by the closed world assumption could be added to the original database in principle. However, in practice it is virtually impossible to maintain a list of every fact that does not hold.

Because of the importance of nonmonotonic reasoning in artificial intelligence, a growing number of researchers have studied the formal underpinnings of nonmonotonic reasoning. The best symbolic approaches developed in the seventies were collected in a special issue of *Artificial Intelligence* (volume 13, 1980) on nonmonotonic logics. Ginsberg has compiled some of the most prominent papers on nonmonotonic reasoning which appeared before 1987 in a single edition [37]. From this work, default logic by Reiter [122], the family of circumscription type formalizations [92, 93], McDermott and Doyle style logics [96, 95], and autoepistemic logic [98] stand out as leading logical formalizations of nonmonotonic reasoning. The books of Etherington [26], Besnard [6], and Brewka [11] provide introductions to nonmonotonic reasoning.

The aim of this work is to strengthen the theoretical foundations of symbolic nonmonotonic reasoning. For that purpose we strive for a framework where nonmonotonic reasoning can be analysed as sentences in some formal language with well defined syntax and semantics. Such a declarative characterization of nonmonotonic reasoning would clarify the concepts used in knowledge representation systems, facilitate the comparison of different systems, permit enforcement of correctness and improve efficiency of the systems, e.g. by providing a basis for developing new optimization techniques. We focus on two questions:

- Computational complexity of nonmonotonic reasoning.

  Results on decidability and computational complexity of nonmonotonic reasoning are needed for evaluating the practical applicability of different approaches to nonmonotonic reasoning. If the complexity analysis is performed using a declarative characterization of nonmonotonic reasoning and without any reference to a particular implementation, more general results are obtained and sources of complexity in nonmonotonic reasoning are better understood. This leads to a clearer picture of factors affecting the complexity and indicates how to identify forms of nonmonotonic reasoning amenable to efficient implementation.

- Relationships between different forms of nonmonotonic reasoning.

  A huge number of formalizations of nonmonotonic reasoning have already been proposed. But what are the essential differences between the proposals? Are there some underlying principles of which various attempts are special cases? Clearly, uncovering such principles would clarify the field, disclose relationships between the approaches and lead to better understanding of the foundations of nonmonotonic reasoning.

In this work *autoepistemic logic* is studied as a theoretical foundation of nonmonotonic reasoning. There are several reasons why autoepistemic logic, introduced originally by Moore [98], stands out as a particularly attractive basis for nonmonotonic reasoning. Autoepistemic logic has a uniform syntax while, e.g., default logic has two categories of objects, first-order sentences and default rules. This leads to conceptual simplicity. Default rules typical in nonmonotonic reasoning as well as queries and integrity constraints can be given a uniform treatment. Autoepistemic logic is formalized as a direct extension of the underlying classical calculus where only simple introspection rules are added and, e.g., no modal axioms are needed as in the McDermott and Doyle style nonmonotonic modal logics. Autoepistemic logic appears to have greater expressive power than approaches such as circumscription which are based on minimal models. Furthermore, there is already a substantial body of results showing that autoepistemic logic offers a unified approach to several other modes of nonmonotonic reasoning (or at least substantial fragments of these). It has been shown that autoepistemic logic is closely related to default logic [61, 84], circumscription [63], McDermott and Doyle style nonmonotonic modal logics [135, 80], stable model semantics of logic programs [32], nonmonotonic truth maintenance systems [23], inheritance reasoning [35], and abduction [52].

1.2 Autoepistemic Logic

Autoepistemic logic is a modal logic with an operator L which is interpreted as "is believed". It was originally introduced by Robert Moore [98, 100, 101] as a reconstruction of McDermott and Doyle's nonmonotonic logic [96] to avoid some peculiarities of this logic. Autoepistemic logic models the beliefs of an ideally rational agent who is capable of perfect introspection. The interesting question is to determine what is the set of beliefs of the agent given a set of sentences as the initial assumptions or premises of the agent. The initial assumptions are required to be included in the set of beliefs. The agent's rationality means that the agent believes only logical consequences of the initial assumptions and her/his beliefs. The agent is capable of using both positive introspection (if χ is a belief, so is $L\chi$) and negative introspection (if χ is not a belief, then $\neg L\chi$ is). The agent is also ideal in the sense that she/he believes every logical consequence of her/his beliefs. The following example illustrates the concept of autoepistemic reasoning.

Example 1.3 Consider an autoepistemic sentence

$$\neg Lp \rightarrow \neg p. \tag{1.2}$$

The intuitive reading of this sentence is that if the agent does not believe p, then $\neg p$ holds (a kind of a closed world assumption). Given the sentence above as the initial assumption of the agent what is the set of beliefs of the agent?

Does, for example, $\neg p$ belong to the set of beliefs? In fact $\neg p$ is one of the beliefs of the agent. This can be shown informally as follows. The sentence p is not one of the beliefs of the agent because it is not a logical consequence of the initial assumption and the set of beliefs. Thus by negative introspection $\neg Lp$ is one of the beliefs of the agent and $\neg p$ is a logical consequence of this belief and the initial assumption.

If p is added to the initial premises, $\neg p$ no longer belongs to the set of beliefs of the agent. So autoepistemic reasoning is nonmonotonic in the sense that the set of beliefs does not grow monotonically when the set of premises is extended. ■

The informal description of the set of beliefs of the agent given above is circular: the set of beliefs is defined using the set of beliefs. This leads Moore [100] to use a fixed point definition. The sets of beliefs are defined as fixed points of an operator given with the aid of the logical consequence relation used by the agent. Fixed points are called *stable expansions* of the initial assumptions.

In this work we propose autoepistemic logic as a theoretical basis of nonmonotonic reasoning needed in knowledge-based systems. Autoepistemic logic fits this framework well: the autoepistemic agent corresponds to the inference engine of the knowledge-based system and the initial assumptions correspond to the knowledge base of the system. We investigate a general setting where autoepistemic logic is defined on top of a given classical logic. Moore's [100] original formalization of autoepistemic logic is based on the classical propositional calculus. Konolige [61] shows how to extend the definition to the first-order case where quantification into the scope of the belief operator is not allowed. We study autoepistemic reasoning in a more general setting where the underlying logic is not necessarily a first-order calculus but could be, e.g., any modal logic. We generalize Moore's definition of propositional autoepistemic logic to the point where given a classical logic CL an autoepistemic logic $\mathrm{CL}_{\mathrm{ae}}$ can be built on top of CL using an approach similar to that of Konolige [61]. This approach is quite different from McDermott and Doyle style nonmonotonic modal logics where the operator L is governed partly by simple negative introspection and partly by modal axioms. Here the operator L follows the simple and straightforward introspection rules of autoepistemic reasoning and only the underlying consequence relation used by the agent is changed. Not very surprisingly, stable expansions in $\mathrm{CL}_{\mathrm{ae}}$, in the autoepistemic logic induced by a classical logic CL, are defined by a similar fixed point equation as stable expansions in propositional autoepistemic logic. Only the consequence relation used in the fixed point equation has to be changed to reflect the change in the underlying logic.

The main contributions of the work fall into the following areas.

- Finitary characterization of stable expansions.

  Although the fixed point definition of stable expansions is compact and elegant, it provides no direct answers to some fundamental questions about autoepistemic reasoning. How many expansions can a set of premises have? When does a set of premises have at least one expansion or exactly one expansion? What is the structure of stable expansions? Can stable expansions be characterized finitely although they are infinite sets of sentences? How is such a finitary characterization derived from a stable expansion? How is the stable expansion constructed when given its characterization?

  We develop a finitary characterization of stable expansions and answer these questions with the aid of the characterization in the general setting. For each stable expansion a characterizing set, a kernel, is identified. The kernel is captured using a logical consequence relation induced by the underlying classical calculus. It is shown that there is a bijective mapping between the kernels and the stable expansions. Furthermore, it is shown how to construct the corresponding stable expansion from its kernel. The characterization provides a method for verifying the number of stable expansions of a set of premises and an upper bound for the number of stable expansions is established.

- Tightly grounded autoepistemic reasoning.

  The circular nature of the fixed point definition of Moore style stable expansions suggests that a stable expansion might contain beliefs justified by circular arguments. These kinds of weakly grounded beliefs are indeed possible. It seems that this circularity is closely related to the fixed point equation used for defining autoepistemic reasoning. The fixed point definition of stable expansions is also non-constructive. There is no direct notion of an autoepistemic proof similar to that in classical logical calculi.

  We propose a new, more constructive method for defining the sets of beliefs of an ideally rational agent with full introspection capabilities. It tries to carry the notion of a proof to autoepistemic logic. Instead of a fixed point equation, autoepistemic reasoning is defined as a sequence of introspection steps using an enumeration of sentences. The resulting enumeration-based expansions are a proper subclass of Moore style expansions. By considering a restricted class of enumerations, an interesting class of expansions is captured. These *L-hierarchic expansions* are very tightly grounded in the premises and do not suffer from problems typical of other approaches where tighter groundedness has been attempted. For example, L-hierarchic expansions are not dependent on the syntactic representation of the premises. A finitary characterization for enumeration-based and L-hierarchic expansions is developed.

- Decidability and complexity of autoepistemic reasoning.

  The finitary characterizations of expansions provide adequate tools for studying the computational properties of autoepistemic reasoning. We consider the three classes expansions: Moore style, enumeration-based, and L-hierarchic expansions. A number of decidability and complexity results are obtained in the general setting. It is shown that autoepistemic reasoning based on each of the three types of expansions is decidable if the logical consequence relation induced by the underlying classical calculus is decidable. In the propositional case, decision problems related to the three classes of expansions are shown to be complete problems with respect to the second level of the polynomial time hierarchy. This implies that propositional autoepistemic reasoning is strictly harder than classical propositional reasoning unless the polynomial time hierarchy collapses.

- Relationships between autoepistemic logic and other forms of nonmonotonic reasoning.

  We extend results on the relationships between autoepistemic logic and other forms of nonmonotonic reasoning. In particular, we argue that eliminating weakly grounded expansions does not reduce the coverage of autoepistemic logic as a unified framework for nonmonotonic reasoning. To substantiate the claim we show that Reiter's default logic, circumscription, answer set semantics of logic programs with classical negation, truth maintenance, and abduction can be embedded into autoepistemic logic based on L-hierarchic expansions. This indicates that autoepistemic logic based on L-hierarchic expansions provides a promising unified theoretical basis for nonmonotonic reasoning.

- A unified decision method for nonmonotonic reasoning.

  We develop a decision method for autoepistemic reasoning. The method is conceptually simple, isolates the subtasks corresponding to the sources of complexity in autoepistemic reasoning and reveals targets for optimizations. The method provides a unified reasoning engine for the large range of forms of nonmonotonic reasoning covered by autoepistemic logic.

1.3 Related Work

Autoepistemic logic is close to the logic of knowledge and ignorance of Halpern and Moses [42]. Also Levesque's [66] proposal is a kind of an autoepistemic logic where, however, the belief operator cannot appear in the set of premises.

Autoepistemic logic is most directly related to McDermott and Doyle's [96] nonmonotonic logic as it was designed to avoid some problems of that logic. McDermott and Doyle style nonmonotonic logics were further developed first by McDermott [95] and then by Shvarts [135]. Marek et al. [80] present a comprehensive study of the approach initiated by McDermott and Doyle. Schwarz [129] develops minimal model semantics for nonmonotonic modal logics. Konolige [60] proposes a hierarchical version of autoepistemic logic. Morris [102] introduces the notion of a stable closure, which is an extension of the notion of a stable expansion in the cases where there are no stable expansions. Levesque [68, 70] puts forward a monotonic logic to capture the patterns of autoepistemic reasoning. Schwarz [127, 130] introduces an autoepistemic logic of knowledge. Lifschitz [74] presents an autoepistemic circumscriptive logic. Przymusinski [116, 118] proposes a three-valued autoepistemic logic which is closely related to the well-founded semantics of logic programs [148, 149]. Przymusinski [117], Kakas and Mancarella [53] and Kalinski [54, 55] develop new variants of autoepistemic logic where the negative introspection rule is weaker than that in Moore's autoepistemic logic. Gelfond [31] introduces a notion of strong introspection.

Besides autoepistemic logic there are many interesting alternatives to be used as a logical foundation for nonmonotonic reasoning including default logic, McDermott and Doyle style nonmonotonic modal logics, and nonmonotonic logics based on two modalities [77, 78, 75]. Reiter's default logic captures many important forms of nonmonotonic reasoning. However, modal approaches seem to provide a more uniform and flexible logical basis for nonmonotonic reasoning. Furthermore, default logic can be embedded in several modal systems using simple local translations. McDermott and Doyle style nonmonotonic modal logics are closely related to autoepistemic logic. In fact, Shvarts [135] shows that the original autoepistemic logic can be seen as a particular McDermott and Doyle style nonmonotonic modal logic. In McDermott and Doyle style logics a fixed point equation is used for defining possible sets of correct nonmonotonic conclusions. This equation is similar to that employed in autoepistemic logic. Consequently, McDermott and Doyle style logics suffer from the same kinds of problems of weak groundedness as the original autoepistemic logic. The groundedness problem is investigated in depth in Chapter 4 and it turns out that there seems to be no direct way of eliminating weakly grounded beliefs when using this type of a fixed point equation. Lin and Shoham [77, 78] propose a formalization of nonmonotonic reasoning which employs two modal operators and a notion of minimal knowledge. They show how to embed default logic and original autoepistemic logic into their system. Lifschitz [75] presents a new formulation of the approach of Lin and Shoham which is not restricted to the propositional case. Lifschitz [75] as well as Schwarz and Truszczyński [131] point out that this minimal knowledge approach has a logically counterintuitive property: it is not conservative with respect to adding explicit definitions. Based on the work of Schwarz [129] on minimal model semantics, Schwarz and Truszczyński [131] propose a novel notion of minimal knowledge. They use a single modal operator and show that resulting minimal knowledge models correspond to expansions in a McDermott and Doyle style logic based on the modal logic **S4F**. The new notion does not suffer from the unintuitive property of the earlier approach. Schwarz and Truszczyński [131] indicate that the expressiveness of the new unimodal approach is essentially the same as that of the bimodal approach. In particular, they show that default logic and autoepistemic logic can be embedded into it. As the new notion of minimal knowledge corresponds to expansions in a McDermott and Doyle style nonmonotonic modal logic, it is prone to problems of weakly grounded beliefs. Lin and Shoham [76] introduce an argument system based approach to nonmonotonic reasoning. This approach covers, e.g., autoepistemic logic.

- Finitary characterization of stable expansions.

  The finitary characterization of stable expansions presented in this work is a variant of the finitary characterization developed by the author earlier [105, 106, 107]. The basic idea is that a stable expansion of a set of premises can be characterized using the $L\phi$ subformulae of the premises. Moore [101], Shvarts [135], and Stärk [137] have proposed similar characterizations based on the same idea. The finitary characterization of stable expansions presented in this work differs slightly from the earlier work by the author [105, 106, 107] as well as from the proposals of Moore [101], Shvarts [135], and Stärk [137]. The earlier approaches characterize stable expansions using the $L\phi$ subformulae of the premises but in the characterization presented in this work only the $L\phi$ subformulae of the premises which are not in the scope of another L operator are needed. For premises with nested L operators, this leads to a more compact finitary characterization and a tighter upper bound for the number of stable expansions. While the earlier characterizations address expansions in propositional autoepistemic logic, the finitary characterization presented in this work is shown to capture expansions in the general case where a given classical logic CL induces an autoepistemic logic $\mathrm{CL}_{\mathrm{ae}}$.

  Shvarts [135] shows that his approach can be used for devising finitary characterizations of expansions in several McDermott and Doyle style nonmonotonic modal logics. Shvarts extends this work together with Marek and Truszczyński [80]. Marek and Truszczyński [83, 87] present a finitary characterization of stable expansions in terms of their L free parts. Moore [99] and Halpern and Moses [42] observe that stable sets can be characterized using complete **S5**-structures. Gelfond [29] observes that the concept of stratification introduced in the context of logic programs [13, 1, 147] can be carried over to autoepistemic logic as a condition guaranteeing a unique stable expansion. Marek and Truszczyński [83, 87] extend Gelfond's notion of stratification. A more detailed comparison of the different approaches is presented in Chapter 3.

- Tightly grounded autoepistemic reasoning.

  The enumeration-based method of defining more tightly grounded expansions which is presented in this work (Chapter 5) has its roots in the ideas [104] elaborated when studying the computational complexity of McDermott and Doyle's nonmonotonic logic. Enumeration-based and L-hierarchic expansions were introduced first in the propositional case and results about the relationship between these classes of expansions and McDermott and Doyle style nonmonotonic logic as well as default logic were established [108].

  The first attempts to devise more tightly grounded versions of autoepistemic reasoning were developed with the aim of establishing a correspondence between autoepistemic reasoning and reasoning in Reiter's default logic. Konolige [61, 62] proposes the classes of moderately grounded and strongly grounded expansions. Marek and Truszczyński [84, 85, 86] continue the work and introduce several classes of expansions: iterative, strongly iterative, and robust expansions. Shvarts [135] and Marek et al. [80] indicate that some weakly grounded expansions can be eliminated by considering appropriate McDermott and Doyle style nonmonotonic modal logics. Tiomkin and Kaminski [140, 141] introduce nonmonotonic ground logic which is a special case of McDermott and Doyle style nonmonotonic modal logic where negative introspection is applied only to formulae without any modal operators. Kaminski [56], Truszczyński [146], and Konolige [65] continue this approach. A critical survey of the various approaches is presented in Chapter 4.

- Decidability and complexity of autoepistemic reasoning.

  Several researchers have made valuable contributions to the analysis of the decidability and complexity of autoepistemic reasoning. The decidability of propositional autoepistemic logic based on Moore style expansions has been established independently by several researchers [83, 87, 101, 103]. The author has shown that first-order autoepistemic logic where quantifying-in is not allowed is decidable if and only if the underlying first-order consequence relation is decidable [107]. Moore [101] as well as Junker and Konolige [50] report a similar result. Marek and Truszczyński [83, 87] show that the problem of existence of a Moore style expansion is **NP**-complete for a restricted subclass of propositional autoepistemic logic. In a later paper [88] they establish further complexity results on restricted classes of autoepistemic logic. The author has shown that decision problems related to Moore style expansions are on the second level of the polynomial time hierarchy [105, 107]. Gottlob [39, 40] extends this result and establishes the corresponding completeness results. He also shows similar completeness results with respect to the second level of the polynomial time hierarchy for decision problems related to N-expansions which coincide with iterative expansions. Reiter's default logic is closely related to several variants of autoepistemic logic. Gottlob [39, 40] and Stillman [139] have independently established completeness results with respect to the second level of the polynomial time hierarchy for reasoning in Reiter's default logic. These results complement the findings of Kautz and Selman [58, 59] and Stillman [138] which indicate that even very restricted subclasses of default reasoning are intractable. Eiter and Gottlob [22] establish complexity results for autoepistemic reasoning based on Konolige's moderately grounded expansions. Employing the approach of Shvarts [135] to characterizing expansions in McDermott and Doyle style logics Marek et al. [80] devise decision methods for several nonmonotonic modal logics.

- Relationships between autoepistemic logic and other forms of nonmonotonic reasoning.

The relationships between autoepistemic logic and other forms of nonmonotonic reasoning have been investigated by several researchers. Gelfond [30] demonstrates how Moore's autoepistemic logic can be used to formalize commonsense reasoning. Shvarts [135] shows that Moore style autoepistemic logic can be seen as a McDermott and Doyle's style nonmonotonic modal logic. Marek et al. [80] extend the relationships between nonmonotonic modal logics and the connections to Moore style autoepistemic logic. The first results on the close relationship between autoepistemic logic and Reiter's default logic were obtained by Konolige [61] who proposes a translation between autoepistemic logic and default logic. This translation is studied further by Marek and Truszczyński [84, 89]. Several other embeddings of default logic into variants of autoepistemic logic, McDermott and Doyle style nonmonotonic modal logics and other modal systems have been studied [41, 56, 75, 77, 78, 85, 86, 108, 111, 145, 144]. Konolige [63] shows how to capture circumscription in autoepistemic logic. The connections between different semantics of logic programs and autoepistemic logic have been investigated intensively [10, 23, 29, 32, 55, 81, 82, 116, 118, 117]. Autoepistemic logic is shown to provide a logical basis for nonmonotonic truth maintenance systems [23, 27, 120]. Gelfond and Przymusinska [35] investigate the formalization of inheritance reasoning in autoepistemic logic. Kakas and Mancarella [52] suggest a connection between abductive reasoning and autoepistemic logic.

- A unified decision method for nonmonotonic reasoning.

  The results mentioned above indicate that a decision procedure for autoepistemic logic could serve as a unified decision method for a large range of forms of nonmonotonic reasoning. The author has presented an analytic tableau based decision procedure for propositional Moore style autoepistemic logic [103]. Marek and Truszczyński [83, 87] devise a decision method for Moore style autoepistemic logic which is based on normal form transformations. Junker and Konolige [50] present a decision method for first-order default logic and autoepistemic logic where quantifying-in is not allowed. The method employs a truth maintenance system and is based on the finitary characterization of stable expansions using $L\phi$ subformulae of the premises. The premises are required to be in a normal form where there are no nested L operators and the formula whose membership in the expansions of the premises is determined cannot contain any L operators. Marek and Truszczyński [88] show how the problem of deciding whether a formula ϕ is in every stable expansion of a set of premises can be reduced to propositional derivability. The premises are required to be in a normal form where there are no nested L operators and ϕ cannot contain any L operators. Using the reduction they establish further complexity results for restricted subclasses of propositional autoepistemic reasoning. Stärk [137] shows that propositional autoepistemic logic based on Moore style stable expansions is decidable using his characterization of stable expansions. Marek et al. [80] present a decision method for propositional autoepistemic logic based on the characterization of stable expansions proposed by Shvarts [135]. Moore [101] devises a decision procedure for propositional autoepistemic logic based on the possible world characterization of stable sets. Moore [101] also states that if the underlying first-order logic is decidable, then the autoepistemic logic built on top of it is decidable. He gives a decision procedure which uses truth assignments and a transformation procedure of the premises.

The rest of the work is organized as follows. In Chapter 2 we extend the formalization of autoepistemic logic to the general setting where we are given an underlying classical logical calculus and we want to build an autoepistemic logic on top of it. Chapter 3 develops a finitary characterization of Moore style stable expansions in the general setting. Chapter 4 surveys critically the approaches to strengthen the definition of autoepistemic reasoning in order to capture expansions which are more tightly grounded in the premises. In Chapter 5 a new method for defining autoepistemic reasoning is introduced. The method solves problems related to the weak groundedness of Moore style stable expansions. In Chapter 6 results on the decidability and computational complexity of autoepistemic reasoning are developed. In Chapter 7 it is shown how other forms of nonmonotonic reasoning can be embedded into autoepistemic logic. Chapter 8 represents a decision method for autoepistemic reasoning and Chapter 9 contains the concluding remarks.

Chapter 2

Moore Style Autoepistemic Logic

Autoepistemic logic models the beliefs of an ideally rational agent who is capable of perfect introspection. The language of autoepistemic logic contains an explicit operator L for belief. The intuitive interpretation of $L\phi$ is that "the agent believes ϕ". Perfect introspection means that the agent is capable of using both

- positive introspection (if χ is a belief, so is $L\chi$) and
- negative introspection (if χ is not a belief, then $\neg L\chi$ is).

The fundamental question in autoepistemic logic is to determine the set of beliefs of the agent given a set of sentences as the initial assumptions or premises of the agent. The initial assumptions form the basis of the beliefs and the agent is assumed to be rational and ideal. A straightforward interpretation of these requirements leads to the following conditions on a set of beliefs of the agent.

1. The initial assumptions are included in the set of beliefs.
2. The agent's rationality means that the agent believes only logical consequences of the initial premises and her/his beliefs.
3. An ideal agent believes every logical consequence of her/his beliefs.

The conditions above are circular: the set of beliefs is characterized using the set of beliefs. Furthermore, it is not clear what is the logical consequence relation in question and how it should reflect the full introspection capabilities of the agent. Hence it is not entirely straightforward to provide a proper characterization of a set of beliefs of the agent.

2.1 Autoepistemic Logic $\mathrm{CL_{ae}}$

We study autoepistemic logic in a general setting where we are given a classical logical calculus CL and we are interested in the autoepistemic logic $\mathrm{CL_{ae}}$ based on CL. In this chapter we show how to define $\mathrm{CL_{ae}}$ by adding an autoepistemic layer on top of the given logic CL. The main task is to develop a precise definition of the sets of beliefs of an agent who is reasoning ideally and rationally according to the given logic CL and is capable of perfect introspection. Our definition generalizes Moore's [100] formalization of propositional autoepistemic logic, i.e. autoepistemic logic on top of the classical propositional calculus. It is based on the idea of decoupling the interpretation of the $L\phi$ formulae from the interpretation of other formulae. This idea has also been used, e.g., by Konolige [61].

Moore's [100] definition of propositional autoepistemic logic is based on the notion of a propositional interpretation of autoepistemic formulae. A propositional interpretation is an assignment of truth values to the autoepistemic formulae which is consistent with the usual truth recursion for the propositional calculus. An autoepistemic interpretation I of an autoepistemic theory Δ is a propositional interpretation that satisfies the following condition:

$$L\phi \text{ is true in } I \text{ if and only if } \phi \in \Delta.$$

The basic idea is that the autoepistemic theory is the set of beliefs of the agent. So $L\phi$ is true (the agent believes ϕ) if ϕ is in the set of beliefs of the agent.

An alternative way of seeing Moore's propositional interpretation of autoepistemic formulae is to decouple a propositional interpretation to a pair $\langle I, T\rangle$ where I is a truth assignment to the atomic formulae and T is a set of formulae which is used only to give truth values to the $L\phi$ formulae: $L\phi$ is true in $\langle I, T\rangle$ if and only if $\phi \in T$. Then $\langle I, T\rangle$ is an autoepistemic interpretation of Δ if $T = \Delta$. The decoupling approach is used as the basis for defining autoepistemic logic in the general setting considered in this work.

Let $\mathcal{L}$ be the language of CL. We extend $\mathcal{L}$ by adding a monadic operator L not appearing in $\mathcal{L}$ and obtain an autoepistemic language $\mathcal{L}_{\mathrm{ae}}$ which is the language of $\mathrm{CL_{ae}}$. The language $\mathcal{L}_{\mathrm{ae}}$ is defined recursively like $\mathcal{L}$ but with an extra formation rule:

$$\text{if } \phi \in \mathcal{L}_{\mathrm{ae}} \text{ is a sentence, then } L\phi \in \mathcal{L}_{\mathrm{ae}}.$$

Thus we consider autoepistemic logics where quantification into a modal context is not permitted (e.g. formulae of the form $\forall x LP(x)$ are not allowed). Problems related to allowing quantification into the scope of the L operator are briefly discussed in Section 2.2. To simplify the treatment we consider only closed formulae. So from here on we assume that $\mathcal{L}$ is the set of all closed formulae of CL and that $\mathcal{L}_{\mathrm{ae}}$ contains only closed formulae.

We assume that the semantics of CL is defined in the usual way. A set of CL-structures $\mathcal{S}$ is given. Then a relation $\models_{\mathrm{cl}}$ between the structures and sentences in $\mathcal{L}$ is defined. $I \models_{\mathrm{cl}} \phi$ is means that a sentence ϕ is true in a structure I. For a set of sentences Σ, $I \models_{\mathrm{cl}} \Sigma$ holds if and only if $I \models_{\mathrm{cl}} \phi$ holds for every $\phi \in \Sigma$. Such a structure is said to be a model of Σ. We also assume that CL contains the classical negation $\neg$: $\neg\phi$ is true if and only if ϕ is not true. This is because in autoepistemic logic $\mathrm{CL}_{\mathrm{ae}}$ we must be able to talk about not believing.

The next problem is to extend the semantics of CL to the richer language of $\mathrm{CL}_{\mathrm{ae}}$. Here the decoupling approach is employed. The structures in which the sentences of $\mathcal{L}_{\mathrm{ae}}$ are interpreted are divided into two parts. One part is a regular CL-structure which is used to interpret the operators of CL. The other is a set of sentences containing the beliefs of the agent. Thus the set of AE-structures $\mathcal{S}_{\mathrm{ae}}$ is the set of pairs $\langle I, T\rangle$ where $I \in \mathcal{S}$ and $T \subseteq \mathcal{L}_{\mathrm{ae}}$. A relation $\models_{\mathrm{ae}}$ between AE-structures and sentences in $\mathcal{L}_{\mathrm{ae}}$ is defined inductively on the formation of the sentences as $\models_{\mathrm{cl}}$ with the addition of a valuation rule for the new $L\phi$ sentences:

$$\langle I, T\rangle \models_{\mathrm{ae}} L\phi \text{ if and only if } \phi \in T.$$

$\langle I, T\rangle \models_{\mathrm{ae}} \phi$ means that a sentence ϕ is true in an AE-structure $\langle I, T\rangle$. So $L\phi$ is true in $\langle I, T\rangle$ if and only if ϕ is one of the beliefs of the agent given by the set T.

When determining the truth of a sentence containing no L operators the set of beliefs of the agent is not needed at all. Hence for all $\phi \in \mathcal{L}$,

$$\langle I, T\rangle \models_{\mathrm{ae}} \phi \text{ if and only if } I \models_{\mathrm{cl}} \phi. \tag{2.1}$$

This is because when evaluating the truth of a sentence $\phi \in \mathcal{L}_{\mathrm{ae}}$ in an AE-structure $\langle I, T\rangle$, the only time the set of beliefs T is used is when ϕ has a subformula of the form $L\chi$. Note that $L\chi$ sentences are treated like atomic sentences in the propositional calculus. They have a static truth value in an AE-structure depending only on whether χ is in the belief set or not. So the truth value of $L\chi$ does not depend on the internal structure of χ or the formula in which $L\chi$ is embedded.

Example 2.1 Let the underlying logic CL be the classical propositional calculus. So the set of CL-structures $\mathcal{S}$ is the set of truth assignments to atomic formulae of $\mathcal{L}$ and the valuation rules are the following.

1. $\langle I, T\rangle \models_{\rm ae} \phi$ if and only if $I \models_{\rm cl} \phi$ when ϕ is atomic.
2. $\langle I, T\rangle \models_{\rm ae} L\phi$ if and only if $\phi \in T$.
3. $\langle I, T\rangle \models_{\rm ae} \neg\phi$ if and only if $\langle I, T\rangle \not\models_{\rm ae} \phi$
4. $\langle I, T\rangle \models_{\rm ae} \phi \wedge \psi$ if and only if $\langle I, T\rangle \models_{\rm ae} \phi$ and $\langle I, T\rangle \models_{\rm ae} \psi$

Thus the basic semantics is obtained from the semantics of the underlying logic. A new valuation rule for $L\phi$ formulae is added (2.) and the set of structures used to give truth values to formulae is changed from the set of CL-structure S to a set of pairs $\langle I, T\rangle$ where $I \in S$ and $T \subseteq \mathcal{L}_{\rm ae}$. ∎

The decoupling approach induces a logical consequence relation $\models_{\rm ae}$ which is very similar to the logical consequence relation $\models_{\rm cl}$ of the underlying logic.

Definition 2.1 *$\Sigma \models_{\rm ae} \phi$ holds if and only if for all* CL-*structures I and for all sets of sentences T, $\langle I, T\rangle \models_{\rm ae} \Sigma$ implies $\langle I, T\rangle \models_{\rm ae} \phi$.*

This consequence relation is monotonic, treats $L\phi$ formulae like atomic formulae in the propositional calculus and behaves otherwise like the consequence relation of the underlying logic.

The consequence relation $\models_{\rm ae}$ does not, however, capture the introspection capabilities of the agent. Introspection can be obtained by parameterizing the consequence relation with the set of beliefs of the agent. We define for an agent having a set of beliefs Δ a consequence relation $\models_\Delta$ which is the underlying consequence relation $\models_{\rm ae}$ extended by introspection. The new consequence relation $\models_\Delta$ characterizes the conclusions drawn by an agent whose set of beliefs is Δ.

Definition 2.2 *$\Sigma \models_\Delta \phi$ if and only if for every* CL-*structure I, $\langle I, \Delta\rangle \models_{\rm ae} \Sigma$ implies $\langle I, \Delta\rangle \models_{\rm ae} \phi$.*

Introspection from the set Δ means that for any Σ, if ϕ is a belief of the agent ($\phi \in \Delta$), then $\Sigma \models_\Delta L\phi$ and if ϕ is not a belief of the agent ($\phi \notin \Delta$), then $\Sigma \models_\Delta \neg L\phi$.

The consequence relation $\models_\Delta$ gives the conclusions drawn by an ideal agent with perfect introspection capabilities having Δ as her/his set of beliefs. The requirements that the agent is rational and ideal remain to be formalized.

Moore [100] uses the concepts of soundness and semantical completeness. Soundness tries to capture the notion of rationality of the beliefs of the agent and completeness defines how an ideal agent reasons. Rationality means that the agent believes only logical consequences of the premises and her/his beliefs. Moore defines soundness in the following way. An autoepistemic theory Δ is sound with respect to a set of premises Σ if and only if every autoepistemic interpretation of Δ which is a model of Σ is a model of Δ. This means that a set of beliefs Δ is sound with respect to a set of premises Σ if every belief in Δ is true when ever every premise in Σ is true given that Δ is the set of beliefs under consideration. Thus Δ is sound with respect to Σ when every sentence in Δ follows from Σ using introspection from the set Δ. This suggests a notion of soundness also in the general setting using the consequence relation $\models_\Delta$.

Definition 2.3 *A set of sentences Δ is* sound *with respect to a set of premises Σ if and only if for all $\phi \in \Delta$, $\Sigma \models_\Delta \phi$.*

Moore defines Δ to be semantically complete if and only if Δ contains every formula that is true in every autoepistemic interpretation of Δ which is a model of Δ. A semantically complete set of beliefs contains everything that must be true given that the whole set of beliefs is true and it is the set of beliefs under consideration. Thus the semantic completeness of Δ means that if a sentence ϕ follows from the set of beliefs Δ using introspection from Δ, then ϕ must be in the set of beliefs Δ. This can be generalized straightforwardly using the consequence relation $\models_\Delta$.

Definition 2.4 *A set of sentences Δ is* semantically complete *if and only if for all $\phi \in \mathcal{L}_{\mathrm{ae}}$, $\Delta \models_\Delta \phi$ implies that $\phi \in \Delta$.*

Moore [100] argues that the possible sets of beliefs that an ideally rational agent might hold, given Σ as the agent's premises, ought to be just the extensions of Σ that are semantically complete and sound with respect to Σ. Moore calls such sets *stable expansions* of Σ. We adopt the same principle.

Definition 2.5 *A set of sentences Δ is a stable expansion of the sentences Σ if and only if the following conditions hold:*

1. $\Sigma \subseteq \Delta$,
2. *Δ is sound with respect to Σ and*
3. *Δ is semantically complete.*

Stable expansions can be characterized very elegantly using a fixed point equation.

Proposition 2.6 *A set of sentences Δ is a stable expansion of $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ if and only if Δ satisfies the following fixed point equation.*

$$\Delta = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \models_\Delta \phi\} \tag{2.2}$$

Proof. ($\Rightarrow$) Let Δ be a stable expansion of Σ. Then as Δ is sound with respect to Σ, $\Delta \subseteq \{\phi \mid \Sigma \models_\Delta \phi\}$. We prove that if Δ is semantically complete and $\Sigma \subseteq \Delta$, then $\{\phi \mid \Sigma \models_\Delta \phi\} \subseteq \Delta$. Let $\Sigma \models_\Delta \phi$. As $\models_\Delta$ is monotonic and $\Sigma \subseteq \Delta$, $\Delta \models_\Delta \phi$ and by semantic completeness $\phi \in \Delta$. Thus $\{\phi \mid \Sigma \models_\Delta \phi\} \subseteq \Delta$.

($\Leftarrow$) Let $\Delta = \{\phi \mid \Sigma \models_\Delta \phi\}$ hold. Then $\Sigma \subseteq \Delta$ and Δ is sound with respect to Σ as $\Delta \subseteq \{\phi \mid \Sigma \models_\Delta \phi\}$. Let $\Delta \models_\Delta \phi$. Assume $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma$. Then also $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Delta$ as for every $\psi \in \Delta$, $\Sigma \models_\Delta \psi$, i.e. for every $\psi \in \Delta$, $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma$ implies $\langle I, \Delta\rangle \models_{\mathrm{ae}} \psi$. Thus $\langle I, \Delta\rangle \models_{\mathrm{ae}} \phi$ as $\Delta \models_\Delta \phi$. We have shown that for every I, $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma$ implies $\langle I, \Delta\rangle \models_{\mathrm{ae}} \phi$. Thus $\Sigma \models_\Delta \phi$ and $\phi \in \Delta$ which shows that Δ is semantically complete. $\square$

An alternative, more syntactic characterization of stable expansions can be given using the consequence relation $\models_{\mathrm{ae}}$. This result is implied by the next proposition which gives the correspondence between the relations $\models_\Delta$ and $\models_{\mathrm{ae}}$.

We adopt the following notations. Let $\Delta \subseteq \mathcal{L}_{\mathrm{ae}}$ be a set of formulae.

$$L\Delta = \{L\phi \mid \phi \in \Delta\} \tag{2.3}$$

$$\neg\Delta = \{\neg\phi \mid \phi \in \Delta\} \tag{2.4}$$

$$\overline{\Delta} = \mathcal{L}_{\mathrm{ae}} - \Delta \tag{2.5}$$

We use frequently the notation $\neg L\overline{\Delta}$ which denotes the set $\{\neg L\phi \mid \phi \in \mathcal{L}_{\mathrm{ae}} - \Delta\}$.

Proposition 2.7 $\Sigma \models_\Delta \phi$ *if and only if* $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$*.*

Proof. ($\Rightarrow$) Let $\Sigma \models_\Delta \phi$. Let $\langle I, T\rangle$ be an AE-structure such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup L\Delta \cup \neg L\overline{\Delta}$. Then $T = \Delta$ which can be shown as follows. If $\psi \in \Delta$, then $L\psi \in L\Delta$ and thus $\langle I, T\rangle \models_{\mathrm{ae}} L\psi$ which implies that $\psi \in T$. Similarly, if $\psi \notin \Delta$, then $\neg L\psi \in \neg L\overline{\Delta}$ and thus $\langle I, T\rangle \models_{\mathrm{ae}} \neg L\psi$ which implies that $\psi \notin T$. So $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup L\Delta \cup \neg L\overline{\Delta}$ implies $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup L\Delta \cup \neg L\overline{\Delta}$ and thus $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma$. As $\Sigma \models_\Delta \phi$, then $\langle I, \Delta\rangle \models_{\mathrm{ae}} \phi$. We have shown that for every AE-structure $\langle I, T\rangle$, $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup L\Delta \cup \neg L\overline{\Delta}$ implies $\langle I, T\rangle \models_{\mathrm{ae}} \phi$. Therefore $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$.

($\Leftarrow$) Let $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$ hold. Let I be an CL-structure such that $\langle I, \Delta \rangle \models_{\mathrm{ae}} \Sigma$. Then $\langle I, \Delta \rangle \models_{\mathrm{ae}} \Sigma \cup L\Delta \cup \neg L\overline{\Delta}$ as $\langle I, \Delta \rangle \models_{\mathrm{ae}} L\Delta \cup \neg L\overline{\Delta}$ for every CL-structure I. Then as $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$ holds, $\langle I, \Delta \rangle \models_{\mathrm{ae}} \phi$ holds. We have proved that for every CL-structure I such that $\langle I, \Delta \rangle \models_{\mathrm{ae}} \Sigma$, $\langle I, \Delta \rangle \models_{\mathrm{ae}} \phi$ holds. Thus $\Sigma \models_{\Delta} \phi$. □

Proposition 2.7 implies a fixed point characterization of stable expansions similar to that in Moore's propositional autoepistemic logic.

Proposition 2.8 *A set of sentences Δ is a stable expansion of Σ if and only if Δ satisfies the following fixed point equation.*

$$\Delta = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi\} \tag{2.6}$$

Moore [100] shows that stable expansions can be characterized using the concepts of stability and groundedness. Stability is originally due to Stalnaker [136].

Definition 2.9 *A set of sentences Δ is* stable *if and only if it satisfies the following conditions:*

1. *If $\Delta \models_{\mathrm{ae}} \psi$, then $\psi \in \Delta$.*
2. *If $\phi \in \Delta$, then $L\phi \in \Delta$.*
3. *If $\phi \in \overline{\Delta}$, then $\neg L\phi \in \Delta$.*

Definition 2.10 *A set of sentences Δ is* grounded *in a set of premises Σ if and only if*

$$\Delta \subseteq \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi\}. \tag{2.7}$$

It is easy to show that stable expansions can be characterized using the concepts of groundedness and stability also in the general case.

Proposition 2.11 *A set of sentences Δ is grounded in Σ if and only if Δ is sound with respect to Σ.*

Proof. ($\Rightarrow$) Let Δ be grounded in Σ. If $\phi \in \Delta$, then $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$. By Proposition 2.7 $\Sigma \models_{\Delta} \phi$ holds. Thus Δ is sound with respect to Σ.

($\Leftarrow$) Let Δ be sound with respect to Σ. If $\phi \in \Delta$, then $\Sigma \models_{\Delta} \phi$. By Proposition 2.7 $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$ holds. Hence Δ is grounded in Σ. □

Proposition 2.12 *A set of sentences Δ is stable if and only if Δ is semantically complete.*

Proof. ($\Rightarrow$) Let Δ be stable. If $\Delta \models_\Delta \phi$, then by Proposition 2.7 $\Delta \cup L\Delta \cup \neg L\overline{\Delta} \models_{\text{ae}} \phi$. As Δ is stable, $L\Delta \cup \neg L\overline{\Delta} \subseteq \Delta$ and thus $\Delta \models_{\text{ae}} \phi$ by the monotonicity of $\models_{\text{ae}}$. This implies by the stability of Δ that $\phi \in \Delta$ and that Δ is semantically complete.

($\Leftarrow$) Let Δ be semantically complete. If $\Delta \models_{\text{ae}} \phi$, then $\Delta \models_\Delta \phi$ and by the semantical completeness of Δ $\phi \in \Delta$. If $\phi \in \Delta$, then $\Delta \models_\Delta L\phi$ and by the semantical completeness of Δ $L\phi \in \Delta$. If $\phi \in \overline{\Delta}$, then $\Delta \models_\Delta \neg L\phi$ and by the semantical completeness of Δ $\neg L\phi \in \Delta$. Thus Δ is stable. □

As a direct corollary we obtain a characterization of stable expansions in terms of stability and groundedness.

Proposition 2.13 *A set of sentences Δ is a stable expansion of a set of premises Σ if and only if*

1. $\Sigma \subseteq \Delta$,
2. *Δ is grounded in Σ and*
3. *Δ is a stable set.*

2.2 Quantifying-In

The extension of autoepistemic logic to allow quantification into the scope of the operator L is not unproblematic. Moore [101] argues that it is not conceptually clear how to define autoepistemic reasoning if full quantification is permitted. The example he considers is the following. Suppose the set of premises is

$$\{P(a), P(b), \forall x(\neg LP(x) \rightarrow \neg P(x))\} \tag{2.8}$$

and we are asked to decide whether $\neg P(c)$ follows from this set of premises. It seems that it should as the third premise says that for all objects x if the agent does not believe $P(x)$, then $\neg P(x)$ holds. But if $c = a$ or $c = b$, $P(c)$ follows from the set of premises and thus the agent believes it. The problem is that the quantifier ranges over objects in the domain and not over the names of the objects in the language of the logic. So if the agent believes that c has the property P even though under another name, this blocks the conclusion $\neg P(c)$. Thus if we are very careful, perhaps, we should infer from the set of premises (2.8) $((c \neq a) \wedge (c \neq b)) \rightarrow \neg P(c)$ instead of $\neg P(c)$.

Full quantification in autoepistemic reasoning is an interesting research topic. Konolige [63] considers first-order autoepistemic logic where quantification into the scope of the L operator is allowed but the nesting of the L operators is not. Thus, e.g., formulae of the form $LL\phi$ are not allowed. In a later paper Konolige [64] investigates three alternative semantics for quantification in autoepistemic logic. Levesque [68, 70] treats also quantification into the scope of the belief operator in his logic which is closely related to autoepistemic logic. He allows also nested belief operators. Levesque understands quantification substitutionally using a set of standard names.

In this work we do not consider full quantification primaly because of its unfavorable computational properties. Konolige [63, 64] shows that if full quantification is allowed, autoepistemic logic is able to capture minimization over countable model structures. Schlipf [126] shows that this kind of minimization is highly uncomputable. On the other hand, it seems that in many cases full quantification is not essential and can be avoided. For example, in the context of logic programs Herbrand interpretations are used [32, 149]. Here a quantified sentence can be seen as a shorthand for all its substitution instances where quantified variables are substituted by terms of the Herbrand base.

2.3 Supplementary Concepts and Results

We end the chapter by defining some additional concepts and providing two propositions which will to be used extensively in this work.

Definition 2.14 *For a set* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$, $\mathrm{Cn}(\Sigma) = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \models_{\mathrm{ae}} \phi\}$.

For a set $\Sigma \subseteq \mathcal{L}$, $\mathrm{Cn}_{\mathrm{cl}}(\Sigma) = \{\phi \in \mathcal{L} \mid \Sigma \models_{\mathrm{cl}} \phi\}$.

Definition 2.15 *For a formula* $\phi \in \mathcal{L}_{\mathrm{ae}}$, $\mathrm{Sf}^{\mathrm{L}}(\phi)$ *denotes the set of all subformulae of the form* $L\chi$ *of* ϕ *and* $\mathrm{Sf}^{\mathrm{qL}}(\phi)$ *denotes the set of all* $L\chi$ *quasi-subformulae of* ϕ*. Quasi-subformulae are subformulae in the usual sense except that* $L\phi$ *formulae are treated as atomic sentences having no further quasi-subformulae. Hence* $\mathrm{Sf}^{\mathrm{qL}}(\phi)$ *denotes the set of all* $L\chi$ *subformulae of* ϕ *which are not in the scope of an* L *operator in* ϕ*. For a set of formulae* Σ, $\mathrm{Sf}^{\mathrm{L}}(\Sigma) = \bigcup_{\phi\in\Sigma} \mathrm{Sf}^{\mathrm{L}}(\phi)$ *and similarly for* $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$.

A sentence ϕ *is said to be* quasi-atomic *if it begins with an* L *operator. Quasi-atomic formulae and their negations are called* quasi-literals.

For a quasi-literal χ, $\text{q-atom}(\chi)$ *denotes the corresponding quasi-atomic formula, i.e. for a quasi-atomic formula* χ, $\text{q-atom}(\chi) = \text{q-atom}(\neg\chi) = \chi$*. For a set of quasi-literals* Λ, $\text{q-atom}(\Lambda) = \{\text{q-atom}(\chi) \mid \chi \in \Lambda\}$.

Example 2.2 Let the underlying logic be the classical propositional calculus. The set of $L\chi$ subformulae of a formula $\phi \in \mathcal{L}_{\mathrm{ae}}$, $\mathrm{Sf}^{\mathrm{L}}(\phi)$, is defined inductively as follows.

1. $\mathrm{Sf}^{\mathrm{L}}(\phi) = \emptyset$ if ϕ is atomic.
2. $\mathrm{Sf}^{\mathrm{L}}(\neg\phi) = \mathrm{Sf}^{\mathrm{L}}(\phi)$.
3. $\mathrm{Sf}^{\mathrm{L}}(\phi \wedge \psi) = \mathrm{Sf}^{\mathrm{L}}(\phi) \cup \mathrm{Sf}^{\mathrm{L}}(\psi)$.
4. $\mathrm{Sf}^{\mathrm{L}}(L\phi) = \{L\phi\} \cup \mathrm{Sf}^{\mathrm{L}}(\phi)$.

The set $\mathrm{Sf}^{\mathrm{qL}}(\phi)$ is defined similarly except that $\mathrm{Sf}^{\mathrm{qL}}(L\phi) = \{L\phi\}$. ■

Definition 2.16 *The belief depth* $\mathrm{BD}(\phi)$ *of a sentence* $\phi \in \mathcal{L}_{\mathrm{ae}}$ *is the maximal depth of nesting of* L *operators in* ϕ.

1. *For an atomic sentence* ϕ, $\mathrm{BD}(\phi) = 0$,
2. $\mathrm{BD}(L\phi) = 1 + \mathrm{BD}(\phi)$,
3. *and for an operator op of* CL,

$$\mathrm{BD}(op(\phi_1, \ldots, \phi_n)) = max(\mathrm{BD}(\phi_1), \ldots, \mathrm{BD}(\phi_n)).$$

Example 2.3 Let the underlying logic be the classical propositional calculus and let $\Sigma = \{Lp, q \wedge L\neg L(Lp \wedge \neg q)\}$.

$$\begin{aligned}
\mathrm{Sf}^{\mathrm{L}}(\Sigma) &= \{Lp, L\neg L(Lp \wedge \neg q), L(Lp \wedge \neg q)\} \\
\mathrm{Sf}^{\mathrm{qL}}(\Sigma) &= \{Lp, L\neg L(Lp \wedge \neg q)\}
\end{aligned}$$

$$\text{q-atom}(\neg L(Lp \wedge \neg q))) = L(Lp \wedge \neg q)$$

$$\begin{aligned}
\mathrm{BD}(L\neg L(Lp \wedge \neg q)) &= 1 + \mathrm{BD}(\neg L(Lp \wedge \neg q)) \\
&= 1 + \mathrm{BD}(L(Lp \wedge \neg q)) \\
&= 2 + \mathrm{BD}((Lp \wedge \neg q)) \\
&= 2 + max(\mathrm{BD}(Lp), \mathrm{BD}(\neg q)) \\
&= 2 + max(1, 0) \\
&= 3
\end{aligned}$$

■

A key property of the underlying consequence relation $\models_{\mathrm{ae}}$ is stated in the following proposition.

Proposition 2.17 *If there is a set of quasi-literals* Λ *containing no formula and its negation and* $\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi$, *then there is a set* $\Lambda' \subseteq \Lambda$ *such that* $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \phi$ *and* $\text{q-atom}(\Lambda') \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\phi\})$.

Proof. The proof of the proposition is based on a simple observation that for any CL-structure I and any sets $T, T' \subseteq \mathcal{L}_{\mathrm{ae}}$ if for every $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$, $\chi \in T$ if and only if $\chi \in T'$, then

$$\langle I, T\rangle \models_{\mathrm{ae}} \phi \text{ if and only if } \langle I, T'\rangle \models_{\mathrm{ae}} \phi. \tag{2.9}$$

This observation is valid because when evaluating a sentence ϕ in an AE-structure $\langle I, T\rangle$ the only time the set T is needed is when ϕ has a quasi-subformula $L\chi$. If the two structures $\langle I, T\rangle$ and $\langle I, T'\rangle$ with the same CL-structure I agree on every such quasi-subformula, then they agree on the whole sentence ϕ.

Let $\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi$ hold and let

$$\Lambda' = \{\chi \in \Lambda \mid \text{q-atom}(\chi) \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})\}. \tag{2.10}$$

Then $\Lambda' \subseteq \Lambda$ and $\text{q-atom}(\Lambda') \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$ holds. We show that for Λ', $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \phi$ holds as well. Let $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda'$. Assume that $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$. Let

$$T' = T \cup \{\chi \mid L\chi \in \Lambda - \Lambda'\} - \{\chi \mid \neg L\chi \in \Lambda - \Lambda'\}. \tag{2.11}$$

As Λ contains no formula and its negation, if $L\chi \in \Lambda - \Lambda'$, then $\chi \in T'$ and if $\neg L\chi \in \Lambda - \Lambda'$, then $\chi \notin T'$. Thus $\langle I, T'\rangle \models_{\mathrm{ae}} \Lambda - \Lambda'$. We show that for all $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$, $\chi \in T$ if and only if $\chi \in T'$. Let $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$. Let $\chi \in T$ and assume that $\chi \notin T'$. Then $\neg L\chi \in \Lambda - \Lambda'$, which is a contradiction because $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$ and $\neg L\chi \in \Lambda$ imply that $\neg L\chi \in \Lambda'$. Thus $\chi \in T'$. Similarly, let $\chi \in T'$ and assume that $\chi \notin T$. Then $L\chi \in \Lambda - \Lambda'$ which is a contradiction because $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$ and $L\chi \in \Lambda$ imply that $L\chi \in \Lambda'$. Thus $\chi \in T$.

Hence if $\mathrm{Sf}^{\,\mathrm{qL}}(\Upsilon) \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$, for all $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Upsilon)$, $\chi \in T$ if and only if $\chi \in T'$. This implies by (2.9) that for any set Υ such that $\mathrm{Sf}^{\,\mathrm{qL}}(\Upsilon) \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$, $\langle I, T\rangle \models_{\mathrm{ae}} \Upsilon$ if and only if $\langle I, T'\rangle \models_{\mathrm{ae}} \Upsilon$. So as $\mathrm{Sf}^{\,\mathrm{qL}}(\Sigma) \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$ and $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma$, $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma$. Similarly, as $\mathrm{Sf}^{\,\mathrm{qL}}(\phi) \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$, $\langle I, T'\rangle \not\models_{\mathrm{ae}} \phi$ and as $\mathrm{Sf}^{\,\mathrm{qL}}(\Lambda') \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\phi\})$, $\langle I, T'\rangle \models_{\mathrm{ae}} \Lambda'$. This means that $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda$ but $\langle I, T'\rangle \not\models_{\mathrm{ae}} \phi$ which contradicts the fact that $\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi$ holds. Thus $\langle I, T\rangle \models_{\mathrm{ae}} \phi$ holds for every AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda'$ holds. Hence $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \phi$. □

We define a simple consequence relation $\models_{\mathrm{L}}$, which is given recursively using the underlying consequence relation $\models_{\mathrm{ae}}$. The idea here is to incorporate the introspection principles of autoepistemic reasoning into the consequence relation so that $L\phi$ sentences are no longer treated like atomic sentences.

Definition 2.18 *Given a set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ and a sentence $\phi \in \mathcal{L}_{\mathrm{ae}}$,*

$$\Sigma \models_{\mathrm{L}} \phi \textit{ if and only if } \Sigma \cup \mathrm{SB}_{\Sigma}(\phi) \models_{\mathrm{ae}} \phi$$

$$\textit{where}\quad \mathrm{SB}_{\Sigma}(\phi) = \{L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi) \mid \Sigma \models_{\mathrm{L}} \chi\} \cup \{\neg L\chi \in \neg \mathrm{Sf}^{\,\mathrm{qL}}(\phi) \mid \Sigma \not\models_{\mathrm{L}} \chi\}.$$

Example 2.4 Let the underlying logic be the classical propositional calculus and let p be an atomic sentence. Then $\emptyset \models_{\mathrm{L}} \neg Lp$ holds. This can be verified as follows. As $\mathrm{Sf}^{\,\mathrm{qL}}(p) = \emptyset$, $\mathrm{SB}_{\emptyset}(p) = \emptyset$. So $\emptyset \not\models_{\mathrm{L}} p$ holds because $\emptyset \not\models_{\mathrm{ae}} p$ holds. This implies that $\mathrm{SB}_{\emptyset}(\neg Lp) = \{\neg Lp\}$. Thus $\emptyset \models_{\mathrm{L}} \neg Lp$ holds.

On the other hand, $\{p\} \models_{\mathrm{L}} p$ holds because $\{p\} \models_{\mathrm{ae}} p$ holds. This implies that $\mathrm{SB}_{\{p\}}(\neg Lp) = \{Lp\}$. Thus $\{p\} \models_{\mathrm{L}} \neg Lp$ does not hold. ■

Example 2.4 shows that $\models_{\mathrm{L}}$ is a nonmonotonic consequence relation.

The last proposition of this chapter states a property which is useful in showing that two stable sets with a common L free part are equal.

Proposition 2.19 *Let the underlying logic be a first-order predicate calculus. Let S_1 and S_2 be two stable sets such that $S_1 \cap \mathcal{L} \subseteq S_2 \cap \mathcal{L}$. Then if for all χ such that $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$, $\chi \in S_1$ if and only if $\chi \in S_2$, then $\phi \in S_1$ implies $\phi \in S_2$.*

Proof. Let $\phi \notin S_2$. Konolige [61, Proposition 3.7] shows that a formula ψ of first-order autoepistemic logic without nested L operators can be transformed into an equivalent normal form $d_1 \wedge \ldots \wedge d_n$ where each d_i is of the form

$$\gamma \vee L\alpha_1 \vee \ldots \vee L\alpha_m \vee \neg L\beta_1 \vee \ldots \vee \neg L\beta_l$$

where $\gamma, \alpha_i, \beta_i \in \mathcal{L}$. The transformation treats $L\chi$ subformulae of ψ like atomic formulae and the same transformation can be used to put an arbitrary formula $\phi \in \mathcal{L}_{\mathrm{ae}}$ into an equivalent normal form $\phi' = d_1 \wedge \ldots \wedge d_n$ such that each d_i is of the form

$$\gamma \vee L\alpha_1 \vee \ldots \vee L\alpha_m \vee \neg L\beta_1 \vee \ldots \vee \neg L\beta_l$$

where $\gamma \in \mathcal{L}$ and $\alpha_i, \beta_i \in \mathcal{L}_{\mathrm{ae}}$ and, furthermore, $\mathrm{Sf}^{\,\mathrm{qL}}(\phi) = \mathrm{Sf}^{\,\mathrm{qL}}(\phi')$.

It holds that $\phi \in S_2$ if and only if $\phi' \in S_2$. So $\phi' \notin S_2$. Thus there exists a disjunction $d_j \notin S_2$. Halpern and Moses [42] observe that a stable set S has the following properties.

$$L\chi \vee \chi' \in S \quad \text{if and only if} \quad \chi \in S \text{ or } \chi' \in S \tag{2.12}$$
$$\neg L\chi \vee \chi' \in S \quad \text{if and only if} \quad \chi \notin S \text{ or } \chi' \in S \tag{2.13}$$

Let $L\chi \vee \chi' \in S$. If $\chi \notin S$, then $\neg L\chi \in S$ as S is stable and thus $\chi' \in S$ as S is closed under first-order consequence. The other direction is immediate. Hence (2.12) holds.

Let $\neg L\chi \vee \chi' \in S$. If $\chi \in S$, then $L\chi \in S$ as S is stable and thus $\chi' \in S$ as S is closed under first-order consequence. The other direction is immediate. Hence (2.13) holds.

By (2.12) and (2.13) $d_j \notin S_2$ if and only if $\gamma \notin S_2, \alpha_1 \notin S_2, \ldots, \alpha_m \notin S_2, \beta_1 \in S_2, \ldots, \beta_l \in S_2$. The normal form transformation does not affect the $L\chi$ formulae and thus $\mathrm{Sf}^{\,\mathrm{qL}}(\phi) = \mathrm{Sf}^{\,\mathrm{qL}}(\phi')$. Then for each α_i and β_i, $L\alpha_i \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$ and $L\beta_i \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$. This implies $\alpha_1 \notin S_1, \ldots, \alpha_m \notin S_1, \beta_1 \in S_1, \ldots, \beta_l \in S_1$. As $\gamma \in \mathcal{L}$ and $\gamma \notin S_2$, $\gamma \notin S_1$. Thus $d_j \notin S_1$ which implies $\phi' \notin S_1$. As ϕ is equivalent to ϕ', $\phi \notin S_1$. Hence $\phi \in S_1$ implies $\phi \in S_2$. $\square$

Chapter 3

A Finitary Characterization of Moore Style Stable Expansions

A stable expansion is the key concept in autoepistemic reasoning. But stable expansions are hard to handle and reason about. Firstly, stable expansions are infinite. Since a stable expansion is a stable set, if χ is in a stable expansion, then so are, e.g., $L\chi, LL\chi, LLL\chi \ldots$ Secondly, stable expansions are defined non-constructively using a fixed point equation (2.2) involving a context sensitive consequence relation $\models_\Delta$. The fixed point equation (2.6) for stable expansions is little easier to reason about as the context is lifted to the set of premises, but the problem of non-constructiveness still remains. Furthermore, there are many interesting questions about stable expansions which are hard to answer on the basis of the fixed point definition.

- Are there sets of premises having a stable expansion?
- How many stable expansions can a set of premises have?
- Which kind of premises have at least one expansion or exactly one expansion?
- Can an infinite stable expansion be represented finitely? How is a finitary representation derived from a stable expansion? How is a stable expansion constructed when given its finitary characterization?

In this chapter our goal is to answer these kinds of questions about the structure and properties of stable expansions by developing a simple finitary characterization of stable expansions. The finitary characterization provides also the basis for the complexity analysis and computational treatment of autoepistemic reasoning to be developed.

There is a number of attempts to find a finitary characterization of Moore style stable expansions. Most of them stay within the propositional case only but there are some finitary characterizations in the first-order case without quantifying-in. Some of the characterizations seem to rely on quite restricting assumptions on the underlying logic, e.g. a particular normal form is required. So they are not directly applicable in the general setting of this work.

The finitary characterization of stable expansions presented in this chapter is a modification of an earlier characterization developed by the author [105, 106, 107]. Characterizations which are quite similar to the earlier characterization have been developed independently by Moore [101], Shvarts [135], and Stärk [137]. At the end of the chapter we compare these characterizations to the one to be presented.

Marek and Truszczyński [87] present a finitary characterization of stable expansions in terms of their L free parts. The characterization applies to stable expansions of sets of premises in a particular normal form and captures only consistent stable expansions.

Moore [99] and Halpern and Moses [42] observe that in the propositional case stable sets can be characterized using complete **S5**-structures. This makes it possible to characterize a stable set using a finite structure provided that the language of the underlying logic in question has only finitely many atomic formulae. As stable expansions are stable sets, it is possible to characterize stable expansions using **S5**-structures.

Sets of premises with a single stable expansion are of special interest because then the notion of autoepistemic reasoning is clearly defined: ϕ is a correct autoepistemic conclusion from a set of premises if and only if it belongs to the unique stable expansion of the premises. Gelfond [29] shows that the concept of stratification introduced in the context of logic programs can be used as a condition guaranteeing a unique stable expansion. Marek and Truszczyński [87] generalize Gelfond's notion of stratification.

We present a finitary characterization of stable expansions that relies on sets of sentences constructed from the $L\chi$ quasi-subformulae of the premises satisfying a special fullness condition. Full sets defined below serve as kernels of stable expansions; a full set uniquely characterizes a stable expansion.

Definition 3.1 *For a set of sentences Σ, a set $\Lambda \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ is Σ-full if the following two conditions hold for every $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$:*

1. $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ *if and only if* $L\chi \in \Lambda$.

2. $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$ *if and only if* $\neg L\chi \in \Lambda$.

As an immediate consequence of the definition of a *full* set we note the following.

Proposition 3.2 *Let* Λ *be a* Σ*-full set. Then for every* $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$*, either* $L\chi \in \Lambda$ *or* $\neg L\chi \in \Lambda$.

Clearly, the definition of a full set is circular and leads to a fixed point equation.

Proposition 3.3 Λ *is a* Σ*-full set if and only if* $\Lambda = \mathrm{L}^{+}_{\Sigma}(\Lambda) \cup \mathrm{L}^{-}_{\Sigma}(\Lambda)$
where $\mathrm{L}^{+}_{\Sigma}(\Lambda) = \{L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma) \mid \Sigma \cup \Lambda \models_{\mathrm{L}} \chi\}$ *and*
$\mathrm{L}^{-}_{\Sigma}(\Lambda) = \{\neg L\chi \in \neg\mathrm{Sf}^{\mathrm{qL}}(\Sigma) \mid \Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi\}$.

Proof. ($\Rightarrow$) Let Λ be a Σ-full set. If $L\chi \in \mathrm{L}^{+}_{\Sigma}(\Lambda)$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ and thus $L\chi \in \Lambda$ as Λ is a Σ-full set. Similarly, $\neg L\chi \in \mathrm{L}^{-}_{\Sigma}(\Lambda)$ implies that $\neg L\chi \in \Lambda$. Hence $\mathrm{L}^{+}_{\Sigma}(\Lambda) \cup \mathrm{L}^{-}_{\Sigma}(\Lambda) \subseteq \Lambda$. If $L\chi \in \Lambda$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ and thus $L\chi \in \mathrm{L}^{+}_{\Sigma}(\Lambda)$. Similarly, $\neg L\chi \in \Lambda$ implies that $\neg L\chi \in \mathrm{L}^{-}_{\Sigma}(\Lambda)$. Hence $\Lambda = \mathrm{L}^{+}_{\Sigma}(\Lambda) \cup \mathrm{L}^{-}_{\Sigma}(\Lambda)$.

($\Leftarrow$) Let $\Lambda = \mathrm{L}^{+}_{\Sigma}(\Lambda) \cup \mathrm{L}^{-}_{\Sigma}(\Lambda)$ hold. Let $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. If $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$, then $L\chi \in \mathrm{L}^{+}_{\Sigma}(\Lambda) \subseteq \Lambda$. If $L\chi \in \Lambda$, then $L\chi \in \mathrm{L}^{+}_{\Sigma}(\Lambda)$ and thus $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. Similarly, if $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$, then $\neg L\chi \in \mathrm{L}^{-}_{\Sigma}(\Lambda) \subseteq \Lambda$. If $\neg L\chi \in \Lambda$, then $\neg L\chi \in \mathrm{L}^{-}_{\Sigma}(\Lambda)$ and thus $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$. Hence Λ satisfies the conditions for a Σ-full set. □

Example 3.1 Let the underlying logic CL be the propositional calculus. Let $\Sigma = \{Lp_1 \to p_1, Lp_2 \to p_2\}$ where p_1 and p_2 are atomic sentences. Then $\mathrm{Sf}^{\mathrm{qL}}(\Sigma) = \{Lp_1, Lp_2\}$. The set $\Lambda = \{Lp_1, \neg Lp_2\}$ is Σ-full as $\Sigma \cup \Lambda \models_{\mathrm{L}} p_1$ and $\Sigma \cup \Lambda \not\models_{\mathrm{L}} p_2$. For example, sets $\{Lp_1, \neg Lp_1, \neg Lp_2\}$ and $\{\neg Lp_2\}$ are not Σ-full. ■

Next we prove two fundamental lemmata relating full sets and stable expansions. The proofs are based on Proposition 2.17 and the fact that the consequence relation $\models_{\mathrm{ae}}$ is monotonic, i.e. if $\Sigma \models_{\mathrm{ae}} \phi$ and $\Sigma \subseteq \Sigma'$, then $\Sigma' \models_{\mathrm{ae}} \phi$. The first lemma states that for every stable expansion, there is a corresponding full set. The following proposition contains the key result needed in the proof of the lemma.

Proposition 3.4 *Let* $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$ *hold. Then for all* $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$*,* $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ *if and only if* $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ *where* $\Lambda = \{\chi \in L\Delta \cup \neg L\overline{\Delta} \mid \text{q-atom}(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$.

Proof. We prove the proposition by induction on the belief depth of χ, $\mathrm{BD}(\chi)$. Let $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. Let $\mathrm{BD}(\chi) = 0$. Then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ if and only if $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ and as $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$, $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. Let $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ hold. By Proposition 2.17 there is a set $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and $\text{q-atom}(\Lambda') \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\chi\}) = \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. For every $\chi' \in \Lambda'$, as $\chi' \in L\Delta \cup \neg L\overline{\Delta}$, then $\chi' \in \Lambda$. Thus $\Lambda' \subseteq \Lambda$ and $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$.

Assume that for all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ with $\mathrm{BD}(\chi) < i$,

$$\Sigma \cup \Lambda \models_{\mathrm{L}} \chi \text{ if and only if } \Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi.$$

Let $\mathrm{BD}(\chi) = i$. Let $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. Then $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\chi) \models_{\mathrm{ae}} \chi$. If $L\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi'$ and $L\chi' \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$. As $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, then $L\chi' \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. Because $\mathrm{BD}(\chi') < i$, $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi'$ by the inductive hypothesis. Thus $\chi' \in \Delta$ and $L\chi' \in L\Delta$. Similarly, $\neg L\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi')$ implies $\neg L\chi' \in \neg L\overline{\Delta}$. Thus $\mathrm{SB}_{\Sigma\cup\Lambda}(\chi) \subseteq L\Delta \cup \neg L\overline{\Delta}$ and $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. On the other hand, let $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ hold. By Proposition 2.17 there is a set $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and for every $\chi' \in \Lambda'$, $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\chi\})$. If $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$, then $\chi' \in \Lambda$ because $\chi' \in L\Delta \cup \neg L\overline{\Delta}$. If $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$, then as $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. This implies $\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$ in the following way. If $\chi' = L\chi''$, then $\chi'' \in \Delta$ and as $\mathrm{BD}(\chi'') < i$, by the inductive hypothesis $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi''$ and thus $\chi' = L\chi'' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$. Similarly, if $\chi' = \neg L\chi''$, then $\chi'' \notin \Delta$ and by the inductive hypothesis $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi''$ and thus $\chi' = \neg L\chi'' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$. Hence $\Lambda' \subseteq \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$ and $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. □

Lemma 3.5 *If* $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$, *then*

$$\Lambda = \{\chi \in L\Delta \cup \neg L\overline{\Delta} \mid \text{q-atom}(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\} \tag{3.1}$$

is Σ*-full.*

Proof. Let $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. Let $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. By Proposition 3.4 $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$. So $\chi \in \Delta$ and $L\chi \in L\Delta$. Thus $L\chi \in \Lambda$. On the other hand, let $L\chi \in \Lambda$. So $\chi \in \Delta$. Thus $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ and by Proposition 3.4 $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. Hence Condition 1 of Definition 3.1 holds. If $\neg L\chi \in \Lambda$, then $\chi \notin \Delta$. Then by Proposition 3.4 $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$. If $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$, then $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \not\models_{\mathrm{ae}} \chi$ by Proposition 3.4. Thus $\chi \notin \Delta$ and $\neg L\chi \in \Lambda$. Hence Condition 2 holds as well and Λ is Σ-full. □

Lemma 3.5 shows that for every stable expansion of Σ, there is a Σ-full set. We aim to show that the converse holds as well. But first we note that if a full set is contained in the set of beliefs induced by a set of sentences, then it agrees completely with the set of beliefs.

Proposition 3.6 *Let Δ be a set of sentences of $\mathcal{L}_{\mathrm{ae}}$ and let Λ be a Σ-full set such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Then for every χ such that* $\text{q-atom}(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$, *if $\chi \in L\Delta \cup \neg L\overline{\Delta}$, then $\chi \in \Lambda$.*

Proof. Let $\text{q-atom}(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. Let $\chi \in L\Delta$. Then $\chi = L\chi'$. Assume that $L\chi' \notin \Lambda$. Then by Proposition 3.2 $\neg L\chi' \in \Lambda$ and thus $\neg L\chi' \in L\Delta \cup \neg L\overline{\Delta}$. So $\chi' \notin \Delta$ and $L\chi' \notin L\Delta$, a contradiction. Hence $L\chi' \in \Lambda$.

Let $\chi \in \neg L\overline{\Delta}$. Then $\chi = \neg L\chi'$. Assume that $\neg L\chi' \notin \Lambda$. Then $L\chi' \in \Lambda$ and thus $L\chi' \in L\Delta \cup \neg L\overline{\Delta}$. So $\chi' \in \Delta$ and $\neg L\chi' \notin \neg L\overline{\Delta}$, a contradiction. Hence $\neg L\chi' \in \Lambda$. □

The other fundamental lemma states that for each full set, there is a stable expansion containing the full set.

Lemma 3.7 *Let Σ be a set of sentences of $\mathcal{L}_{\mathrm{ae}}$ and Λ a Σ-full set. Then for*

$$\Delta = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \Lambda \models_{\mathrm{L}} \phi\} \tag{3.2}$$

$\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$ and $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$.

Proof. First we prove that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. If $L\chi \in \Lambda$, then as Λ is Σ-full, $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. Thus $\chi \in \Delta$ and $L\chi \in L\Delta$. Similarly, if $\neg L\chi \in \Lambda$, then as Λ is Σ-full, $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$. Thus $\chi \notin \Delta$ and $\neg L\chi \in \neg L\overline{\Delta}$.

We show that $\phi \in \Delta$ if and only if $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$. Let $\phi \in \Delta$. Then $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\phi) \models_{\mathrm{ae}} \phi$. If $L\chi \in \mathrm{SB}_{\Sigma\cup\Lambda}(\phi)$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. Thus $\chi \in \Delta$ and $L\chi \in L\Delta$. Similarly, if $\neg L\chi \in \mathrm{SB}_{\Sigma\cup\Lambda}(\phi)$, then $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi$, $\chi \notin \Delta$ and $\neg L\chi \in \neg L\overline{\Delta}$. Thus $\mathrm{SB}_{\Sigma\cup\Lambda}(\phi) \subseteq L\Delta\cup\neg L\overline{\Delta}$. Hence $\Lambda\cup\mathrm{SB}_{\Sigma\cup\Lambda}(\phi) \subseteq L\Delta\cup\neg L\overline{\Delta}$ holds which implies $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. On the other hand, let $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$. By Proposition 2.17 there is a set $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \phi$ and for every $\chi' \in \Lambda'$, $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\phi\})$. If $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$, then $\chi' \in \Lambda$ by Proposition 3.6. If $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\phi)$, then $\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\phi)$ which can be shown in the following way. If $\chi' = L\chi''$, then $\chi'' \in \Delta$ and thus $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi''$ and so $\chi' = L\chi'' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\phi)$. Similarly, if $\chi' = \neg L\chi''$, then $\chi'' \notin \Delta$, $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi'$ and thus $\chi' = \neg L\chi'' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\phi)$. So $\Lambda' \subseteq \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\phi)$ and $\Sigma \cup \Lambda \models_{\mathrm{L}} \phi$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus $\phi \in \Delta$. Hence $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$. □

As a consequence of Lemmata 3.5 and 3.7 we are able to provide a necessary and sufficient condition ensuring that a set of premises has at least one stable expansion.

Theorem 3.8 *A set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ has a stable expansion if and only if there exists a Σ-full set.*

Proof. If Σ has a stable expansion, then by Lemma 3.5 there is a Σ-full set Λ. If there is a Σ-full set Λ, then by Lemma 3.7 there is a stable expansion. $\square$

Using the theorem we can formally show that premises having a stable expansion exist.

Example 3.2 Let the underlying logic CL be the propositional calculus. Let $\Sigma = \{\neg Lp \to \neg p\}$, where p is an atomic sentence. As $\mathrm{Sf}^{\mathrm{qL}}(\Sigma) = \{Lp\}$, there are two possible candidates for Σ-full sets: $\Lambda_1 = \{Lp\}$ and $\Lambda_2 = \{\neg Lp\}$. The set $\Lambda_1 = \{Lp\}$ is not Σ-full because $\Sigma \cup \Lambda_1 \not\models_{\mathrm{L}} p$ but $\Lambda_2 = \{\neg Lp\}$ is Σ-full because $\Sigma \cup \Lambda_2 \not\models_{\mathrm{L}} p$. So Σ has a stable expansion $\Delta = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \Lambda_2 \models_{\mathrm{L}} \phi\}$ by Lemma 3.7. $\blacksquare$

In fact there is a one-to-one correspondence between full sets and stable expansions.

Theorem 3.9 *Let Σ be a set of sentences of $\mathcal{L}_{\mathrm{ae}}$. For every set of sentences $\Delta \subseteq \mathcal{L}_{\mathrm{ae}}$ such that $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$, there exists exactly one Σ-full set Λ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$.*

Proof. Lemma 3.5 implies that for every $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$, there is a Σ-full set Λ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Assume there is another Σ-full set Λ' such that $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$ and $\Lambda' \neq \Lambda$ holds. Thus there exists $\chi \in \Lambda' - \Lambda$ (or $\chi \in \Lambda - \Lambda'$). As $\chi \in \Lambda'$, q-atom$(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. As $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$, $\chi \in L\Delta \cup \neg L\overline{\Delta}$ and by Proposition 3.6 $\chi \in \Lambda$ which is a contradiction. Thus $\Lambda' = \Lambda$. $\square$

Theorem 3.10 *Let Σ be a set of sentences of $\mathcal{L}_{\mathrm{ae}}$. For every Σ-full set Λ, there is exactly one set of sentences Δ such that $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$ and $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$.*

Proof. Lemma 3.7 implies that there exists a set $\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Assume there is another set $\Delta' = \mathrm{Cn}(\Sigma \cup L\Delta' \cup \neg L\overline{\Delta'})$ such that $\Lambda \subseteq L\Delta' \cup \neg L\overline{\Delta'}$. We prove that for every $\psi \in \mathcal{L}_{\mathrm{ae}}$, $\psi \in \Delta$ if and only if $\psi \in \Delta'$ by induction on $\mathrm{BD}(\psi)$. This shows that the two stable expansions are equal.

Let $\mathrm{BD}(\psi) = 0$. Let $\psi \in \Delta$. So $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \psi$. By Proposition 2.17 there is a set $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \psi$ and for every $\chi \in \Lambda'$, q-atom$(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\psi\})$. As the $\mathrm{BD}(\psi) = 0$, q-atom$(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. Then by Proposition 3.6 as $\chi \in L\Delta \cup \neg L\overline{\Delta}$, $\chi \in \Lambda$. So $\chi \in L\Delta' \cup \neg L\overline{\Delta'}$. Hence $\Lambda' \subseteq L\Delta' \cup \neg L\overline{\Delta'}$ which implies $\Sigma \cup L\Delta' \cup \neg L\overline{\Delta'} \models_{\mathrm{ae}} \psi$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus $\psi \in \Delta'$. Similarly, $\psi \in \Delta'$ implies $\psi \in \Delta$.

Assume that for every χ with $\mathrm{BD}(\chi) < i$, $\chi \in \Delta$ if and only if $\chi \in \Delta'$. Let $\mathrm{BD}(\psi) = i$. Let $\psi \in \Delta$. So $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \psi$. By Proposition 2.17

there exists a set $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \psi$ and for every $\chi \in \Lambda'$, q-atom$(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\psi\})$. If q-atom$(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$, then by Proposition 3.6 as $\chi \in L\Delta \cup \neg L\overline{\Delta}$, then also $\chi \in \Lambda$ holds. So $\chi \in L\Delta' \cup \neg L\overline{\Delta'}$. Otherwise, q-atom$(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\psi)$. The formula q-atom$(\chi)$ is of the form $L\chi'$ and $\mathrm{BD}(\chi') < i$. If $\chi \in L\Delta$, then $\chi = L\chi'$ and $\chi' \in \Delta$. By the inductive hypothesis $\chi' \in \Delta'$ and thus $\chi = L\chi' \in L\Delta'$. In the other case $\chi \in \neg L\overline{\Delta}$. Then $\chi = \neg L\chi'$ and $\chi' \notin \Delta$. By the inductive hypothesis $\chi' \notin \Delta'$ and thus $\chi = \neg L\chi' \in \neg L\overline{\Delta'}$. So $\Lambda' \subseteq L\Delta' \cup \neg L\overline{\Delta'}$ and by the monotonicity of $\models_{\mathrm{ae}}$ $\Sigma \cup L\Delta' \cup \neg L\overline{\Delta'} \models_{\mathrm{ae}} \psi$. Hence $\psi \in \Delta'$. Similarly, $\psi \in \Delta'$ implies $\psi \in \Delta$. □

Given a Σ-full set Λ, the set $\{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \Lambda \models_{\mathrm{L}} \phi\}$ is a stable expansion of Σ by Lemma 3.7 and by Theorem 3.10 it is the unique stable expansion containing Λ. Thus a sentence ϕ belongs to the expansion induced by Λ if and only if $\Sigma \cup \Lambda \models_{\mathrm{L}} \phi$.

Corollary 3.11 *Let Σ be a set of sentences of $\mathcal{L}_{\mathrm{ae}}$ and Λ a Σ-full set. Then*

$$\mathrm{SE}_{\Sigma}(\Lambda) = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \Lambda \models_{\mathrm{L}} \phi\}$$

is the unique stable expansion Δ of Σ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$.

Full sets also provide a technique for determining the number of stable expansions of a set of premises.

Theorem 3.12 *For a set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$, there is a bijective mapping from the set of Σ-full sets to the set of stable expansions of Σ.*

Proof. Let F be the set of Σ-full sets and SE the set of stable expansions of Σ. Let $f : F \mapsto SE$ be a mapping defined as follows: for every $\Lambda \in F$,

$$f(\Lambda) = \mathrm{SE}_{\Sigma}(\Lambda) = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \Lambda \models_{\mathrm{L}} \phi\}. \tag{3.3}$$

By Lemma 3.7 for every $\Lambda \in F$, $f(\Lambda) \in SE$. So f is in fact a mapping. The mapping f is injective as if $f(\Lambda) = f(\Lambda')$ for any $\Lambda, \Lambda' \in F$, then by Theorem 3.9 $\Lambda = \Lambda'$ as for every $\Lambda \in F, \Lambda \subseteq Lf(\Lambda) \cup \neg L\overline{f(\Lambda)}$ by Lemma 3.7. The mapping f is also surjective and thus bijective as for every $\Delta \in SE$, there is a set $\Lambda \in F$ such that $\Delta = f(\Lambda)$. By Lemma 3.5 for each $\Delta \in SE$, there is a set $\Lambda \in F$ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$ and by Theorem 3.10 Δ is the only stable expansion such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Thus $\Delta = f(\Lambda)$. □

Corollary 3.13 *A set of sentences of $\mathcal{L}_{\mathrm{ae}}$ has exactly as many stable expansions as it has full sets.*

Example 3.3 Let the underlying logic CL be the propositional calculus. Let $\Sigma = \{Lp \rightarrow p, \neg Lp \rightarrow q\}$, where p and q are atomic. There are two possible candidates for Σ-full sets: $\Lambda_1 = \{Lp\}$ and $\Lambda_2 = \{\neg Lp\}$. Both are Σ-full. For example, Λ_1 is full as $\Sigma \cup \Lambda_1 \models_{\mathrm{L}} p$. So Σ has exactly two stable expansions: $\mathrm{SE}_\Sigma(\{Lp\})$ and $\mathrm{SE}_\Sigma(\{\neg Lp\})$. ■

As a corollary of the previous theorems we can offer a (partial) answer to the question about a criterion for the sets of premises with exactly one stable expansion. Thus the full set approach provides a practical technique for examining sets of premises and proving that they have exactly one stable expansion.

Corollary 3.14 *A set of sentences of $\mathcal{L}_{\mathrm{ae}}$ has exactly one stable expansion if and only if it has exactly one full set.*

As a straightforward consequence it follows that sets of premises containing only sentences without the L operator have exactly one stable expansion. This is because such sets of premises have only one full set, the empty set. For the first-order case this result has already been established by Konolige [61].

Corollary 3.15 *A set of sentences $\Sigma \subseteq \mathcal{L}$ has exactly one stable expansion* $\mathrm{SE}_\Sigma(\emptyset)$.

Note that in the propositional case $\mathrm{SE}_\Sigma(\emptyset) = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \models_{\mathrm{L}} \phi\}$ gives a compact characterization of Marek's operator E [79] which produces the unique stable expansion of an L free set of premises.

Theorem 3.16 *Let the underlying logic be a first-order predicate calculus. Let Σ be a set of sentences of $\mathcal{L}$. Then $\mathrm{SE}_\Sigma(\emptyset)$ is the unique stable set S such that $\mathrm{Cn}_{\mathrm{cl}}(\Sigma) = S \cap \mathcal{L}$.*

Proof. As $\mathrm{SE}_\Sigma(\emptyset)$ is a stable expansion, it is a stable set and clearly $\mathrm{Cn}_{\mathrm{cl}}(\Sigma) = \mathrm{SE}_\Sigma(\emptyset) \cap \mathcal{L}$. Assume that there is another stable set S such that $\mathrm{Cn}_{\mathrm{cl}}(\Sigma) = S \cap \mathcal{L}$. We show that for all $\phi \in \mathcal{L}_{\mathrm{ae}}$, $\phi \in \mathrm{SE}_\Sigma(\emptyset)$ if and only if $\phi \in S$ by induction on $\mathrm{BD}(\phi)$. If $\mathrm{BD}(\phi) = 0$, then the assertion clearly holds. Assume that for all $\phi \in \mathcal{L}_{\mathrm{ae}}$ such that $\mathrm{BD}(\phi) < i$, $\phi \in \mathrm{SE}_\Sigma(\emptyset)$ if and only if $\phi \in S$. Let $\mathrm{BD}(\phi) = i$. If $\phi \in \mathrm{SE}_\Sigma(\emptyset)$, then $\Sigma \models_{\mathrm{L}} \phi$ and thus $\Sigma \cup \mathrm{SB}_\Sigma(\phi) \models_{\mathrm{ae}} \phi$. By the inductive hypothesis $\mathrm{SB}_\Sigma(\phi) \subseteq S$ and as $\Sigma \subseteq S$, $\phi \in S$. Also from the inductive hypothesis follows that $\phi \in S$ implies $\phi \in \mathrm{SE}_\Sigma(\emptyset)$ by Proposition 2.19. Thus $S = \mathrm{SE}_\Sigma(\emptyset)$. □

Because of the fixed point definition an upper bound for the number of stable expansions is difficult to determine. However, using the finitary characterization a tight upper bound can be given. The author [106, 107] has previously shown that there is at most 2^n stable expansions of a set of premises where n is the cardinality of the set of $L\phi$ subformulae of the premises. Here we show that this result can be improved with the aid the new finitary characterization. If the set $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ is finite, there is only a finite number of candidates for Σ-full sets and thus there can be only finitely many stable expansions of Σ. A Σ-full set contains either $L\chi$ or $\neg L\chi$ for each $L\chi$ quasi-subformula of Σ. Thus there can be only 2^n stable expansions of a set of premises Σ for which the cardinality of $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ is n.

Corollary 3.17 *A set of sentences Σ such that the cardinality of* $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ *is n has at most 2^n stable expansions.*

Example 3.4 We illustrate the upper bound on the number of stable expansions by an example. Let the underlying logic CL be the propositional calculus. Let $\Sigma = \{Lp_1 \to p_1, Lp_2 \to p_2, \ldots, Lp_n \to p_n\}$ where $p_1, \ldots, p_n$ are atomic. Then $\mathrm{Sf}^{\mathrm{qL}}(\Sigma) = \{Lp_1, \ldots, Lp_n\}$ and its cardinality is n. The set Σ has 2^n stable expansions as it has 2^n full sets:

$$\{\neg Lp_1, \neg Lp_2, \ldots, \neg Lp_n\}, \{Lp_1, \neg Lp_2, \ldots, \neg Lp_n\}, \ldots, \{Lp_1, Lp_2, \ldots, Lp_n\}.$$

■

A full set can also be captured in terms of the underlying consequence relation $\models_{\mathrm{ae}}$ but then all $L\phi$ subformulae of the premises have to be considered.

Proposition 3.18 *Let Σ be set of sentences of $\mathcal{L}_{\mathrm{ae}}$. If $\Upsilon \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and for every $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$,*

1. *$\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ if and only if $L\chi \in \Upsilon$ and*
2. *$\Sigma \cup \Upsilon \not\models_{\mathrm{ae}} \chi$ if and only if $\neg L\chi \in \Upsilon$*

then $\Lambda = \{\phi \in \Upsilon \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$ is a Σ-full set and for all $\chi \in \mathcal{L}_{\mathrm{ae}}$,

$$\Sigma \cup \Lambda \models_{\mathrm{L}} \chi \textit{ if and only if } \Sigma \cup \Upsilon \models_{\mathrm{L}} \chi.$$

Hence $\mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(\Upsilon)$.

Proof. We prove first that for all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$,

$$\Sigma \cup \Lambda \models_{\mathrm{L}} \chi \text{ if and only if } \Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi \tag{3.4}$$

by induction on $\mathrm{BD}(\chi)$. This implies directly that Λ is a Σ-full set.

Let $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. If $\mathrm{BD}(\chi) = 0$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ if and only if $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ and as $\Lambda \subseteq \Upsilon$, $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. Let $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ hold. By Proposition 2.17 there is a set $\Lambda' \subseteq \Upsilon$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and for every $\chi' \in \Lambda'$, q-atom$(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\chi\}) = \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. As $\chi' \in \Upsilon$, then $\chi' \in \Lambda$ and $\Lambda' \subseteq \Lambda$. Thus $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$.

Assume that for all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ with $\mathrm{BD}(\chi) < i$,

$$\Sigma \cup \Lambda \models_{\mathrm{L}} \chi \text{ if and only if } \Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi.$$

Let $\mathrm{BD}(\chi) = i$. Let $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$. Then $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\chi) \models_{\mathrm{ae}} \chi$. If $L\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi'$ and $L\chi' \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$. As $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, then $L\chi' \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. Because $\mathrm{BD}(\chi') < i$, $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi'$ by the inductive hypothesis. Thus by Condition 1 $L\chi' \in \Upsilon$. Similarly, by Condition 2 $\neg L\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi')$ implies that $\neg L\chi' \in \Upsilon$. Thus $\mathrm{SB}_{\Sigma\cup\Lambda}(\chi) \subseteq \Upsilon$ and $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. On the other hand, let $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ hold. By Proposition 2.17 there is a set $\Lambda' \subseteq \Upsilon$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and for every $\chi' \in \Lambda'$, q-atom$(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\chi\})$. If q-atom$(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$, then $\chi' \in \Lambda$ because $\chi' \in \Upsilon$. If q-atom$(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$, then as $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, q-atom$(\chi') \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. This implies that $\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$ in the following way. If $\chi' = L\chi''$, then by Condition 1 $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi''$ and as $\mathrm{BD}(\chi'') < i$, by the inductive hypothesis $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi''$. Thus $\chi' = L\chi'' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$. Similarly, if $\chi' = \neg L\chi''$, then by Condition 2 and the inductive hypothesis $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi''$ and thus $\chi' = \neg L\chi'' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$. Hence $\Lambda' \subseteq \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$ and $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$.

Then we prove that for all $\chi \in \mathcal{L}_{\mathrm{ae}}$,

$$\Sigma \cup \Lambda \models_{\mathrm{L}} \chi \text{ if and only if } \Sigma \cup \Upsilon \models_{\mathrm{L}} \chi \tag{3.5}$$

by induction on $\mathrm{BD}(\chi)$.

If $\mathrm{BD}(\chi) = 0$, then $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ if and only if $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ and as $\Lambda \subseteq \Upsilon$, $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus $\Sigma \cup \Upsilon \models_{\mathrm{L}} \chi$ holds. Let $\Sigma \cup \Upsilon \models_{\mathrm{L}} \chi$ hold. Then $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ holds and by Proposition 2.17 there is a set $\Lambda' \subseteq \Upsilon$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and for every $\chi' \in \Lambda'$, q-atom$(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\chi\}) = \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$. As $\chi' \in \Upsilon$, then $\chi' \in \Lambda$ and $\Lambda' \subseteq \Lambda$. Thus $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$.

Assume that for all $\chi \in \mathcal{L}_{\mathrm{ae}}$ with $\mathrm{BD}(\chi) < i$,

$$\Sigma \cup \Lambda \models_{\mathrm{L}} \chi \text{ if and only if } \Sigma \cup \Upsilon \models_{\mathrm{L}} \chi.$$

Let $\mathrm{BD}(\chi) = i$. By the inductive hypothesis $\mathrm{SB}_{\Sigma\cup\Lambda}(\chi) = \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$. Let $\Sigma\cup\Lambda \models_{\mathrm{L}} \chi$. Then $\Sigma\cup\Lambda\cup\mathrm{SB}_{\Sigma\cup\Lambda}(\chi) \models_{\mathrm{ae}} \chi$. As $\Lambda\cup\mathrm{SB}_{\Sigma\cup\Lambda}(\chi) \subseteq \Upsilon\cup\mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$, by the monotonicity of $\models_{\mathrm{ae}}$ $\Sigma \cup \Upsilon \models_{\mathrm{L}} \chi$. On the other hand, let $\Sigma \cup \Upsilon \models_{\mathrm{L}} \chi$ hold. Then $\Sigma \cup \Upsilon \cup \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi) \models_{\mathrm{ae}} \chi$. Assume that $\Upsilon \cup \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ contains a formula and its negation. As neither Υ nor $\mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ can contain a formula and its negation, it must be the case that for some $L\chi'$, either $L\chi' \in \Upsilon$ and

$\neg L\chi' \in \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ or $\neg L\chi' \in \Upsilon$ and $L\chi' \in \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$. If $L\chi' \in \Upsilon$, then $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi'$. By (3.4) $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi'$ holds as $L\chi' \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. If $L\chi' \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$, then $\mathrm{BD}(\chi') < i$ and by the inductive hypothesis $\Sigma \cup \Upsilon \models_{\mathrm{L}} \chi'$. Hence $\neg L\chi' \notin \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$. So it must be the case that $\neg L\chi' \in \Upsilon$ and $L\chi' \in \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$. Then $\Sigma \cup \Upsilon \not\models_{\mathrm{ae}} \chi'$. By (3.4) $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi'$ holds as $L\chi' \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. If $L\chi' \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$, then $\mathrm{BD}(\chi') < i$ and by the inductive hypothesis $\Sigma \cup \Upsilon \not\models_{\mathrm{L}} \chi'$. Hence $L\chi' \notin \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$, a contradiction. Thus $\Upsilon \cup \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ contains no formula and its negation. Then by Proposition 2.17 there is a set $\Lambda' \subseteq \Upsilon \cup \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and for every $\chi' \in \Lambda'$, $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\chi\})$. We show that $\Lambda' \subseteq \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$. Let $\chi' \in \Lambda'$. If $\chi' \in \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$, then $\chi' \in \mathrm{SB}_{\Sigma\cup\Lambda}(\chi) = \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$. If $\chi' \notin \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$, then $\chi' \in \Upsilon$. If $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$, then $\chi' \in \Lambda$. Otherwise, $\text{q-atom}(\chi') \in \mathrm{Sf}^{\mathrm{qL}}(\chi)$. Then $\mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ contains $\text{q-atom}(\chi')$ or its negation. As $\Upsilon \cup \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi)$ contains no formula and its negation, $\chi' \in \mathrm{SB}_{\Sigma\cup\Upsilon}(\chi) = \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$. Hence $\Lambda' \subseteq \Lambda \cup \mathrm{SB}_{\Sigma\cup\Lambda}(\chi)$ and $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. □

The finitary characterization of stable expansions in terms of full sets presented in this chapter is a modification of an earlier characterization developed by the author [105, 106, 107]. Proposition 3.18 reveals the close relationship between full sets presented in this chapter and the earlier characterization. The basic idea of the earlier characterization is that a stable expansion of a set of premises can be characterized using the $L\phi$ subformulae of the premises. A decision procedure for autoepistemic logic [103] is based on this idea. Fixed points of the nonmonotonic consequence operator of McDermott and Doyle's nonmonotonic logic can be characterized using this idea [104]. The earlier characterizing sets [105, 106, 107], which were also called full sets, can be defined as follows. For a set of premises Σ, a set $\Upsilon \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ such that for every $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ the two conditions given in Proposition 3.18

1. $\Sigma \cup \Upsilon \models_{\mathrm{ae}} \chi$ if and only if $L\chi \in \Upsilon$ and
2. $\Sigma \cup \Upsilon \not\models_{\mathrm{ae}} \chi$ if and only if $\neg L\chi \in \Upsilon$

are satisfied serves as a kernel or a characterizing set for a Moore style stable expansion of Σ. By comparing the conditions in Proposition 3.18 and Definition 3.1 one observes that the earlier characterizing sets differ from the full sets presented in this chapter in two respects.

1. The earlier characterization is constructed from the $L\phi$ subformulae of the premises instead of the $L\phi$ quasi-subformulae.
2. The underlying consequence relation $\models_{\mathrm{ae}}$ is used instead of $\models_{\mathrm{L}}$.

The results obtained in this chapter for the general case can be established using the earlier characterization with the exception of the following point. The full sets presented in this work reveal that $L\phi$ subformulae of the premises which are in the scope of an L operator do not play a leading role when characterizing stable expansions. The full sets presented in this chapter provide a more compact characterization of stable expansions where only $L\phi$ quasi-subformulae of the premises are used and $L\phi$ subformulae in the scope of another L operator are disregarded. This implies a tighter upper bound for the number of stable expansions. Using the earlier characterization, it can be proved that a set of premises has at most 2^n stable expansions where n is the number of $L\phi$ subformulae of the premises [106, 107]. Using the new full set characterization, the upper bound is $2^{n'}$ where n' is the number of $L\phi$ quasi-subformulae of the premises. A tighter upper bound is obtained because the number of $L\phi$ quasi-subformulae of the premises can be strictly lower than that of $L\phi$ subformulae if the premises contain nested L operators.

Other researchers have independently exploited the same idea of using the $L\phi$ subformulae of the premises for characterizing stable expansions. Shvarts [135] studies McDermott and Doyle style nonmonotonic modal logics and develops finitary characterizations for expansions in several nonmonotonic modal logics using this idea. As a result of this study a finitary characterization for stable expansions is obtained, too. Shvarts shows that propositional autoepistemic logic can be seen as a McDermott and Doyle style logic based on the modal logic **K45** where consistent stable expansions correspond to **K45**-expansions. He provides a finitary characterization for **K45**-expansions and thus for stable expansions which is only slightly different from the earlier full set approach [105, 106, 107]. In his approach a stable expansion is characterized by a subset Φ of the $L\phi$ subformulae of the premises Σ which is required to be *propositional admissible* for Σ, i.e. to satisfy the following three conditions.

1. $\Sigma \cup \neg\Phi \cup (\mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi) \cup \{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi\}$ is propositionally consistent.
2. For all $L\chi \in \Phi$, $\Sigma \cup \neg\Phi \cup (\mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi) \cup \{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi\} \not\models_{\mathrm{ae}} \chi$.
3. For all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi$, $\Sigma \cup \neg\Phi \cup (\mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi) \models_{\mathrm{ae}} \chi$.

Clearly this is very close to the earlier definition of full sets where $\neg\Phi \cup (\mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Phi)$ is the full set and Conditions 2 and 3 correspond to the two conditions in Proposition 3.18. Condition 1 has no counterpart in the full set approach and because of this difference the characterization of Shvarts captures only the consistent stable expansions. This is not a serious shortcoming because inconsistent stable expansions could be given a separate treatment.

Shvarts [135] devises finitary characterizations for expansions in other nonmonotonic modal logic by suitably modifying Condition 3. Shvarts extends his work together with Marek and Truszczyński [80]. They develop finitary characterizations for further nonmonotonic modal logics and show how the finitary characterizations can be used as a basis for decision procedures. The difference between the full set approach and that of Shvarts can be explained by the fact that Shvarts aims to provide finitary characterizations also for expansions in McDermott and Doyle style logics based on modal logics containing the axiom schema **T**. In these logics expansions are not necessarily Moore style stable expansions and thus do not always have a corresponding full set. Shvarts captures the different structure of the expansions caused by the presense of the axiom schema **T** by using the new consistency condition 1 and a slightly modified condition 2 for the $\neg L\phi$ formulae in the characterizing set. We discuss this point further in Section 7.3.

Stärk [137] has proposed a characterization of stable expansions using $L\phi$ subformulae of the premises and admissible labelings similar to those used in a decision procedure for propositional autoepistemic logic [103]. Stärk's characterization is identical to the earlier full set approach [105, 106, 107]. Stärk captures a stable expansion as a labeling which is a mapping λ from $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ to $\{0, 1\}$. A labeling λ correspond to a full set in the following way: $\lambda(\phi) = 1$ means that $L\phi$ belongs to the full set and $\lambda(\phi) = 0$ means that $\neg L\phi$ belongs to the full set. The admissibility condition for the labeling corresponds exactly to the two conditions in Proposition 3.18.

Moore [101] has proposed a characterization of stable expansions in first-order autoepistemic logic without quantifying-in. The characterization is based on truth assignments to $L\phi$ subformulae of the premises and a transformation procedure of the premises. The intuition of the approach seems to be very close to the earlier full set approach [105, 106, 107]. The main difference is that Moore's approach is more semantically oriented being based on truth assignments whereas the other approach is a syntactic one where no transformation of premises is required.

Chapter 4

Groundedness of Moore Style Stable Expansions

A Moore style stable expansion is defined by stating a condition (the fixed point equation) which refers to the whole expansion. This leaves open the possibility that a stable expansion could contain beliefs which are circularly justified. The soundness or groundedness condition for stable expansions tries to capture the requirement that the beliefs of the agent are based on the given set of premises. However, it can be argued that the condition proposed by Moore is rather weak. In this chapter we explain the problem and survey critically the proposed solutions.

Example 4.1 Consider the following premise which illustrates the problem. The formula p is atomic.

$$\Sigma = \{Lp \rightarrow p\} \tag{4.1}$$

The premise has two stable expansions whose corresponding full sets are $\Lambda_1 = \{\neg Lp\}$ and $\Lambda_2 = \{Lp\}$. The stable expansion $\mathrm{SE}_\Sigma(\{Lp\})$ contains the belief p on the basis of the following argument. The agent first adopts the belief p, thus obtains Lp by positive introspection and from this and the premise deduces p. This kind of a belief based on a circular argument is often considered to be too weakly grounded and a stronger form of groundedness in the premises is required. ■

Konolige [61] presents two stronger notions of groundedness leading to *moderately grounded* and *strongly grounded* expansions. His basic motivation is to establish a correspondence between reasoning in Reiter's [122] default logic and autoepistemic reasoning. He considers a particular translation between default theories and sets of autoepistemic sentences (see Section 7.1) and attempts to identify a class of stable expansions corresponding to extensions in default logic under this translation. Konolige observes that the fixed point equation of a stable expansion (2.6) can be rewritten in the form

$$\Delta = \{\phi \mid \Sigma \cup L\Delta_0 \cup \neg L\overline{\Delta}_0 \models_{\mathrm{SS}} \phi\} \tag{4.2}$$

where $\Delta_0 = \Delta \cap \mathcal{L}$, $\overline{\Delta}_0 = \overline{\Delta} \cap \mathcal{L} = \mathcal{L} - \Delta$ and $\models_{\mathrm{SS}}$ is a consequence relation defined similar to $\models_{\mathrm{ae}}$ with respect to autoepistemic structures $\langle I, T\rangle$ except that the sets of sentences T are restricted to vary over stable sets instead of all sets of sentences. Thus using a stronger type of the consequence relation ($\models_{\mathrm{SS}}$), it is possible to reduce the self-referential formulae in the equation to those with the L operator applied to formulae in $\mathcal{L}$. To eliminate circularly justified beliefs, Konolige introduces the concept of moderately grounded sets.

Definition 4.1 (Konolige [61]) *A set of sentences $\Delta \subseteq \mathcal{L}_{\mathrm{ae}}$ is* moderately grounded *in Σ if it obeys the fixed point equation*

$$\Delta = \{\phi \mid \Sigma \cup L\Sigma \cup \neg L\overline{\Delta}_0 \models_{\mathrm{SS}} \phi\}. \tag{4.3}$$

Konolige shows that a set of formulae Δ moderately grounded in Σ is a stable expansion of Σ which is minimal in the following sense: there is no stable set S which contains Σ such that $S \cap \mathcal{L} \subset \Delta \cap \mathcal{L}$, i.e. the L free formulae in S are a proper subset of the L free formulae in Δ. Such expansions are said to be *stable set minimal* for Σ.

Moderately grounded expansions do not quite capture extensions in default logic and, moreover, are still rather weakly grounded as can be seen from the following example presented by Konolige [61].

Example 4.2 Consider the following set of premises Σ where p, q are atomic formulae.

$$\Sigma = \{Lp \rightarrow p, \neg Lp \rightarrow q\} \tag{4.4}$$

The set Σ has two stable expansions $\mathrm{SE}_\Sigma(\{Lp\})$ and $\mathrm{SE}_\Sigma(\{\neg Lp\})$ which both are stable set minimal but the first one contains the belief p which is based on the belief p, i.e. on the same circular argument as in Example 4.1. ■

To capture extensions in default logic Konolige introduces the class of strongly grounded expansions. The class is defined only for autoepistemic sentences in a normal form:

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \tag{4.5}$$

where α, β_i, and ω are all in $\mathcal{L}$ and any of the disjuncts except for ω may be absent. Konolige [61] shows that any set of autoepistemic sentences has a **K45**-equivalent set of sentences each of which is in the normal form (4.5).

Definition 4.2 (Konolige [61, 62]) *Let* Σ *be a set of sentences in the normal form and let* Δ *be a stable expansion of* Σ*. Let* $\Sigma' = \{\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \in \Sigma \mid \beta_1 \notin \Delta, \ldots, \beta_n \notin \Delta\}$*. Then* Δ *is* strongly grounded *in* Σ *if and only if*

$$\Delta = \{\phi \mid \Sigma' \cup L\Sigma' \cup \neg L\overline{\Delta}_0 \models_{\mathrm{SS}} \phi\}. \tag{4.6}$$

Konolige shows that strongly grounded expansions are a strict subclass of moderately grounded expansions. Furthermore, strongly grounded expansions correspond to extensions in default logic under the Konolige translation but there are some unnatural aspects in this notion of groundedness as discussed by Konolige [61] and Marek and Truszczyński [86]. Different expansions can be grounded in different subsets of the premises although is seems natural to expect that an expansion is grounded in the whole set of premises. For example, consider the set of premises

$$\Sigma = \{\neg Lp \rightarrow q, \neg Lq \rightarrow p\} \tag{4.7}$$

which has two expansions $\mathrm{SE}_\Sigma(\{Lp, \neg Lq\})$ and $\mathrm{SE}_\Sigma(\{\neg Lp, Lq\})$. The first is strongly grounded in $\Sigma' = \{\neg Lq \rightarrow p\}$ and the second in $\Sigma' = \{\neg Lp \rightarrow q\}$.

Konolige [61] points out that strong groundedness is a syntactic property of the premises. This means that classically equivalent sets of premises may have different strongly grounded expansions. For example, $(Lp \rightarrow p) \wedge (\neg Lp \rightarrow p)$ is equivalent to p in propositional logic but the former has none and the latter has one strongly grounded expansion. Hence, strongly grounded expansions depend on the syntactic representation of the premises.

Marek and Truszczyński [84, 86] present an interesting notion of groundedness in their study of the relationship between default logic and autoepistemic logic. It is based on a monotone operator A defined as follows.

$$A(\Sigma) = \mathrm{Cn}(\Sigma \cup L\Sigma) \tag{4.8}$$

The operator A is equipped with a context parameter $\Delta \subseteq \mathcal{L}_{\mathrm{ae}}$ and given a set of premise $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ the operator A^Δ is iterated in the following way.

$$\begin{aligned}
A_0^\Delta(\Sigma) &= \mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta}) && (4.9)\\
A_{n+1}^\Delta(\Sigma) &= A(A_n^\Delta(\Sigma)) = \mathrm{Cn}(A_n^\Delta(\Sigma) \cup LA_n^\Delta(\Sigma)) && (4.10)\\
A^\Delta(\Sigma) &= \bigcup_{n=0}^{\infty} A_n^\Delta(\Sigma) && (4.11)
\end{aligned}$$

Definition 4.3 (Marek and Truszczyński [84]) *A set of sentences* Δ *is an* iterative expansion *of* Σ *if and only if*

$$\Delta = A^\Delta(\Sigma). \tag{4.12}$$

The class of iterative expansions is an attractive candidate for the sets of beliefs adopted by a fully introspective agent. In particular, because in a later paper Marek and Truszczyński [86] show that Reiter's default logic can be seen as a special case of autoepistemic logic based on iterative expansions under a suitable translation of default theories into sets of autoepistemic sentences (see Section 7.1). However, iterative expansions may be regarded too weakly grounded because they are not necessarily stable set minimal. The following example due to Marek and Truszczyński [86] illustrates this.

Example 4.3 Consider a premise

$$\Sigma = \{\neg L \neg Lp \rightarrow p\} \tag{4.13}$$

where p is atomic. The set Σ has two stable expansions $\mathrm{SE}_\Sigma(\{\neg L \neg Lp\})$ and $\mathrm{SE}_\Sigma(\{L \neg Lp\})$ which are both iterative but only the latter is stable set minimal. Technically the problem seems to be that the definition of $A^\Delta(\Sigma)$ allows the adoption of a belief in the stable expansion (in the context Δ) which is justified (indirectly) by the same belief. This is what happens in the former expansion. If p is put to the context Δ but $\neg Lp$ is not, then by negative introspection we have that $\neg L \neg Lp \in A_0^\Delta(\Sigma)$ and $p \in A_0^\Delta(\Sigma)$. Thus $Lp \in A_1^\Delta(\Sigma)$. Hence $p \in A^\Delta(\Sigma)$ and $Lp \in A^\Delta(\Sigma)$. It cannot be the case that $\neg Lp \in A^\Delta(\Sigma)$ (unless $A^\Delta(\Sigma)$ is inconsistent, which it is not). ■

Marek and Truszczyński [84] show how to strengthen the notion of iterative expansions to capture extensions in default logic under the Konolige translation. This notion is defined only for sets of premises in the normal form (4.5).

Definition 4.4 *(Marek and Truszczyński [84]) Let Σ be a set of sentences in the normal form. Let Δ be a stable expansion of Σ. Then Δ is a* strong iterative expansion *of Σ if and only if*

$$\Delta = A^\Delta(GD(\Sigma, \Delta)) \tag{4.14}$$

where $GD(\Sigma, \Delta) = \{\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \in \Sigma \mid \alpha \in \Delta \text{ and } \beta_1 \notin \Delta, \ldots, \beta_n \notin \Delta\}$.

Marek and Truszczyński [84] define yet another class of expansions to capture extensions in default logic under the Konolige translation. These expansions are called *robust* and they are also defined only for premises in the normal form.

Konolige's strongly grounded expansions as well as Marek and Truszczyński's strong iterative and robust expansions are all defined only for premises in the normal form. In particular, no nested belief operators are allowed in the

$$\begin{tikzcd} \boxed{\begin{array}{l}\text{Strongly grounded}\\ \text{Strong iterative}\\ \text{Robust}\end{array}} \arrow[r] \arrow[d] & \boxed{\text{Iterative}} \arrow[d] \\ \boxed{\begin{array}{l}\text{Moderately grounded}\\ \text{Stable set minimal}\end{array}} \arrow[r] & \boxed{\text{Stable expansion}} \end{tikzcd}$$

Figure 4.1: Relationships between groundedness concepts.

premises. The three classes of expansions coincide. Given a set of autoepistemic premises in the normal form, there is a corresponding default theory under the Konolige translation. For an expansion belonging to any of the three classes, its L free part is an extension of the corresponding default theory. This implies that there is an expansion belonging to each of the three classes such that the extension of the corresponding default theory is the L free part of the expansion. This means that all these expansions have the same L free part which implies that they coincide. Expansions in these three classes share the same unnatural properties as discussed in connection with Konolige's strongly grounded expansions. In particular, they depend on the syntactic representation of the premises.

Figure 4.1 displays a summary of the various groundedness conditions and their relationships. The arrow is interpreted as implies. For example, strong groundedness implies stable set minimality.

It seems clear that in order to obtain more tightly grounded expansions the construction of an expansion should start from the formulae without the L operator. Marek [79] presents such a construction (the operator E) which builds a stable set out of its L free part. The construction can be described compactly using the consequence relation $\models_{\mathrm{L}}$ as discussed in Chapter 3. However, the method does not seem to be applicable to building expansions from premises containing L operators because expansions cannot be constructed using such a layered method where, roughly speaking, all formulae having the same depth of L operators are treated at the same time. Consider, e.g., the set of premises $\{\neg Lq \rightarrow q, \neg LLp \rightarrow q\}$.

Example 4.1 suggests that the weak groundedness of Moore style expansions is caused by the principle of direct positive introspection employed in the fixed point equation (2.6). McDermott and Doyle's [96, 95] scheme of defining nonmonotonic modal logics offers an alternative approach where direct positive introspection is replaced by appropriate modal axioms and inference rules. The scheme can be used to defining a nonmonotonic modal logic based on a given monotonic modal logic S. The central notion is that of an S-*expansion* which is defined by the fixed point equation (4.15) where $\vdash_S$ is the derivability relation of the underlying modal logic S.

$$\Delta = \{\phi \mid \Sigma \cup \neg L\overline{\Delta} \vdash_S \phi\} \tag{4.15}$$

The S-expansions are closely related to autoepistemic expansions (see Section 7.3) and some weakly grounded expansions can be eliminated by choosing a suitable underlying modal logic S. However, circularly justified beliefs are not excluded completely. Marek and Truszczyński [84] note that $A^{\Delta}(\Sigma)$ is the set of all formulae provable from $\Sigma \cup \neg L\overline{\Delta}$ using classical propositional logic and the necessitation rule

$$\text{from } \phi \text{ infer } L\phi. \tag{4.16}$$

Hence, an iterative expansion can be defined as an **N**-expansion where the underlying modal logic **N** is an extension of the propositional calculus with no modal axioms but the rule of necessitation. Example 4.3 illustrates the problem of circularly justified beliefs in **N**-expansions (iterative expansions). This problem of **N**-expansions is immediately inherited by other McDermott and Doyle style nonmonotonic modal logics because of the following property of nonmonotonic modal logics: for any modal logics S_1, S_2 contained in **S5**, if S_1 is contained in S_2, then any S_1-expansion is a S_2-expansion [80]. (For example, Chellas [14] and Hughes and Cresswell [46] provide thorough introductions to modal logics.)

In Example 4.3 the weakly grounded expansion seems to hinge on the fact that $\neg L\neg Lp$ is obtained by direct negative introspection. A straightforward technique to eliminate such a possibility is to restrict the set of formulae subject to direct introspection. Tiomkin and Kaminski [140, 141] propose *nonmonotonic ground logics* where negative introspection is applied only to L free formulae. The resulting sets of beliefs, *ground S-expansions*, are defined using the following fixed point equation

$$\Delta = \{\phi \mid \Sigma \cup \neg L\overline{\Delta}_0 \vdash_S \phi\} \tag{4.17}$$

where $\vdash_S$ is the derivability relation of the underlying modal logic S. Kaminski [56], Truszczyński [146], Konolige [65], and Schwarz [128] have continued this approach. Tiomkin and Kaminski [140, 141] consider nonmonotonic ground logics based on modal logics $\mathbf{T}, \mathbf{S4}$, and **S5**. For these modal logics, S-expansions are not necessarily stable expansions. For example, the premise $\neg Lp \rightarrow p$ has no stable expansions. But all the underlying logics $\mathbf{T}, \mathbf{S4}$, and **S5** contain the axiom schema **T**

$$L\phi \rightarrow \phi \tag{4.18}$$

which produces an S-expansion containing p. To remain within stable expansions, the axiom schema **T** has to be abandoned. Konolige's [61] moderately grounded expansions are an example of these kinds of ground S-expansions. They are based on the modal logic **K45**. The ground **K45**-expansions are rather weakly grounded as discussed in Example 4.2. In fact it seems that even ground expansions based on the modal logic **K** have similar problems of weak groundedness as discussed in Example 4.2. Consider the premises

$$\Sigma = \{\neg Lp \to q, \neg L \neg Lp \to p\}. \tag{4.19}$$

Surprisingly, p follows from Σ and a result of negative introspection $\neg Lq$ using only the axiom schema **K**

$$L(\phi \to \psi) \to (L\phi \to L\psi). \tag{4.20}$$

Here is a proof.

1.	$\neg Lp \to q$	(premise)
2.	$\neg L \neg Lp \to p$	(premise)
3.	$\neg Lq$	(premise)
4.	$L(\neg Lp \to q)$	1 (necessitation rule)
5.	$L(\neg Lp \to q) \to (L\neg Lp \to Lq)$	(K)
6.	$L\neg Lp \to Lq$	4, 5 (propositional reasoning)
7.	$\neg L \neg Lp$	3, 6 (propositional reasoning)
8.	p	2, 7 (propositional reasoning)

Thus it seems that Σ has two ground **K**-expansions: one containing p and the other containing q. The former is quite weakly grounded in the premises as the belief p appears to be circularly justified.

In the fixed point equation (4.17) negative introspection is applied only to L free formulae and principles such as the axiom schema **K** (4.20) (or some weaker principle such as that proposed by Konolige [65]) are needed to accomplish further negative introspection. For example, if an expansion contains $L\phi$ and is not inconsistent, we should be able to infer that $\neg L \neg L\phi$ is in the expansion. The set of premises (4.19) provides an example which suggests that the existence of expansions containing circularly justified beliefs is unavoidable when using the fixed point equation (4.17) if nested L operators are allowed in the premises.

Chapter 5

Enumeration-Based Expansions

As shown in the previous chapter, it is not entirely straightforward to devise a satisfying characterization of the sets of beliefs of an agent with full introspection capabilities if tight groundedness in the premises is required. If the agent is ideal, rational and capable of full introspection, the sets of beliefs should be Moore style stable expansions. However, stable expansions may contain weakly grounded beliefs. To achieve tight groundedness the circularly justified beliefs must be eliminated. Hence only stable set minimal sets of beliefs should be considered. However, Example 4.2 shows that the minimality requirement is too weak to exclude all circularly justified beliefs. Iterative expansions proposed by Marek and Truszczyński [84] are a step in the right direction but are still too weakly grounded as they are not necessarily stable set minimal. Strong iterative expansions (which are the same as strongly grounded and robust expansions) are stable set minimal but are not satisfactory, especially, because of their dependence on the syntactic representation of premises. The variants of autoepistemic logic and McDermott and Doyle style nonmonotonic modal logics considered in the literature are based on a fixed point equation. As discussed in Chapter 4, circularly justified beliefs are difficult to exclude in these logics, in particular, if nested L operators are allowed in the premises. This suggests that the problems of weak groundedness might be connected to the use of a fixed point definition.

The author has proposed a new, more constructive method of defining sets of beliefs of an introspective agent in the propositional case [108]. The definition is based on enumerations of sentences instead of a fixed point equation. Davis [16] used this idea in connection with McDermott and Doyle's [96] nonmonotonic logic to characterize the intersection of the fixed points of the nonmonotonic consequence operator in McDermott and Doyle's logic. The author has extended the work of Davis and shown that also individual fixed points in McDermott and Doyle's logic can be characterized using enumerations of sentences [104]. The enumeration-based method tries to carry the notion of a proof to autoepistemic logic by formalizing autoepistemic reasoning as a sequence of introspection steps of an ideally rational agent. An enumeration of sentences specifies the order of the introspection steps.

In this chapter we show that the enumeration-based definition can be applied in the more general setting of this work where an underlying classical calculus is given and an autoepistemic logic is defined on top of that. We show that enumeration-based expansions are a proper subclass of Moore style stable expansions. We put forward a subclass of enumeration-based expansions, *L-hierarchic expansions*, as an interesting candidate for the class of tightly grounded sets of beliefs of an ideally rational fully introspective agent. L-hierarchic expansions are shown to be stable set minimal in the first-order case but are not dependent on the syntactic representation of the premises. We provide a finitary characterization for enumeration-based expansions and establish a close connection between iterative expansions and enumeration-based expansions.

The idea is to build a set of introspection beliefs $\mathbf{B}^{\varepsilon}(\Sigma)$ from premises Σ by applying introspection to sentences in the order given by an enumeration ε. The set $\mathbf{B}^{\varepsilon}(\Sigma)$ together with Σ induces the set of beliefs $\mathbf{SE}^{\varepsilon}(\Sigma)$ of a fully introspective agent having initial assumptions Σ after introspecting sentences in the order ε.

Definition 5.1 *Let* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$. *Let* $\varepsilon = \psi_1, \psi_2, \ldots$ *be an enumeration of sentences in* $\mathcal{L}_{\mathrm{ae}}$. *Let* $\mathbf{B}_0^{\varepsilon}(\Sigma) = \emptyset$.

Define $\mathbf{B}_{i+1}^{\varepsilon}(\Sigma)$ *for* $i = 0, 1, \ldots$ *as follows:*

$$\mathbf{B}_{i+1}^{\varepsilon}(\Sigma) = \begin{cases} \mathbf{B}_i^{\varepsilon}(\Sigma) \cup \{L\psi_{i+1}\} & \textit{if } \Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_{i+1} \\ \mathbf{B}_i^{\varepsilon}(\Sigma) \cup \{\neg L\psi_{i+1}\} & \textit{otherwise} \end{cases}$$

For a finite enumeration $\varepsilon = \psi_1, \psi_2, \ldots, \psi_n$,

$$\mathbf{B}^{\varepsilon}(\Sigma) = \mathbf{B}_n^{\varepsilon}(\Sigma)$$

and otherwise

$$\mathbf{B}^{\varepsilon}(\Sigma) = \bigcup_{i=0}^{\infty} \mathbf{B}_i^{\varepsilon}(\Sigma).$$

Finally

$$\mathbf{SE}^{\varepsilon}(\Sigma) = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi\}.$$

To guarantee that the introspection principles of autoepistemic reasoning are respected, i.e. that the set of beliefs $\mathbf{SE}^{\varepsilon}(\Sigma)$ induced by an introspection order ε satisfies the fixed point equation (2.6) for a stable expansion, we must require the enumeration ε to be acceptable with respect to the set of premises. This is because $\mathbf{SE}^{\varepsilon}(\Sigma)$ is not necessarily a stable expansion of Σ for every enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ as shown in the following example.

Example 5.1 Let the underlying logic be the propositional calculus. Consider a set of premises

$$\Sigma = \{\neg Lp \rightarrow q\} \tag{5.1}$$

and take an enumeration $\varepsilon = q, p, \ldots$ Then $\mathbf{B}_1^{\varepsilon}(\Sigma) = \{\neg Lq\}$ as $\Sigma \not\models_{\mathrm{ae}} q$ and $\mathbf{B}_2^{\varepsilon}(\Sigma) = \{\neg Lq, \neg Lp\}$ as $\Sigma \cup \mathbf{B}_1^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} p$. Thus $q \in \mathbf{SE}^{\varepsilon}(\Sigma)$ and $\neg Lq \in \mathbf{SE}^{\varepsilon}(\Sigma)$ but $Lq \notin \mathbf{SE}^{\varepsilon}(\Sigma)$. So $\mathbf{SE}^{\varepsilon}(\Sigma)$ does not satisfy the fixed point equation (2.6) defining stable expansions. ■

Definition 5.2 *An enumeration ε is* Σ-acceptable *if there is no i and no formula ϕ such that $\neg L\phi \in \mathbf{B}_i^{\varepsilon}(\Sigma)$ but $\Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$. A set of sentences Δ is an enumeration-based expansion of Σ if and only if $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$ for a Σ-acceptable enumeration ε of $\mathcal{L}_{\mathrm{ae}}$.*

We assume that the underlying consequence relation $\models_{\mathrm{ae}}$ is compact, i.e. if $\Sigma \models_{\mathrm{ae}} \phi$, there exists a finite subset Σ' of Σ such that $\Sigma' \models_{\mathrm{ae}} \phi$. For compact logics, the acceptability condition implies that results of introspection in the order ε cannot be overruled.

Proposition 5.3 *For all Σ-acceptable enumerations ε, if $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$, then $\psi_i \notin \mathbf{SE}^{\varepsilon}(\Sigma)$.*

Proof. Let $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ hold. Thus $\neg L\psi_i \in \mathbf{B}_i^{\varepsilon}(\Sigma)$. Assume that $\psi_i \in \mathbf{SE}^{\varepsilon}(\Sigma)$. Then $\Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$. As $\models_{\mathrm{ae}}$ is compact and monotonic, there exists $k \geq i$ such that $\Sigma \cup \mathbf{B}_k^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$. Because $k \geq i$, $\neg L\psi_i \in \mathbf{B}_i^{\varepsilon}(\Sigma)$ implies $\neg L\psi_i \in \mathbf{B}_k^{\varepsilon}(\Sigma)$. This leads to a contradiction as ε is Σ-acceptable. Thus $\psi_i \notin \mathbf{SE}^{\varepsilon}(\Sigma)$. □

Using this proposition it is straightforward to show that enumeration-based expansions are a subclass of Moore style stable expansions.

Theorem 5.4 *For all Σ-acceptable enumerations ε of $\mathcal{L}_{\mathrm{ae}}$, $\mathbf{SE}^{\varepsilon}(\Sigma)$ is a stable expansion of Σ.*

Proof. We prove that $\mathbf{SE}^{\varepsilon}(\Sigma) = \{\phi \mid \Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi\}$ is a stable expansion of Σ by showing that

$$\mathbf{B}^{\varepsilon}(\Sigma) = L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}. \tag{5.2}$$

($\subseteq$) If $L\psi_i \in \mathbf{B}^\varepsilon(\Sigma)$, then $\Sigma \cup \mathbf{B}^\varepsilon_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$ and as $\mathbf{B}^\varepsilon_{i-1}(\Sigma) \subseteq \mathbf{B}^\varepsilon(\Sigma)$, $\psi_i \in \mathbf{SE}^\varepsilon(\Sigma)$ by the monotonicity of $\models_{\mathrm{ae}}$ and thus $L\psi_i \in L\mathbf{SE}^\varepsilon(\Sigma)$. If $\neg L\psi_i \in \mathbf{B}^\varepsilon(\Sigma)$, then $\Sigma \cup \mathbf{B}^\varepsilon_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ and by Proposition 5.3 $\psi_i \notin \mathbf{SE}^\varepsilon(\Sigma)$. Thus $\neg L\psi_i \in \neg L\overline{\mathbf{SE}^\varepsilon(\Sigma)}$. This shows that $\mathbf{B}^\varepsilon(\Sigma) \subseteq L\mathbf{SE}^\varepsilon(\Sigma) \cup \neg L\overline{\mathbf{SE}^\varepsilon(\Sigma)}$.

($\supseteq$) If $\neg L\psi_i \in \neg L\overline{\mathbf{SE}^\varepsilon(\Sigma)}$, then $\Sigma \cup \mathbf{B}^\varepsilon(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ holds. This implies $\Sigma \cup \mathbf{B}^\varepsilon_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ by the monotonicity of $\models_{\mathrm{ae}}$ as $\mathbf{B}^\varepsilon_{i-1}(\Sigma) \subseteq \mathbf{B}^\varepsilon(\Sigma)$. Thus $\neg L\psi_i \in \mathbf{B}^\varepsilon(\Sigma)$. If $L\psi_i \in L\mathbf{SE}^\varepsilon(\Sigma)$, then $\psi_i \in \mathbf{SE}^\varepsilon(\Sigma)$ and by Proposition 5.3 $\Sigma \cup \mathbf{B}^\varepsilon_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$ and $L\psi_i \in \mathbf{B}^\varepsilon(\Sigma)$. Hence $\mathbf{B}^\varepsilon(\Sigma) = L\mathbf{SE}^\varepsilon(\Sigma) \cup \neg L\overline{\mathbf{SE}^\varepsilon(\Sigma)}$ and thus

$$\mathbf{SE}^\varepsilon(\Sigma) = \{\phi \mid \Sigma \cup L\mathbf{SE}^\varepsilon(\Sigma) \cup \neg L\overline{\mathbf{SE}^\varepsilon(\Sigma)} \models_{\mathrm{ae}} \phi\} \tag{5.3}$$

holds. □

In the fixed point approach correct reasoning is defined by presenting a condition (the fixed point equation) directly on the whole set of correct conclusions. The enumerations attempt to carry the notion of a proof to autoepistemic logic. In the enumeration-based approach autoepistemic reasoning is formalized as a sequence of introspection steps. The acceptability condition can also be verified step by step as if $\neg L\phi \in \mathbf{B}^\varepsilon_i(\Sigma)$, then ϕ is some ψ_j and $j \leq i$.

Enumerations are not capable of capturing every Moore style stable expansion. Consider the propositional case and the premise $\Sigma = \{Lp \rightarrow p\}$. It has two stable expansions: $\mathrm{SE}_\Sigma(\{Lp\})$ containing p and $\mathrm{SE}_\Sigma(\{\neg Lp\})$ not containing p. Only the latter is characterizable using enumerations because in enumeration-based expansions Lp cannot be used in deriving p directly. This leads one to assume that enumerations might produce minimal expansions. However, this is not the case as shown by the following example.

Example 5.2 Let the underlying logic be the propositional calculus. Consider a premise

$$\Sigma = \{\neg L\neg Lp \rightarrow p\} \tag{5.4}$$

where p is an atomic formula. The two stable expansions $\mathrm{SE}_\Sigma(\{L\neg Lp\})$ and $\mathrm{SE}_\Sigma(\{\neg L\neg Lp\})$ of Σ are both characterizable using enumerations. The former not containing p using an enumeration $p, \neg Lp, \ldots$ and the latter, non-minimal expansion containing p using an enumeration $\neg Lp, p, \ldots$ Let us look what happens in the second enumeration ε: $\mathbf{B}^\varepsilon_1(\Sigma) = \{\neg L\neg Lp\}$ as $\Sigma \not\models_{\mathrm{ae}} \neg Lp$, but then $\mathbf{B}^\varepsilon_2(\Sigma) = \{\neg L\neg Lp, Lp\}$ as $\Sigma \cup \mathbf{B}^\varepsilon_1(\Sigma) \models_{\mathrm{ae}} p$. ■

Thus enumeration-based expansions are not necessarily minimal with respect to their L free parts because it is possible to introspect a sentence ϕ without first introspecting the subformulae of ϕ which can effect the result of the introspection of ϕ. In the case of Example 5.2 it is possible to introspect $\neg Lp$ and obtain $\neg L\neg Lp$ without first checking the status of p. In order to obtain tightly grounded beliefs, it seems natural to demand that all subformulae of a given sentence ϕ that can effect the result of the introspection of ϕ must be examined before ϕ. This new groundedness requirement can be incorporated easily to the definition of enumeration-based expansions. We require the enumerations to be *L-hierarchic*. This means that a sentence ϕ can appear in the enumeration only after all ψ such that $L\psi$ is a subformula of ϕ have appeared in the enumeration.

Definition 5.5 *Let ε be an enumeration $\psi_1, \psi_2, \ldots$ of a set of formulae $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$. The enumeration ε is* L-hierarchic *if and only if for all $\psi_i, \psi_j \in \Sigma$, $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\psi_j)$ implies $i < j$.*

Definition 5.6 *A set of sentences Δ is an* L-hierarchic expansion *of Σ if and only if $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$ for some Σ-acceptable L-hierarchic enumeration ε of $\mathcal{L}_{\mathrm{ae}}$.*

It turns out that the simple requirement on the order of the sentences in the enumerations guarantees that in the first-order case L-hierarchic expansions are stable set minimal, i.e. minimal with respect to their L free parts.

Theorem 5.7 (Niemelä [108]) *Let the underlying logic be a first-order calculus. Every L-hierarchic expansion $\mathbf{SE}^{\varepsilon}(\Sigma)$ of Σ is a stable set minimal stable expansion of Σ.*

Proof. We denote $\mathbf{SE}^{\varepsilon}(\Sigma)$ by Δ. Assume that there exists a stable set S containing Σ such that $S \cap \mathcal{L} \subseteq \Delta \cap \mathcal{L}$. We prove that $\Delta = S$ by showing that for all $i = 1, 2, \ldots$, $\psi_i \in \Delta$ if and only if $\psi_i \in S$ by induction on i. This proves that Δ is stable set minimal for Σ.

Let $i = 1$. ($\Rightarrow$) If $\psi_1 \in S$, then as ε is L-hierarchic, $\psi_1 \in \mathcal{L}$ and $\psi_1 \in \Delta$. ($\Leftarrow$) If $\psi_1 \in \Delta$, then by Proposition 5.3 $\Sigma \models_{\mathrm{ae}} \psi_1$ as ε is Σ-acceptable. Thus $\psi_1 \in S$ as $\Sigma \subseteq S$ and S is stable.

Assume that for all $j = 1, 2, \ldots, i-1$, $\psi_j \in \Delta$ if and only if $\psi_j \in S$. ($\Rightarrow$) If $\psi_i \in \Delta$, then $\Sigma \cup \mathbf{B}^{\varepsilon}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$ as ε is Σ-acceptable. If $L\psi_j \in \mathbf{B}^{\varepsilon}_{i-1}(\Sigma)$, then $\psi_j \in \Delta$. Thus by the inductive hypothesis $\psi_j \in S$ and as S is stable $L\psi_j \in S$. If $\neg L\psi_j \in \mathbf{B}^{\varepsilon}_{i-1}(\Sigma)$, then $\psi_j \notin \Delta$. Thus $\psi_j \notin S$ and as S is stable, $\neg L\psi_j \in S$. Hence $\mathbf{B}^{\varepsilon}_{i-1}(\Sigma) \subseteq S$. Thus $\psi_i \in S$ as $\Sigma \cup \mathbf{B}^{\varepsilon}_{i-1}(\Sigma) \subseteq S$ and S is stable. ($\Leftarrow$) As ε is L-hierarchic, then if $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\psi_i)$, χ is some ψ_j and $j < i$. Thus by the inductive hypothesis $\psi_j \in \Delta$ if and only if $\psi_j \in S$. Hence for all $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\psi_i)$, $\chi \in \Delta$ if and only if $\chi \in S$ which implies by Proposition 2.19 that $\psi_i \in S$ implies $\psi_i \in \Delta$. Thus $\Delta = S$. $\square$

Now we turn to the problem of characterizing enumeration-based expansions finitely. We prove an important theorem which shows that for sets of premises with finitely many $L\phi$ subformulae, it is sufficient to examine finite enumerations of sentences in order to find an acceptable enumeration of $\mathcal{L}_{\mathrm{ae}}$. But first we note that for a finite enumeration the test for acceptability can be stated in the following simple form.

Proposition 5.8 *Let $\varepsilon_n = \psi_1, \ldots, \psi_n$ be an enumeration of a finite set. Then ε_n is Σ-acceptable if and only if ε_n satisfies the condition:*

$$\textit{for all } \neg L\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma), \Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \phi. \tag{5.5}$$

Proof. ($\Rightarrow$) Let ε_n be Σ-acceptable. If $\neg L\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then ϕ is some ψ_i in ε_n and $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. So by Proposition 5.3 $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \phi$ holds. ($\Leftarrow$) Let Condition (5.5) hold. Assume that ε_n is not Σ-acceptable. Then there is $i \leq n$ such that $\neg L\phi \in \mathbf{B}_i^{\varepsilon_n}(\Sigma)$ but $\Sigma \cup \mathbf{B}_i^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \phi$. But as $\mathbf{B}_i^{\varepsilon_n}(\Sigma) \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\models_{\mathrm{ae}}$ is monotonic, $\neg L\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \phi$, a contradiction. □

Using Proposition 3.18 it can be shown that for every acceptable enumeration of the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ there is a corresponding full set.

Proposition 5.9 *Let Σ be a set of sentences and let $\varepsilon = \psi_1, \psi_2, \ldots$ be a Σ-acceptable enumeration of the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. Then*

$$\Lambda = \{\phi \in \mathbf{B}^{\varepsilon}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\} \tag{5.6}$$

is a Σ-full set and for all $\chi \in \mathcal{L}_{\mathrm{ae}}$, $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ if and only if $\Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{L}} \chi$.

Proof. The proposition is a consequence of Proposition 3.18. If $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, then χ is some ψ_i. If $\Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$, then as ε is Σ-acceptable, by Proposition 5.3 $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$ and thus $L\psi_i \in \mathbf{B}^{\varepsilon}(\Sigma)$. Also $L\psi_i \in \mathbf{B}^{\varepsilon}(\Sigma)$ implies $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$ and by the monotonicity of $\models_{\mathrm{ae}}$ $\Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$. Similarly, $\Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ if and only if $\neg L\psi_i \in \mathbf{B}^{\varepsilon}(\Sigma)$. Thus $\mathbf{B}^{\varepsilon}(\Sigma)$ satisfies the conditions given in Proposition 3.18. Therefore, $\Lambda = \{\phi \in \mathbf{B}^{\varepsilon}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$ is a Σ-full set and for all $\chi \in \mathcal{L}_{\mathrm{ae}}$, $\Sigma \cup \Lambda \models_{\mathrm{L}} \chi$ if and only if $\Sigma \cup \mathbf{B}^{\varepsilon}(\Sigma) \models_{\mathrm{L}} \chi$. □

Theorem 5.10 *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ be a set of sentences such that $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is finite. If an enumeration $\varepsilon_n = \psi_1, \ldots, \psi_n$ of the finite set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ is Σ-acceptable, then*

$$\Lambda = \{\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$$

is Σ-full and there exists an Σ-acceptable enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ which is an extension of ε_n such that $\mathbf{SE}^{\varepsilon}(\Sigma)$ is the unique stable expansion Δ of Σ for which $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$ (i.e. $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(\mathbf{B}_n^{\varepsilon_n}(\Sigma))$). Furthermore, if ε_n is L-hierarchic, then also ε is.

Proof. Let ε_n be Σ-acceptable. Proposition 5.9 implies that $\Lambda = \{\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)\}$ is a Σ-full set. Let ε be an extension of ε_n such that for all $i, j > n$ if $L\psi_j \in \mathrm{Sf}^{\,\mathrm{L}}(\psi_i)$, then $j < i$. Let ε_n be L-hierarchic. Assume ε is not. So there is some i, j such that $L\psi_j \in \mathrm{Sf}^{\,\mathrm{L}}(\psi_i)$ but $j > i$. If $i > n$, then $j > n$ which leads to a contradiction. Thus $i \leq n$. So $L\psi_i \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$ and thus $L\psi_j \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$. Hence $j \leq n$ which leads to contradiction as ε_n is L-hierarchic. Thus ε is L-hierarchic.

Assume that ε is not Σ-acceptable. Then there exists some ψ_j and i such that $\Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j$ but $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$ and thus $\neg L\psi_j \in \mathbf{B}_i^{\varepsilon}(\Sigma)$. Let i be the index such that $\Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j$ but $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$. So $\neg L\psi_j \in \mathbf{B}_i^{\varepsilon}(\Sigma)$ and $j \leq i$. By Proposition 2.17 $L\psi_i \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\psi_j\})$. If $i \leq n$, then $j \leq n$. As ε_n is Σ-acceptable, $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$ holds which implies $\Sigma \cup \mathbf{B}_i^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$ and thus $\Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$, a contradiction. So $i > n$ and thus $L\psi_i \notin \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$ which implies $L\psi_i \in \mathrm{Sf}^{\,\mathrm{qL}}(\psi_j)$. Now $j \leq n$ because otherwise $j > i$. But then $L\psi_j \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$ and thus $L\psi_i \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$, a contradiction. Thus ε is Σ-acceptable.

If $L\psi_i \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_i$. As ε is an extension of ε_n, $\psi_i \in \mathbf{SE}^{\varepsilon}(\Sigma)$ and thus $L\psi_i \in L\mathbf{SE}^{\varepsilon}(\Sigma)$. If $\neg L\psi_i \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. As ε is an extension of ε_n and Σ-acceptable, $\psi_i \notin \mathbf{SE}^{\varepsilon}(\Sigma)$ by Proposition 5.3 and thus $\neg L\psi_i \in \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$. Hence $\mathbf{B}_n^{\varepsilon_n}(\Sigma) \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ holds and thus $\Lambda \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ holds. By Corollary 3.11 $\mathrm{SE}_{\Sigma}(\Lambda)$ is the unique stable expansion Δ of Σ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. As $\mathbf{SE}^{\varepsilon}(\Sigma)$ is a stable expansion by Theorem 5.4, $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda)$. By Proposition 5.9 $\mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(\mathbf{B}_n^{\varepsilon_n}(\Sigma))$. □

When an acceptable enumeration is projected to the $L\phi$ subformulae of the premises, it remains acceptable.

Theorem 5.11 *Let Σ be a set of sentences. For every Σ-acceptable (L-hierarchic) enumeration ε of $\mathcal{L}_{\mathrm{ae}}$, there exists a Σ-acceptable (L-hierarchic) enumeration $\varepsilon_L = \psi_1, \psi_2, \ldots$ of the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)\}$ such that*

$$\Lambda = \{\phi \in \mathbf{B}^{\varepsilon_L}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)\} \tag{5.7}$$

is the unique Σ-full set for which $\Lambda \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ holds and thus $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(\mathbf{B}^{\varepsilon_L}(\Sigma))$.

Proof. Let ε be a Σ-acceptable (L-hierarchic) enumeration of $\mathcal{L}_{\mathrm{ae}}$ and let $\varepsilon_L = \psi_1, \ldots$ be the projection of ε to the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)\}$, i.e. an enumeration of the sentences $\{\phi \mid L\phi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)\}$ in ε in the same order they appear in ε. We show that for all $i = 1, 2, \ldots,$

$$\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_L}(\Sigma) \models_{\mathrm{ae}} \psi_i \text{ if and only if } \psi_i \in \mathbf{SE}^{\varepsilon}(\Sigma). \tag{5.8}$$

Assume that (5.8) does not hold and let i the smallest index for which it fails. If $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$, then as (5.8) holds for every $j < i$, $\mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma) \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ holds. This implies $\psi_i \in \mathbf{SE}^{\varepsilon}(\Sigma)$ as $\mathbf{SE}^{\varepsilon}(\Sigma)$ is a stable expansion of Σ. Thus it must be the case that $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ but $\psi_i \in \mathbf{SE}^{\varepsilon}(\Sigma)$. The formula ψ_i is some $\psi_{i'}$ in ε and by Proposition 5.3 $\Sigma \cup \mathbf{B}^{\varepsilon}_{i'-1}(\Sigma) \models_{\mathrm{ae}} \psi_{i'}$. By Proposition 2.17 there is a set $\Lambda \subseteq \mathbf{B}^{\varepsilon}_{i'-1}(\Sigma)$ such that $\Sigma \cup \Lambda \models_{\mathrm{ae}} \psi_{i'}$ and $\text{q-atom}(\Lambda) \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \{\psi_{i'}\})$. As $L\psi_{i'} \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$, $\text{q-atom}(\Lambda) \subseteq \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)$. Thus if $L\psi_{j'} \in \Lambda$, then $j' < i'$ and $\Sigma \cup \mathbf{B}^{\varepsilon}_{j'-1}(\Sigma) \models_{\mathrm{ae}} \psi_{j'}$ holds which implies $\psi_{j'} \in \mathbf{SE}^{\varepsilon}(\Sigma)$. The formula $\psi_{j'}$ is some ψ_j in ε_L and $j < i$. Thus as $\psi_j \in \mathbf{SE}^{\varepsilon}(\Sigma)$ holds, $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{j-1}(\Sigma) \models_{\mathrm{ae}} \psi_j$ holds which implies $L\psi_{j'} \in \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma)$. If $\neg L\psi_{j'} \in \Lambda$, then $j' < i'$ and $\Sigma \cup \mathbf{B}^{\varepsilon}_{j'-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_{j'}$ holds and thus $\psi_{j'} \notin \mathbf{SE}^{\varepsilon}(\Sigma)$ by Proposition 5.3 as ε is Σ-acceptable. The formula $\psi_{j'}$ is some ψ_j in ε_L and $j < i$. Thus as $\psi_j \notin \mathbf{SE}^{\varepsilon}(\Sigma)$ holds, $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{j-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$ holds implying $\neg L\psi_{j'} \in \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma)$. So $\Lambda \subseteq \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma)$ which implies $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$, a contradiction. Hence (5.8) holds for every $i = 1, 2, \ldots$

Condition (5.8) implies that $\mathbf{B}^{\varepsilon_L}_{n}(\Sigma) \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ which in turn implies that ε_L is Σ-acceptable in the following way. Let $\neg L\psi_i \in \mathbf{B}^{\varepsilon_L}_{n}(\Sigma)$. Then $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ and by (5.8) $\psi_i \notin \mathbf{SE}^{\varepsilon}(\Sigma)$. Thus $\Sigma \cup L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)} \not\models_{\mathrm{ae}} \psi_i$ which implies $\Sigma \cup \mathbf{B}^{\varepsilon_L}_{n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ by the monotonicity of $\models_{\mathrm{ae}}$. So by Proposition 5.8 ε_L is Σ-acceptable and by Proposition 5.9 Λ is a Σ-full set. Furthermore, $\Lambda \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ holds as $\mathbf{B}^{\varepsilon_L}(\Sigma) \subseteq L\mathbf{SE}^{\varepsilon}(\Sigma) \cup \neg L\overline{\mathbf{SE}^{\varepsilon}(\Sigma)}$ holds. By Corollary 3.11 $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda)$ and by Proposition 5.9 $\mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(\mathbf{B}^{\varepsilon_L}(\Sigma))$. □

Thus given a set of premises having only finitely many $L\phi$ subformulae, the finite Σ-acceptable enumerations of the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)\}$ characterize enumeration-based expansions. For each such finite enumeration ε, there is an enumeration-based expansion which is an L-hierarchic expansion if ε is L-hierarchic. On the other hand, for each enumeration-based expansion there is a corresponding finite acceptable enumeration of the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma)\}$.

As the first application of the finitary characterization, we show that for premises of a particularly simple type, L-hierarchic expansions and Moore style expansions coincide. We start by establishing a proposition which simplifies the proof.

Proposition 5.12 *Let the underlying logic be a first-order calculus. Let Σ be a set of sentences of the form*

$$\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi \tag{5.9}$$

where $n \geq 0$ and $\psi \in \mathcal{L}$ and let Λ be a set of quasi-literals containing no formula and its negation simultaneously. Then for all $\chi \in \mathcal{L}$,

$$\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi \text{ if and only if } \mathrm{Hds}(\Sigma, \Lambda) \models_{\mathrm{cl}} \chi$$

where $\mathrm{Hds}(\Sigma, \Lambda) = \{\psi \mid \neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi \in \Sigma$ *and* $\neg L\phi_i \in \Lambda$ *for all* $i = 1, \ldots, n\}$.

Proof. ($\Rightarrow$) Let $\mathrm{Hds}(\Sigma, \Lambda) \not\models_{\mathrm{cl}} \chi$ hold. Then there is a CL-structure I such that $I \models_{\mathrm{cl}} \mathrm{Hds}(\Sigma, \Lambda)$ and $I \not\models_{\mathrm{cl}} \chi$. Let $T = \{\phi \mid \neg L\phi \notin \Lambda\}$. Now $\langle I, T\rangle \models_{\mathrm{ae}} \Lambda$ because if $L\phi \in \Lambda$, then as Λ contains no formula and its negation, $\neg L\phi \notin \Lambda$ and thus $\phi \in T$. On the other hand, if $\neg L\phi \in \Lambda$, then $\phi \notin T$. As $\chi \in \mathcal{L}$ and $I \not\models_{\mathrm{cl}} \chi$, $\langle I, T\rangle \not\models_{\mathrm{ae}} \chi$ holds. Let $\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi \in \Sigma$. If $\langle I, T\rangle \not\models_{\mathrm{ae}} \psi$, then as $\psi \in \mathcal{L}$, $I \not\models_{\mathrm{cl}} \psi$ and thus $\psi \notin \mathrm{Hds}(\Sigma, \Lambda)$. This implies that for some i, $1 \leq i \leq n$, $\neg L\phi_i \notin \Lambda$. Hence $\phi_i \in T$ and thus $\langle I, T\rangle \not\models_{\mathrm{ae}} \neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n$. So $\langle I, T\rangle \models_{\mathrm{ae}} \neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi$ and $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma$. We have established that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda$ and $\langle I, T\rangle \not\models_{\mathrm{ae}} \chi$ holds. Thus $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} \chi$ holds.

($\Leftarrow$) Let $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} \chi$ hold. Then there exists an AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda$ and $\langle I, T\rangle \not\models_{\mathrm{ae}} \chi$ holds. Thus $I \not\models_{\mathrm{cl}} \chi$ holds. Let $\psi \in \mathrm{Hds}(\Sigma, \Lambda)$. So there is some $\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi \in \Sigma$ and $\neg L\phi_i \in \Lambda$ for all $i = 1, \ldots, n$. But then as $\langle I, T\rangle \models_{\mathrm{ae}} \Lambda \cup \{\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi\}$, $\langle I, T\rangle \models_{\mathrm{ae}} \psi$ which implies $I \models_{\mathrm{cl}} \psi$ as $\psi \in \mathcal{L}$. We have shown that $I \models_{\mathrm{cl}} \mathrm{Hds}(\Sigma, \Lambda)$ and $I \not\models_{\mathrm{cl}} \chi$ which implies that $\mathrm{Hds}(\Sigma, \Lambda) \not\models_{\mathrm{cl}} \chi$ holds. □

Theorem 5.13 *Let the underlying logic be a first-order calculus. Let* Σ *be a set of sentences of the form*

$$\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_n \rightarrow \psi$$

where $n \geq 0$ *and* $\phi_i, \psi \in \mathcal{L}$ *such that* $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ *is finite. Then* Δ *is a stable expansion of* Σ *if and only if* Δ *is an L-hierarchic expansion of* Σ.

Proof. By Theorem 5.4 all L-hierarchic expansions are also stable expansions. Thus it suffices to show that for each stable expansion there is an L-hierarchic enumeration producing it. Let Δ be a stable expansion of Σ. Then by Lemma 3.5

$$\Lambda = \{\chi \in L\Delta \cup \neg L\overline{\Delta} \mid \text{q-atom}(\chi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\} \tag{5.10}$$

is a Σ-full set and by Corollary 3.11 $\Delta = \mathrm{SE}_{\Sigma}(\Lambda)$. By Theorem 5.10 we have to show that there is a Σ-acceptable L-hierarchic enumeration $\varepsilon_n = \psi_1, \ldots, \psi_n$ of the finite set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ and that

$$\Lambda = \{\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}. \tag{5.11}$$

Then by Theorem 5.10 there exists a Σ-acceptable L-hierarchic enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ which is an extension of ε_n such that $\mathbf{SE}^{\varepsilon}(\Sigma)$ is the unique stable expansion Δ of Σ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Thus $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda) = \Delta$ which implies that Δ is an L-hierarchic expansion of Σ.

Let $\varepsilon_n = \psi_1, \ldots, \psi_n$ be an enumeration of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that if $\psi_i \notin \Delta$ but $\psi_j \in \Delta$, then $i < j$. Clearly, ε_n is L-hierarchic as the sentences in the enumeration contain no L operators. We prove that ε_n is Σ-acceptable and that (5.11) holds by showing that for all $i = 1, \ldots, n$,

$$\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i \text{ if and only if } \psi_i \in \Delta. \tag{5.12}$$

Let $i = 1$. Then $\Sigma \cup \mathbf{B}^{\varepsilon_n}_0(\Sigma) \models_{\mathrm{ae}} \psi_1$ if and only if $\Sigma \models_{\mathrm{ae}} \psi_1$ which implies $\psi_1 \in \Delta$ as Δ is a stable expansion of Σ. If $\psi_1 \in \Delta$, then $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \psi_1$ and for all $i = 1, \ldots, n$, $\psi_i \in \Delta$ by the condition on the enumeration ε_n. By Proposition 5.12 $\mathrm{Hds}(\Sigma, L\Delta \cup \neg L\overline{\Delta}) \models_{\mathrm{cl}} \psi_1$. For every $\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_m \rightarrow \chi \in \Sigma$ and for every $l = 1, \ldots, m$, ϕ_l is some ψ_j in the enumeration and as for all $i = 1, \ldots, n$, $\psi_i \in \Delta$, $\neg L\phi_l \notin \neg L\overline{\Delta}$. Thus $\mathrm{Hds}(\Sigma, L\Delta \cup \neg L\overline{\Delta}) = \mathrm{Hds}(\Sigma, \emptyset)$. So $\mathrm{Hds}(\Sigma, \emptyset) \models_{\mathrm{cl}} \psi_1$ holds which implies $\Sigma \models_{\mathrm{ae}} \psi_1$ by Proposition 5.12.

Let $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$. By the inductive hypothesis $\mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \subseteq L\Delta \cup \neg L\overline{\Delta}$. Thus $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \psi_i$ and $\psi_i \in \Delta$. Let $\psi_i \in \Delta$. Thus $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \psi_i$. By Proposition 5.12 $\mathrm{Hds}(\Sigma, L\Delta \cup \neg L\overline{\Delta}) \models_{\mathrm{cl}} \psi_i$. Let $\chi \in \mathrm{Hds}(\Sigma, L\Delta \cup \neg L\overline{\Delta})$. Then some $\neg L\phi_1 \wedge \cdots \wedge \neg L\phi_m \rightarrow \chi \in \Sigma$ and for all $l = 1, \ldots, m$, ϕ_l is some ψ_j in the enumeration and $\psi_j \notin \Delta$. Then by the condition on the enumeration and as $\psi_i \in \Delta$, $j < i$. Thus by the inductive hypothesis $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{j-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$. So $\neg L\psi_j \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$. We have shown that $\mathrm{Hds}(\Sigma, L\Delta \cup \neg L\overline{\Delta}) \subseteq \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma))$. By the monotonicity of $\models_{\mathrm{cl}}$ $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) \models_{\mathrm{cl}} \psi_i$ and by Proposition 5.12 $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$.

Thus for all $i = 1, \ldots, n$ (5.12) holds. This implies that ε_n is Σ-acceptable in the following way. If $\neg L\psi_i \in \mathbf{B}^{\varepsilon_n}_n(\Sigma)$, then $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ and by (5.12) $\psi_i \notin \Delta$. Hence $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \not\models_{\mathrm{ae}} \psi_i$ and thus $\Sigma \cup \mathbf{B}^{\varepsilon_n}_n(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ as by (5.12) $\mathbf{B}^{\varepsilon_n}_n(\Sigma) \subseteq L\Delta \cup \neg L\overline{\Delta}$. So by Proposition 5.8 ε_n is Σ-acceptable. Furthermore, $\mathbf{B}^{\varepsilon_n}_n(\Sigma) \subseteq L\Delta \cup \neg L\overline{\Delta}$ implies that (5.11) holds. Thus Δ is an L-hierarchic expansion of Σ. □

The definition of iterative expansions fits easily the general setting of this work. Iterative expansions based on a given underlying logic can be defined as iterative expansions in the propositional case (Definition 4.3). We aim to show that acceptable enumerations produce iterative expansions and that for every iterative expansion there is a corresponding enumeration producing it.

Theorem 5.14 *For all Σ-acceptable enumerations ε of $\mathcal{L}_{\mathrm{ae}}$, $\mathbf{SE}^{\varepsilon}(\Sigma)$ is an iterative expansion of Σ.*

Proof. Let $\varepsilon = \psi_1, \psi_2, \ldots$ be a Σ-acceptable enumeration of $\mathcal{L}_{\mathrm{ae}}$. We have to show that for $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$, $A^{\Delta}(\Sigma) = \Delta$ holds. First we show that for all $i = 1, 2, \ldots,$

$$\psi_i \in \Delta \text{ implies } \psi_i \in A^{\Delta}_i(\Sigma) \tag{5.13}$$

by induction on i. If $i = 1$, then $\psi_1 \in \Delta$ implies by Proposition 5.3 $\Sigma \models_{\mathrm{ae}} \psi_1$ as ε is Σ-acceptable. Hence $\psi_1 \in A_0^\Delta(\Sigma) \subseteq A_1^\Delta(\Sigma)$ by the monotonicity of $\models_{\mathrm{ae}}$. Let $i > 1$. Again $\psi_i \in \Delta$ implies $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_i$. First we note that by the monotonicity of $\models_{\mathrm{ae}}$, $A_j^\Delta(\Sigma) \subseteq A_i^\Delta(\Sigma)$ for all i, j such that $0 \leq j \leq i$ and $\Sigma \cup \neg L\overline{\Delta} \subseteq A_{i-1}^\Delta(\Sigma)$ for all $i \geq 1$. If $\neg L\phi \in \mathbf{B}_{i-1}^{\varepsilon}(\Sigma)$, ϕ is some ψ_j such that $j < i$. As ε is Σ-acceptable, $\phi \notin \Delta$ by Proposition 5.3. This implies that $\neg L\phi \in \neg L\overline{\Delta} \subseteq A_{i-1}^\Delta(\Sigma)$. If $L\phi \in \mathbf{B}_{i-1}^{\varepsilon}(\Sigma)$, ϕ is some ψ_j such that $j < i$ and $\psi_j \in \Delta$. By the inductive hypothesis $\psi_j \in A_j^\Delta(\Sigma) \subseteq A_{i-1}^\Delta(\Sigma)$. Hence $L\psi_j \in LA_{i-1}^\Delta(\Sigma)$. We have shown that $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \subseteq A_{i-1}^\Delta(\Sigma) \cup LA_{i-1}^\Delta(\Sigma)$ holds. This implies that $\psi_i \in A_i^\Delta(\Sigma) = Cn(A_{i-1}^\Delta(\Sigma) \cup LA_{i-1}^\Delta(\Sigma))$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus (5.13) holds which implies immediately that $\Delta \subseteq A^\Delta(\Sigma)$ holds.

Then we show by induction on i that for all $i = 0, 1, 2, \ldots$, $A_i^\Delta(\Sigma) \subseteq \Delta$ which implies $A^\Delta(\Sigma) \subseteq \Delta$ and thus $A^\Delta(\Sigma) = \Delta$. Let $\phi \in A_0^\Delta(\Sigma)$. Then $\Sigma \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi$. By Theorem 5.4, Δ is a stable expansion. Hence, $\phi \in \Delta = \{\phi \mid \Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \phi\}$ by the monotonicity of $\models_{\mathrm{ae}}$ and $A_0^\Delta(\Sigma) \subseteq \Delta$ holds. Let $\phi \in A_i^\Delta(\Sigma)$. Then $A_{i-1}^\Delta(\Sigma) \cup LA_{i-1}^\Delta(\Sigma) \models_{\mathrm{ae}} \phi$. By the inductive hypothesis $A_{i-1}^\Delta(\Sigma) \subseteq \Delta$ and thus $LA_{i-1}^\Delta(\Sigma) \subseteq L\Delta \subseteq \Delta$. By the monotonicity of $\models_{\mathrm{ae}}$, $\Delta \models_{\mathrm{ae}} \phi$ holds. As Δ is a stable expansion, it is closed under $\models_{\mathrm{ae}}$. Hence $\phi \in \Delta$ and $A_i^\Delta(\Sigma) \subseteq \Delta$ holds. □

Theorem 5.15 *For each iterative expansion Δ of a set of premises Σ such that $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is finite, there is a Σ-acceptable enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ such that $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$.*

Proof. By Theorem 5.10 the existence of an enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ inducing an iterative expansion Δ can be shown by constructing an appropriate Σ-acceptable enumeration ε_n of the finite set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. It is sufficient that ε_n satisfies the following two conditions.

(C1) *$\psi_1, \ldots, \psi_{m_1}$ are all the sentences $\psi_i \notin \Delta$ for which $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$.*

(C2) *$\psi_{m_1+1}, \ldots, \psi_n$ are all the sentences $\psi_i \in \Delta$ for which $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ in the order they are produced by the iteration of A^Δ, i.e. if $\psi_i, \psi_j \in \Delta$ and $L\psi_i, L\psi_j \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and there exists k such that $\psi_i \in A_k^\Delta(\Sigma)$ but $\psi_j \notin A_k^\Delta(\Sigma)$, then $i < j$.*

We start by providing an alternative characterization (5.14) of the operator A_i^Δ (4.10) used in the definition of iterative expansions. The characterization makes the role of positive introspection in iterative expansions more explicit. For all n = 0,1,...,

$$A_{n+1}^\Delta(\Sigma) = \mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^\Delta(\Sigma) \cup \ldots \cup A_n^\Delta(\Sigma))). \tag{5.14}$$

We show that (5.14) holds by induction on n. Let $n = 0$. Then

$$\begin{aligned} A_1^\Delta(\Sigma) &= \mathrm{Cn}(A_0^\Delta(\Sigma) \cup LA_0^\Delta(\Sigma)) && (5.15)\\ &= \mathrm{Cn}(\mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta}) \cup LA_0^\Delta(\Sigma)) && (5.16)\\ &= \mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta} \cup LA_0^\Delta(\Sigma)) && (5.17) \end{aligned}$$

where (5.15) holds by the definition of the operator A_i^Δ (4.10), (5.16) holds as by (4.9) $A_0^\Delta(\Sigma) = \mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta})$ and (5.17) holds as for any sets of sentences $\Sigma_1, \Sigma_2 \subseteq \mathcal{L}_{\mathrm{ae}}$, $\mathrm{Cn}(\mathrm{Cn}(\Sigma_1) \cup \Sigma_2) = \mathrm{Cn}(\Sigma_1 \cup \Sigma_2)$.

Assume that $A_n^\Delta(\Sigma) = \mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^\Delta(\Sigma) \cup \ldots \cup A_{n-1}^\Delta(\Sigma)))$.

$$\begin{aligned} &A_{n+1}^\Delta(\Sigma)\\ &\quad= \mathrm{Cn}(A_n^\Delta(\Sigma) \cup LA_n^\Delta(\Sigma)) && (5.18)\\ &\quad= \mathrm{Cn}(\mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^\Delta(\Sigma) \cup \ldots \cup A_{n-1}^\Delta(\Sigma))) \cup\\ &\qquad\quad LA_n^\Delta(\Sigma)) && (5.19)\\ &\quad= \mathrm{Cn}(\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^\Delta(\Sigma) \cup \ldots \cup A_{n-1}^\Delta(\Sigma) \cup A_n^\Delta(\Sigma))) && (5.20) \end{aligned}$$

where (5.18) holds by the definition of the operator A_i^Δ (4.10), (5.19) holds by the inductive hypothesis and (5.20) holds as for any sets of sentences $\Sigma_1, \Sigma_2 \subseteq \mathcal{L}_{\mathrm{ae}}$, $\mathrm{Cn}(\mathrm{Cn}(\Sigma_1) \cup \Sigma_2) = \mathrm{Cn}(\Sigma_1 \cup \Sigma_2)$. Thus (5.14) holds.

Then we prove two facts about ε_n.

(F1) *For all* $i = 1, \ldots, m_1 : \Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$.

Let $i \leq m_1$ be the first index such that $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_i$. Then $\mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \subseteq \neg L\overline{\Delta}$. Since Δ is a stable expansion of Σ, by the monotonicity of $\models_{\mathrm{ae}}$, $\psi_i \in \Delta$, a contradiction. Thus (F1) holds.

(F2) *For all* $i = m_1 + 1, \ldots, n : \Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_i$.

Let i be the first index such that $m_1 < i \leq n$ and $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. As $\psi_i \in \Delta$, there is some minimal k such that $\psi_i \in A_k^\Delta(\Sigma)$. If $k = 0$, then $\Sigma \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \psi_i$ or otherwise by (5.14) $\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^\Delta(\Sigma) \cup \ldots \cup A_{k-1}^\Delta(\Sigma)) \models_{\mathrm{ae}} \psi_i$. By Proposition 2.17 there is a set $\Lambda \subseteq \neg L\overline{\Delta} \cup L(A_0^\Delta(\Sigma) \cup \ldots \cup A_{k-1}^\Delta(\Sigma))$ such that $\Sigma \cup \Lambda \models_{\mathrm{ae}} \psi_i$ and $\text{q-atom}(\Lambda) \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\psi_i\})$. As $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, $\text{q-atom}(\Lambda) \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. If $\neg L\psi_j \in \Lambda$, then $\psi_j \notin \Delta$ and thus by (C1) $j \leq m_1$ and by (F1) $\neg L\psi_j \in \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma)$. If $L\psi_j \in \Lambda$, then $\psi_j \in A_0^\Delta(\Sigma) \cup \ldots \cup A_{k-1}^\Delta(\Sigma)$. Thus by (C2) $j < i$. So $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_j$ and $L\psi_j \in \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma)$. Hence $\Lambda \subseteq \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma)$ and by the monotonicity of $\models_{\mathrm{ae}}$ $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_i$, a contradiction. Thus (F2) holds.

Using the facts (F1) and (F2) we prove first that

$$\mathbf{B}_n^{\varepsilon_n}(\Sigma) \subseteq L\Delta \cup \neg L\overline{\Delta} \tag{5.21}$$

holds. If $L\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then ϕ is some ψ_j in ε_n and $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_j$. By (F1) $j > m_1$ and by (C2) $\psi_j \in \Delta$. Hence $L\phi \in L\Delta$. If $\neg L\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then ϕ is some ψ_j in ε_n and $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$. By (F2) $j \leq m_1$ and by (C1) $\psi_j \notin \Delta$. Hence $L\phi \in \neg L\overline{\Delta}$. Thus (5.21) holds.

Now we can show that ε_n is Σ-acceptable. Let $\neg L\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then ϕ is some ψ_i in ε_n and $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. By (F1) $i \leq m_1$ and by (C1) $\psi_i \notin \Delta$. Assume that $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_i$. Since Δ is a stable expansion and $\models_{\mathrm{ae}}$ is monotonic, by (5.21) $\psi_i \in \Delta$, a contradiction. Hence, by Proposition 5.8 ε_n is Σ-acceptable.

By Theorem 5.10 $\Lambda = \{\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$ is a Σ-full set and there is a Σ-acceptable enumeration ε such that $\mathbf{SE}^{\varepsilon}(\Sigma)$ is the unique stable expansion Δ' of Σ such that $\Lambda \subseteq L\Delta' \cup \neg L\overline{\Delta'}$. By (5.21) $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. As Δ is a stable expansion of Σ, $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$. □

The set $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ has to be finite for (C1) and (C2) to be satisfied by an enumeration of sentences. The question whether there exists a Σ-acceptable enumeration ε which induces Δ for an iterative expansion Δ of a set of premises Σ such that $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is infinite remains open.

The enumeration-based method provides an alternative way of defining iterative expansions. Example 5.2 reveals why iterative expansions are not necessarily stable set minimal. It is possible to introspect a sentence ϕ without first introspecting the subformulae of ϕ which can effect the result of the introspection of ϕ. An L-hierarchic expansion of Σ is a stable set minimal iterative expansion which is not dependent on the syntactic representation of the premises. For example, in the propositional case $\{Lp \rightarrow p, \neg Lp \rightarrow p\}$ and $\{p\}$ have both the same unique L-hierarchic expansion.

L-hierarchic expansions are a proper subclass of iterative expansions because, e.g., $\Sigma = \{\neg L\neg Lp \rightarrow p\}$ has a unique L-hierarchic expansion $\mathrm{SE}_{\Sigma}(\{L\neg Lp\})$ although it has two iterative expansions. However, there are sets of premises for which L-hierarchic expansions and iterative expansions coincide. Theorem 5.13 gives such a class of premises. Next we establish another result on premises containing only formulae of the form

$$(L\alpha_1 \wedge \cdots \wedge L\alpha_m \wedge \neg LL\beta_1 \wedge \cdots \wedge \neg LL\beta_n) \rightarrow \gamma \tag{5.22}$$

where $\alpha_i, \beta_i, \gamma \in \mathcal{L}$. These kinds of formulae are called *T-clauses* by Marek and Truszczyński [86] who show that T-clauses can be used to provide an embedding of Reiter's default logic into autoepistemic logic based on iterative expansions. It turns out that an iterative expansion of a set of T-clauses having only finitely many $L\chi$ subformulae is an L-hierarchic expansion.

Theorem 5.16 *Let the underlying logic be a first-order calculus. For each iterative expansion Δ of a set of T-clauses Σ such that $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is finite, there is an L-hierarchic Σ-acceptable enumeration ε such that $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$.*

Proof. We use the same conditions (C1) and (C2) as in the proof of Theorem 5.15 for an enumeration inducing an iterative expansion with two additional requirements.

(C1') *The sentences $\psi_1, \ldots, \psi_{m_1}$ are ordered L-hierarchicly.*

(C2') *For the sentences $\psi_{m_1+1}, \ldots, \psi_n$, it holds that if $\psi_i \in A_k^{\Delta}(\Sigma)$ if and only if $\psi_i \in A_k^{\Delta}(\Sigma)$ for all $k = 0, 1, \ldots$, then ψ_i, ψ_j are ordered L-hierarchicly.*

As in the proof of Theorem 5.15 Conditions (C1) and (C2) guarantee that ε_n is Σ-acceptable and that there is a Σ-acceptable enumeration ε such that $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$. So it remains to be shown that ε_n is L-hierarchic. Then by Theorem 5.10 ε is L-hierarchic.

Assume that ε_n is not L-hierarchic. So there is some $i, j \leq n$ such that $L\psi_j \in \mathrm{Sf}^{\mathrm{L}}(\psi_i)$ but $j > i$. It must be the case that $j > m_1$ because otherwise $i, j \leq m_1$ and then by (C1') ψ_i and ψ_j would be ordered L-hierarchicly. As $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and Σ is a set of T-clauses, it must the case that $\psi_i = L\psi_j$ and $\psi_j \in \mathcal{L}$. As $j > m_1$, then by (C2) $\psi_j \in \Delta$ and thus $\psi_i \in \Delta$ which implies by (C2) that $i > m_1$.

Thus by (C2) and (C2') there is some k such that $\psi_i \in A_k^{\Delta}(\Sigma)$ but $\psi_j \notin A_k^{\Delta}(\Sigma)$. By (5.14) $\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^{\Delta}(\Sigma) \cup \ldots \cup A_{k-1}^{\Delta}(\Sigma)) \models_{\mathrm{ae}} \psi_i$ but $\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^{\Delta}(\Sigma) \cup \ldots \cup A_{k-1}^{\Delta}(\Sigma)) \not\models_{\mathrm{ae}} \psi_j$ and $\psi_j \notin A_0^{\Delta}(\Sigma) \cup \ldots \cup A_{k-1}^{\Delta}(\Sigma)$. So there is an AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \neg L\overline{\Delta} \cup L(A_0^{\Delta}(\Sigma) \cup \ldots \cup A_{k-1}^{\Delta}(\Sigma))$ but $\langle I, T\rangle \not\models_{\mathrm{ae}} \psi_j$. Let $T' = T - \{\psi_j\}$. Then $\langle I, T'\rangle \models_{\mathrm{ae}} \neg L\overline{\Delta} \cup L(A_0^{\Delta}(\Sigma) \cup \ldots \cup A_{k-1}^{\Delta}(\Sigma))$ and $\langle I, T'\rangle \not\models_{\mathrm{ae}} \psi_i$. So there is some $\phi = (L\alpha_1 \wedge \cdots \wedge L\alpha_m \wedge \neg LL\beta_1 \wedge \cdots \wedge \neg LL\beta_n) \rightarrow \gamma \in \Sigma$ such that $\langle I, T'\rangle \not\models_{\mathrm{ae}} \phi$ because otherwise $\Sigma \cup \neg L\overline{\Delta} \cup L(A_0^{\Delta}(\Sigma) \cup \ldots \cup A_{k-1}^{\Delta}(\Sigma)) \not\models_{\mathrm{ae}} \psi_i$. Thus $\langle I, T'\rangle \models_{\mathrm{ae}} (L\alpha_1 \wedge \cdots \wedge L\alpha_m \wedge \neg LL\beta_1 \wedge \cdots \wedge \neg LL\beta_n)$ but $\langle I, T'\rangle \not\models_{\mathrm{ae}} \gamma$. So for all $l = 1, \ldots, m$, $\alpha_l \in T'$ and thus $\alpha_l \in T$ and for all $l = 1, \ldots, n$, $L\beta_l \notin T'$ and thus $L\beta_l \notin T$. As $\langle I, T\rangle \not\models_{\mathrm{ae}} \gamma$, $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$, a contradiction. Thus ε_n is L-hierarchic. $\square$

Chapter 6

On the Decidability and Computational Complexity of Autoepistemic Reasoning

We have not yet presented a definition for the set of autoepistemic conclusions derivable from a set of premises. A stable expansion of the set of premises is an obvious candidate for the set of autoepistemic conclusions. However, we have shown that there are sets of premises with more than one stable expansion. This leads to the concepts of cautious and brave reasoning. We introduce the derivability relations $\mid\!\sim_{\mathrm{c}}$ and $\mid\!\sim_{\mathrm{b}}$ to capture cautious and brave reasoning, respectively.

Definition 6.1 *$\Sigma \mid\!\sim_{\mathrm{c}} \phi$ if and only if $\phi \in \Delta$ for all stable expansions Δ of Σ.*

$\Sigma \mid\!\sim_{\mathrm{b}} \phi$ if and only if there is a stable expansion Δ of Σ such that $\phi \in \Delta$.

The derivability relation $\mid\!\sim_{\mathrm{b}}$ corresponds the notion of arguability defined by McDermott and Doyle for their non-monotonic logic [96] and the derivability relation $\mid\!\sim_{\mathrm{c}}$ equals to McDermott and Doyle's notion of provability [96].

Analogously, we define derivability relations with respect to enumeration-based expansions and L-hierarchic expansions.

Definition 6.2 *$\Sigma \mid\!\sim_{\mathrm{c(E)}} \phi$ if and only if $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma)$ for all Σ-acceptable enumerations ε of $\mathcal{L}_{\mathrm{ae}}$.*

$\Sigma \mid\!\sim_{\mathrm{c(LE)}} \phi$ if and only if $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma)$ for all L-hierarchic Σ-acceptable enumerations ε of $\mathcal{L}_{\mathrm{ae}}$.

$\Sigma \mathrel{|\!\sim}_{\mathrm{b(E)}} \phi$ *if and only if there is a* Σ*-acceptable enumeration* ε *of* $\mathcal{L}_{\mathrm{ae}}$ *such that* $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma)$.

$\Sigma \mathrel{|\!\sim}_{\mathrm{b(LE)}} \phi$ *if and only if there is an L-hierarchic* Σ*-acceptable enumeration* ε *of* $\mathcal{L}_{\mathrm{ae}}$ *such that* $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma)$.

From the discussion above we conclude that the fundamental decision problems in connection with autoepistemic reasoning are cautious and brave derivability as well as the problem of deciding whether a given set of premises has an expansion.

We assume that some underlying logic CL is given and we study the corresponding autoepistemic logic $\mathrm{CL}_{\mathrm{ae}}$ built on top of CL where quantification into the scope of the L operator is not allowed. When studying the decidability and computational complexity of autoepistemic reasoning we restrict ourselves into finite sets of premises. Furthermore, we assume that there is a sentence $\top \in \mathcal{L}$ which is true in every CL-structure.

Definition 6.3 **DERc** *is the set of pairs* $\langle \Sigma, \psi \rangle$ *where* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *is finite and* $\psi \in \mathcal{L}_{\mathrm{ae}}$ *such that* $\Sigma \mathrel{|\!\sim}_{\mathrm{c}} \psi$.

DERb *is the set of pairs* $\langle \Sigma, \psi \rangle$ *where* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *is finite and* $\psi \in \mathcal{L}_{\mathrm{ae}}$ *such that* $\Sigma \mathrel{|\!\sim}_{\mathrm{b}} \psi$.

SE *is the set of finite sets of sentences* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *such that* Σ *has a stable expansion.*

In the same way, the problems **DERc**(E), **DERb**(E), **SE**(E), **DERc**(LE), **DERb**(LE), and **SE**(LE) are defined based on enumerations and L-hierarchic enumerations employing the derivability relations $\mathrel{|\!\sim}_{\mathrm{c(E)}}$, $\mathrel{|\!\sim}_{\mathrm{b(E)}}$, $\mathrel{|\!\sim}_{\mathrm{c(LE)}}$, and $\mathrel{|\!\sim}_{\mathrm{b(LE)}}$. When studying these problems we use some additional concepts defined below.

Definition 6.4 $\overline{\mathbf{DERc}}$ *is the complement of* **DERc**, *i.e. the set of pairs of sentences* $\langle \Sigma, \psi \rangle$ *where* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *is finite and* $\psi \in \mathcal{L}_{\mathrm{ae}}$ *such that* $\Sigma \mathrel{|\!\not\sim}_{\mathrm{c}} \psi$. *Similarly, we define the complements* $\overline{\mathbf{DERc}(\mathrm{E})}$ *and* $\overline{\mathbf{DERc}(\mathrm{LE})}$.

SAT *is the set of sentences* $\phi \in \mathcal{L}$ *such that* ϕ *is satisfiable in* CL, *i.e. there is a* CL*-structure in which* ϕ *is true.*

CN *is the set of pairs* $\langle \Sigma, \psi \rangle$ *where* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *is finite and* $\psi \in \mathcal{L}_{\mathrm{ae}}$ *such that* $\Sigma \models_{\mathrm{ae}} \psi$.

CNL *is the set of pairs* $\langle \Sigma, \psi \rangle$ *where* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *is finite and* $\psi \in \mathcal{L}_{\mathrm{ae}}$ *such that* $\Sigma \models_{\mathrm{L}} \psi$.

FULL *is the set of pairs* $\langle \Sigma, \Lambda \rangle$ *where* $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ *is finite and* $\Lambda \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ *such that* $\Lambda \cup \{\neg L\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma) - \Lambda\}$ *is a* Σ*-full set.*

We study the decidability and complexity of the problems mentioned above by relating them to the decidability and complexity of reasoning in the underlying logic. We use the concepts of polynomial time (many-one) reducibility, polynomial time Turing reducibility, **NP**-hardness and polynomial time hierarchy. Balcázar et al. [3] and Garey and Johnson [28] provide good sources of basic definitions and results. We briefly survey the concepts needed in the complexity analysis. We assume that the reader is familiar with the definitions of a deterministic and nondeterministic Turing machine as well as an oracle Turing machine where the machine is equipped with an oracle set for which membership can be determined in a single step during computation. We consider decision problems, i.e. problems that have only two possible solutions, the answer yes or the answer no. We view a decision problem abstractly as a set corresponding to those instances of the problem that are answered yes.

Given two sets A_1 and A_2, A_1 is said to be *polynomial time (many-one) reducible* to A_2 if and only if there is a function f computable by a deterministic Turing machine in polynomial time such that for all x, $x \in A_1$ if and only if $f(x) \in A_2$. A set A_1 is said to be *polynomial time (nondeterministic) Turing reducible* to A_2 if and only if there is a (nondeterministic) polynomial time oracle machine with an oracle A_2 such that the machine accepts A_1.

P is the class of sets accepted by polynomial time deterministic Turing machines and **NP** is the class of sets accepted by polynomial time nondeterministic Turing machines. A set A is **NP**-*hard* if every set in **NP** is polynomial time reducible to A and A is **NP**-*complete* if A is **NP**-hard and $A \in \mathbf{NP}$. For a class C of sets, **co-**C denotes the class of the complements of the sets in C.

The polynomial time hierarchy is a hierarchy of problems between the complexity classes **P** and **PSPACE** which is the class of sets accepted by deterministic Turing machines whose work space is polynomially bounded. The polynomial time hierarchy is defined using computation with oracles. For a class of sets C, $\mathbf{P}(C)$ denotes the sets accepted by polynomial time deterministic Turing machines using an oracle for a set in C and $\mathbf{NP}(C)$ denotes the sets accepted by polynomial time nondeterministic Turing machines using an oracle for a set in C.

The polynomial time hierarchy is a structure formed by the classes Σ_k^p, Π_k^p, Δ_k^p for each $k \geq 0$ where

1. $\Sigma_0^p = \Pi_0^p = \Delta_0^p = \mathbf{P}$,
2. $\Sigma_{k+1}^p = \mathbf{NP}(\Sigma_k^p)$ for $k \geq 0$,
3. $\Pi_{k+1}^p = \mathbf{co}\text{-}\Sigma_{k+1}^p$ for $k \geq 0$ and

4. $\Delta^p_{k+1} = \mathbf{P}(\Sigma^p_k)$ for $k \geq 0$.

Observe that $\Sigma^p_1 = \mathbf{NP}$, $\Pi^p_1 = \mathbf{co\text{-}NP}$, $\Delta^p_1 = \mathbf{P}$, and $\Delta^p_2 = \mathbf{P}(\mathbf{NP})$. Hardness and completeness for a class in the polynomial time hierarchy is defined using the polynomial time many-one reducibility the same way as **NP**-hardness and completeness.

We aim to show that the decision problems in autoepistemic logic are polynomial time Turing reducible to the underlying consequence relation $\models_{\text{ae}}$ either deterministically or nondeterministically. This can be done by showing that for each of the problems there exists a polynomial time (deterministic or nondeterministic) oracle machine with an oracle for **CN** which accepts the problem in question [3]. This implies that the problems are decidable if the underlying consequence relation $\models_{\text{ae}}$ is decidable.

When relating autoepistemic reasoning to reasoning in the underlying classical logic CL, we need some results on how truth in a CL-structure relates to truth in an AE-structure. The following proposition gives the required relationships which are all based on the fact that in an AE-structure $\langle I, T\rangle$ the truth value of any L free sentence is completely determined by the CL-structure I, i.e. that (2.1) holds for every $\phi \in \mathcal{L}_{\text{ae}}$.

Proposition 6.5 *Let $\phi \in \mathcal{L}$.*

1. *If ϕ is true in every* CL*-structure, then it is true in every* AE*-structure.*
2. *If $\phi \in$ **SAT**, then for every $T \subseteq \mathcal{L}_{\text{ae}}$ there exists a* CL*-structure I such that $\langle I, T\rangle \models_{\text{ae}} \phi$.*
3. *If $\phi \notin$ **SAT**, then for every* AE*-structure N, $N \not\models_{\text{ae}} \phi$.*

Proposition 6.6 **SE** *is polynomial time (many-one) reducible to* **DERb**, **SE**(E) *is polynomial time reducible to* **DERb**(E)*, and* **SE**(LE) *is polynomial time reducible to* **DERb**(LE)*.*

Proof. To prove that **SE** is polynomial time reducible to **DERb** it is sufficient to show that $\Sigma \in \mathbf{SE}$ if and only if $\langle \Sigma, \top\rangle \in \mathbf{DERb}$. If $\Sigma \in \mathbf{SE}$, then there exists a stable expansion Δ of Σ which also contains $\top$ because $\Sigma \cup L\Delta \cup \neg L\overline{\Delta} \models_{\text{ae}} \top$ by Proposition 6.5. Thus $\langle \Sigma, \top\rangle \in \mathbf{DERb}$. Also $\langle \Sigma, \top\rangle \in \mathbf{DERb}$ implies that Σ has a stable expansion. Other reducibilities can be proved similarly. □

Lemma 6.7 **CNL** *is polynomial time Turing reducible to* **CN**.

FUNCTION $\mathbf{cnl}(\Sigma, \psi)$
INPUT: A finite set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ and a sentence $\psi \in \mathcal{L}_{\mathrm{ae}}$.

```
Λ' := ∅;
FOR every Lχ ∈ Sf^qL(ψ) DO
    IF cnl(Σ, χ) = true THEN
        Λ' := Λ' ∪ {Lχ}
    ELSE
        Λ' := Λ' ∪ {¬Lχ};
IF ⟨Σ ∪ Λ', ψ⟩ ∈ CN THEN
    return true
ELSE
    return false.
```

Figure 6.1: The function **cnl** using an oracle for **CN**.

Proof. Figure 6.1 describes a recursive function **cnl**, which uses an oracle for **CN**. This function gives the required polynomial time Turing machine which reduces **CNL** to **CN**. It is straightforward to show by induction on $\mathrm{BD}(\psi)$ that $\mathbf{cnl}(\Sigma, \psi)$ returns true if and only if $\Sigma \models_{\mathrm{L}} \psi$ by observing that Λ' plays the role of $\mathrm{SB}_{\Sigma}(\psi)$. Thus a Turing machine performing the computation of the function $\mathbf{cnl}(\Sigma, \psi)$ and then accepting the input if and only if the function returns true accepts the set **CNL**. This oracle machine is deterministic and performs its computation in polynomial number of steps. □

Lemma 6.8 **FULL** *is polynomial time Turing reducible to* **CN**.

Proof. We show that **FULL** is polynomial time Turing reducible to **CNL** which implies the lemma by Lemma 6.7. An oracle machine with an oracle for **CNL** is described in Figure 6.2. The machine accepts **FULL** which can be shown in the following way. If the machine accepts the input $\langle \Sigma, \Lambda \rangle$, then for every $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)$, either $L\chi \in \Lambda'$ and $\Sigma \cup \Lambda' \models_{\mathrm{L}} \chi$ or $\neg L\chi \in \Lambda'$ and $\Sigma \cup \Lambda' \not\models_{\mathrm{L}} \chi$. Thus Λ' is Σ-full and $\langle \Sigma, \Lambda \rangle \in \mathbf{FULL}$. If the machine does not accept $\langle \Sigma, \Lambda \rangle$, then there exists $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)$ such that either $L\chi \in \Lambda'$ and $\Sigma \cup \Lambda' \not\models_{\mathrm{L}} \chi$ or $\neg L\chi \in \Lambda'$ and $\Sigma \cup \Lambda' \models_{\mathrm{L}} \chi$. Thus Λ' is not Σ-full and $\langle \Sigma, \Lambda \rangle \notin \mathbf{FULL}$.

Every computation of the machine takes a polynomial number of steps with respect to the length of the input because the main loop is executed at most once for each $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)$ and Σ has only a linear number of $L\chi$ quasi-subformulae with respect to its length. □

Lemma 6.9 **SE**, $\overline{\mathbf{DERc}}$, *and* **DERb** *are polynomial time nondeterministic Turing reducible to* **CN**.

INPUT: A finite set of sentences $\Sigma \in \mathcal{L}_{\mathrm{ae}}$ and a set $\Lambda \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$.
 $\Lambda' := \Lambda \cup \{\neg L\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma) - \Lambda\}$;
 FOR every $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ **DO**
  IF $((L\chi \in \Lambda)$ and $(\langle \Sigma \cup \Lambda', \chi\rangle \in \mathbf{CNL}))$ or
   $((L\chi \notin \Lambda)$ and $(\langle \Sigma \cup \Lambda', \chi\rangle \notin \mathbf{CNL}))$ **THEN**
   skip
  ELSE
   halt in a rejecting state;
 Accept.

Figure 6.2: An oracle machine accepting **FULL**.

Proof. The nondeterministic oracle machine accepting $\overline{\mathbf{DERc}}$ is given in Figure 6.3. As the tests of membership in **FULL** and **CNL** can be accomplished in polynomial time using a deterministic oracle machine with an oracle for **CN** by Lemmata 6.8 and 6.7, the machine accomplishes every computation in polynomial time. If the machine in Figure 6.3 accepts its input, then

$$\Lambda' = \Lambda \cup \{\neg L\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma) - \Lambda\} \tag{6.1}$$

is a Σ-full set and $\Sigma \cup \Lambda' \not\models_{\mathrm{L}} \psi$ holds. By Corollary 3.11 there is a stable expansion $\mathrm{SE}_\Sigma(\Lambda')$ of Σ not containing ψ. Thus $\langle \Sigma, \psi\rangle \in \overline{\mathbf{DERc}}$. On the other hand, if $\langle \Sigma, \psi\rangle$ belongs to $\overline{\mathbf{DERc}}$, there is a stable expansion Δ of Σ not containing ψ and by Lemma 3.5 there is a Σ-full set $\Lambda' \subseteq L\Delta \cup \neg L\overline{\Delta}$. Thus $\langle \Sigma, \Lambda\rangle \in \mathbf{FULL}$ where $\Lambda = \{L\chi \mid L\chi \in \Lambda'\}$. By Corollary 3.11 $\Delta = \mathrm{SE}_\Sigma(\Lambda')$ and $\Sigma \cup \Lambda' \not\models_{\mathrm{L}} \psi$. So there is an accepting computation on the input. Thus the machine in Figure 6.3 accepts $\overline{\mathbf{DERc}}$. Similarly, it can be shown that a machine which behaves as the machine in Figure 6.3 but accepts the input if $\langle \Sigma \cup \Lambda', \psi\rangle \in \mathbf{CNL}$ and otherwise rejects the input accepts the set **DERb**. As **DERb** is polynomial time nondeterministic Turing reducible to **CN**, Proposition 6.6 implies that **SE** is polynomial time nondeterministic Turing reducible to **CN**. □

We use the function **introspect** given in Figure 6.4 when showing that the problems related to enumeration-based expansions are reducible to the underlying consequence relation $\models_{\mathrm{ae}}$.

Proposition 6.10 *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ be a finite set of sentences and ε a finite enumeration of sentences in $\mathcal{L}_{\mathrm{ae}}$. Then* **introspect**$(\Sigma, \varepsilon)$ *returns the set* $\mathbf{B}^\varepsilon(\Sigma)$ *using one query from the oracle for* **CN** *for each sentence in ε.*

Lemma 6.11 *The problems* **SE**(E), $\overline{\mathbf{DERc}(\mathrm{E})}$, *and* **DERb**(E) *as well as* **SE**(LE), $\overline{\mathbf{DERc}(\mathrm{LE})}$, *and* **DERb**(LE) *are polynomial time nondeterministic Turing reducible to* **CN**.

INPUT: A finite set of sentences $\Sigma \subseteq \mathcal{L}_{\rm ae}$ and a sentence $\psi \in \mathcal{L}_{\rm ae}$.
  Choose nondeterministically a set $\Lambda \subseteq \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)$;
  IF $\langle \Sigma, \Lambda \rangle \notin$ **FULL THEN**
    halt in a rejecting state;
  $\Lambda' := \Lambda \cup \{\neg L\chi \mid L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma) - \Lambda\}$;
  IF $\langle \Sigma \cup \Lambda', \psi \rangle \notin$ **CNL THEN**
    accept
  ELSE
    halt in a rejecting state.

Figure 6.3: A nondeterministic oracle machine accepting $\overline{\textbf{DERc}}$.

FUNCTION introspect(Σ, ε)
INPUT: A finite set of sentences $\Sigma \subseteq \mathcal{L}_{\rm ae}$ and
    a finite enumeration ε of sentences in $\mathcal{L}_{\rm ae}$.
  $B := \emptyset$;
  WHILE ε is not empty **DO BEGIN**
    Remove the first formula χ from ε;
    IF $\langle \Sigma \cup B, \chi \rangle \in$ **CN THEN**
      $B := B \cup \{L\chi\}$
    ELSE
      $B := B \cup \{\neg L\chi\}$
  END;
  Return B.

Figure 6.4: The function **introspect** which returns $\mathbf{B}^{\varepsilon}(\Sigma)$ using an oracle for **CN**.

Proof. A nondeterministic oracle machine accepting **DERb**(LE) is described in Figure 6.5. As the tests for membership in **CNL** can be performed in polynomial time using a deterministic oracle machine with an oracle for **CN** by Lemma 6.7, the machine accomplishes every computation in polynomial time. If the machine in Figure 6.5 accepts its input, then there is an L-hierarchic enumeration ε_n of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. By Proposition 6.10 $B = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and by Proposition 5.8 ε_n is Σ-acceptable. Then by Theorem 5.10

$$\Lambda = \{\phi \in B \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\} \tag{6.2}$$

is a Σ-full set and as ε_n is L-hierarchic, there is an L-hierarchic Σ-acceptable enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ such that

$$\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(B). \tag{6.3}$$

As $\Sigma \cup B \models_{\mathrm{L}} \psi$ holds, $\psi \in \mathrm{SE}_{\Sigma}(B) = \mathbf{SE}^{\varepsilon}(\Sigma)$ and thus $\langle \Sigma, \psi \rangle \in$ **DERb**(LE).

On the other hand, if $\langle \Sigma, \psi \rangle \in$ **DERb**(LE), there is an L-hierarchic Σ-acceptable enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ such that $\psi \in \mathbf{SE}^{\varepsilon}(\Sigma)$. By Theorem 5.11 there is an L-hierarchic Σ-acceptable enumeration ε_n of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $\Lambda = \{\phi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma) \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$ is the unique Σ-full set such that $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\Lambda) = \mathrm{SE}_{\Sigma}(\mathbf{B}_n^{\varepsilon_n}(\Sigma))$. Thus $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{L}} \psi$ holds. So there is an accepting computation on the input. Thus the machine in Figure 6.5 accepts **DERb**(LE). Similarly, it can be shown that a machine which behaves as the machine in Figure 6.5 but accepts the input if $\langle \Sigma \cup \Lambda', \psi \rangle \notin$ **CNL** and otherwise rejects the input accepts the set $\overline{\mathbf{DERc}(\mathrm{LE})}$.

Machines that accept **DERb**(E) and $\overline{\mathbf{DERc}(\mathrm{E})}$ can be constructed by removing the test for L-hierarchy. As **DERb**(LE) (**DERb**(E)) is polynomial time nondeterministic Turing reducible to **CN**, Proposition 6.6 implies that **SE**(LE) (**SE**(E)) is polynomial time nondeterministic Turing reducible to **CN**. □

The following two theorems show that the decidability of the underlying consequence relation $\models_{\mathrm{ae}}$ implies the decidability of autoepistemic reasoning.

Theorem 6.12 *If* **CN** *is decidable, then each of the problems* **FULL**, **CNL**, **DERc**, **DERb**, *and* **SE** *is decidable.*

Proof. Let **CN** be decidable. By Lemmata 6.8 and 6.7 **FULL** and **CNL** are decidable. We prove that **DERc** and **DERb** are decidable by modifying the nondeterministic oracle machine in Figure 6.3. A combined decision procedure for both problems is described in Figure 6.6. As the set $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ is finite, it has finitely many subsets. Thus the main loop is executed only a finite number of times and the procedure always halts after a finite number of steps. It can be shown that B-answer = TRUE when the procedure halts if and only if $\Sigma \mathrel{|\!\sim_{\mathrm{b}}} \psi$.

INPUT: A finite set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ and a sentence $\psi \in \mathcal{L}_{\mathrm{ae}}$.
  Choose nondeterministically an enumeration ε_n of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$;
  IF ε_n is not L-hierarchic **THEN**
    halt in a rejecting state;
  $B :=$ **introspect**(Σ, ε_n);
  IF for all $\neg L\chi \in B, \langle \Sigma \cup B, \chi\rangle \notin$ **CN THEN**
    IF $\langle \Sigma \cup B, \psi\rangle \in$ **CNL THEN**
      accept
    ELSE
      halt in a rejecting state
  ELSE
    halt in a rejecting state.

Figure 6.5: A nondeterministic oracle machine accepting **DERb**(LE).

INPUT: A finite set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ and a sentence $\psi \in \mathcal{L}_{\mathrm{ae}}$.
  C-answer:= TRUE; B-answer:= FALSE;
  FOR every set $\Lambda \subseteq \mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ **DO**
    IF $\langle \Sigma, \Lambda\rangle \in$ **FULL THEN**
      BEGIN
        $\Lambda' := \Lambda \cup \{\neg L\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma) - \Lambda\}$;
        IF $\langle \Sigma \cup \Lambda', \psi\rangle \in$ **CNL THEN**
          B-answer:= TRUE
        ELSE
          C-answer:= FALSE
      END.

Figure 6.6: A decision procedure for **DERb** and **DERc**.

Similarly, it can be shown that C-answer = TRUE when the procedure halts if and only if $\Sigma \mid\!\sim_{\mathrm{c}} \psi$. As **DERb** is decidable, Proposition 6.6 implies that **SE** is decidable. □

Theorem 6.13 *If* **CN** *is decidable, then each of the problems* **DERc**(E), **DERb**(E), *and* **SE**(E) *as well as* **DERc**(LE), **DERb**(LE), *and* **SE**(LE) *is decidable.*

Proof. Let **CN** be decidable. We prove that **DERc**(LE) and **DERb**(LE) are decidable by modifying the nondeterministic oracle machine in Figure 6.5. A combined decision procedure for both problems is described in Figure 6.7. As the set $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is finite, the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ has only finitely many enumerations. Thus the main loop is executed only a finite number of times

INPUT: A finite set of sentences $\Sigma \subseteq \mathcal{L}_{\text{ae}}$ and a sentence $\psi \in \mathcal{L}_{\text{ae}}$

```
C-answer:= TRUE; B-answer:= FALSE;
FOR every enumeration ε_n of {φ | Lφ ∈ Sf^L(Σ)} DO
    IF  ε_n is L-hierarchic THEN
        BEGIN
          B := introspect(Σ, ε_n);
          IF for all ¬Lχ ∈ B, ⟨Σ ∪ B, χ⟩ ∉ CN THEN
              IF ⟨Σ ∪ B, ψ⟩ ∈ CNL THEN
                  B-answer:= TRUE
              ELSE
                  C-answer:= FALSE
        END.
```

Figure 6.7: A decision procedure for **DERb**(LE) and **DERc**(LE).

and the procedure always halts after a finite number of steps. It can be shown that B-answer = TRUE when the procedure halts if and only if $\Sigma \mid\!\sim_{\text{b(LE)}} \psi$. Similarly, it can be shown that C-answer = TRUE when the procedure halts if and only if $\Sigma \mid\!\sim_{\text{c(LE)}} \psi$. As **DERb**(LE) is decidable, Proposition 6.6 implies that **SE**(LE) is decidable. The decidability of **DERc**(E), **DERb**(E), and **SE**(E) can be shown similarly. □

Next we establish that satisfiability in CL is polynomial time reducible to the decision problems defined above. This shows that the problems are at least as hard as satisfiability in CL, e.g. an efficient (polynomial time) decision procedure for any of the problems implies an efficient decision procedure for satisfiability in CL. We present a series of lemmata on reducing **SAT** to the other problems.

Lemma 6.14 *For a sentence $\phi \in \mathcal{L}$, $\phi \in$ **SAT** if and only if $\{L\neg L\neg\phi\} \in$ **SE** if and only if $\{L\neg L\neg\phi\} \in$ **SE**(E) if and only if $\{L\neg L\neg\phi\} \in$ **SE**(LE).*

Proof. Let $\phi \in$ **SAT**. We denote the set $\{L\neg L\neg\phi\}$ by Σ. Then $\text{Sf}^{\text{L}}(\Sigma) = \{L\neg L\neg\phi, L\neg\phi\}$. We show that $\varepsilon_2 = \neg\phi, \neg L\neg\phi$ is an L-hierarchic Σ-acceptable enumeration of $\{\phi \mid L\phi \in \text{Sf}^{\text{L}}(\Sigma)\}$. Clearly, ε_2 is L-hierarchic and $\mathbf{B}_2^{\varepsilon_2}(\Sigma) = \{\neg L\neg\phi, L\neg L\neg\phi\}$. Thus $\Sigma \cup \mathbf{B}_2^{\varepsilon_2}(\Sigma) \not\models_{\text{ae}} \neg\phi$ which implies by Proposition 5.8 that ε_2 is Σ-acceptable. By Theorem 5.10 there is an L-hierarchic expansion of Σ. Hence $\Sigma \in \mathbf{SE}(\text{LE}) \subseteq \mathbf{SE}(\text{E}) \subseteq \mathbf{SE}$.

Secondly, let $\phi \notin$ **SAT**. Assume that there is a Σ-full set Λ. Then either $\Lambda = \{\neg L\neg L\neg\phi\}$ or $\Lambda = \{L\neg L\neg\phi\}$. Assume that $\Lambda = \{L\neg L\neg\phi\}$. As $\phi \notin$ **SAT**, then by Proposition 6.5 $\Sigma \cup \Lambda \models_{\text{ae}} \neg\phi$. Thus $\text{SB}_{\Sigma\cup\Lambda}(\neg L\neg\phi) = \{L\neg\phi\}$. Let $T = \{\neg\phi, \neg L\neg\phi\}$. Then for any CL-structure I, $\langle I, T\rangle \models_{\text{ae}} \Sigma \cup \Lambda \cup$

$\mathrm{SB}_{\Sigma\cup\Lambda}(\neg L\neg\phi)$, but $\langle I,T\rangle \not\models_{\mathrm{ae}} \neg L\neg\phi$. Hence $\Sigma\cup\Lambda \not\models_{\mathrm{L}} \neg L\neg\phi$, a contradiction as then Λ is not a Σ-full set. Thus $\Lambda = \{\neg L\neg L\neg\phi\}$. Then $\Sigma\cup\Lambda \models_{\mathrm{ae}} \neg L\neg\phi$, as there is no AE-structure $\langle I,T\rangle$ such that $\langle I,T\rangle \models_{\mathrm{ae}} \Sigma\cup\{\neg L\neg L\neg\phi\}$. This is a contradiction as then $\Sigma\cup\Lambda \models_{\mathrm{L}} \neg L\neg\phi$ which implies that Λ is not a Σ-full set. Hence there is no Σ-full set and thus by Theorem 3.8 $\Sigma \notin \mathbf{SE}$ which implies by Theorem 5.4 $\Sigma \notin \mathbf{SE}(\mathrm{E})$ and $\Sigma \notin \mathbf{SE}(\mathrm{LE})$. □

Lemma 6.15 *For a sentence $\phi \in \mathcal{L}$, $\phi \in$ **SAT** if and only if $\langle\emptyset,\neg L\neg\phi\rangle \in$ **DERc** if and only if $\langle\emptyset,\neg L\neg\phi\rangle \in$ **DERc**(E) if and only if $\langle\emptyset,\neg L\neg\phi\rangle \in$ **DERc**(LE).*

Proof. By Corollary 3.15 $\emptyset$ has exactly one stable expansion Δ. By Theorem 5.10 there is a corresponding L-hierarchic enumeration ε of $\mathcal{L}_{\mathrm{ae}}$ such that $\Delta = \mathbf{SE}^{\varepsilon}(\Sigma)$ and by Theorem 5.4 this is the unique enumeration-based expansion of $\emptyset$. So

$$\begin{array}{lll} \neg L\neg\phi \in \Delta & \text{if and only if} & \langle\emptyset,\neg L\neg\phi\rangle \in \mathbf{DERc} \\ & \text{if and only if} & \langle\emptyset,\neg L\neg\phi\rangle \in \mathbf{DERc}(\mathrm{E}) \\ & \text{if and only if} & \langle\emptyset,\neg L\neg\phi\rangle \in \mathbf{DERc}(\mathrm{LE}) \end{array}$$

and it remains to be shown that $\phi \in \mathbf{SAT}$ if and only if $\neg L\neg\phi \in \Delta$.

Let $\phi \in \mathbf{SAT}$. Then by Proposition 6.5 there is a CL-structure I such that $\langle I,\Delta\rangle \models_{\mathrm{ae}} \phi$. But then also $\langle I,\Delta\rangle \models_{\mathrm{ae}} L\Delta \cup \neg L\overline{\Delta}$ holds. This implies that $\emptyset \cup L\Delta \cup \neg L\overline{\Delta} \not\models_{\mathrm{ae}} \neg\phi$. Thus $\neg\phi \notin \Delta$ and $\neg L\neg\phi \in \neg L\overline{\Delta} \subseteq \Delta$. Let $\phi \notin \mathbf{SAT}$. Then by Proposition 6.5 $\neg\phi$ is true in every AE-structure. Thus $\emptyset\cup L\Delta\cup\neg L\overline{\Delta} \models_{\mathrm{ae}} \neg\phi$ which implies that $\neg\phi \in \Delta$. Then for every CL-structure I, $\langle I,\Delta\rangle \models_{\mathrm{ae}} \emptyset \cup L\Delta \cup \neg L\overline{\Delta}$ but $\langle I,\Delta\rangle \not\models_{\mathrm{ae}} \neg L\neg\phi$. Thus $\emptyset \cup L\Delta \cup \neg L\overline{\Delta} \not\models_{\mathrm{ae}} \neg L\neg\phi$ and $\neg L\neg\phi \notin \Delta$. □

Lemma 6.16 *For a sentence $\phi \in \mathcal{L}$, $\phi \in$ **SAT** if and only if $\langle\emptyset,\neg L\neg\phi\rangle \in$ **CNL**.*

Proof. Let $\phi \in \mathbf{SAT}$. Then by Proposition 6.5 there is an AE-structure in which $\neg\phi$ is not true. Thus $\emptyset \not\models_{\mathrm{ae}} \neg\phi$ and so $\mathrm{SB}_{\emptyset}(\neg L\neg\phi) = \{\neg L\neg\phi\}$ and $\emptyset \models_{\mathrm{L}} \neg L\neg\phi$. Let $\phi \notin \mathbf{SAT}$. Then $\emptyset \models_{\mathrm{ae}} \neg\phi$. Thus $\mathrm{SB}_{\emptyset}(\neg L\neg\phi) = \{L\neg\phi\}$ and $\emptyset \cup \mathrm{SB}_{\emptyset}(\neg L\neg\phi) \not\models_{\mathrm{ae}} \neg L\neg\phi$. Hence $\emptyset \not\models_{\mathrm{L}} \neg L\neg\phi$. □

Lemma 6.17 *For a sentence $\phi \in \mathcal{L}$, $\phi \in$ **SAT** if and only if $\langle\{\neg L\neg\phi\},\emptyset\rangle \in$ **FULL**.*

Proof. Let $\phi \in \mathbf{SAT}$. Then by Proposition 6.5 $\{\neg L\neg\phi\} \not\models_{\mathrm{ae}} \neg\phi$. Thus $\{\neg L\neg\phi\} \not\models_{\mathrm{L}} \neg\phi$ and $\langle\{\neg L\neg\phi\},\emptyset\rangle \in \mathbf{FULL}$. On the other hand, let $\phi \notin \mathbf{SAT}$. Then $\{\neg L\neg\phi\} \models_{\mathrm{ae}} \neg\phi$ which is implies that $\{\neg L\neg\phi\} \models_{\mathrm{L}} \neg\phi$ and thus $\langle\{\neg L\neg\phi\},\emptyset\rangle \notin \mathbf{FULL}$. □

Proposition 6.6 and Lemmata 6.14–6.17 prove the following theorem.

Theorem 6.18 **SAT** *is polynomial time reducible to* **DERc**, **DERb**, **SE**, **DERc**(E), **DERb**(E), **SE**(E), **DERc**(LE), **DERb**(LE), **SE**(LE), **FULL**, *and to* **CNL**.

Corollary 6.19 *Let the underlying logic* CL *be a full first-order calculus. Then* **DERc**, **DERb**, **SE**, **DERc**(E), **DERb**(E), **SE**(E), **DERc**(LE), **DERb**(LE), **SE**(LE), **FULL**, *and* **CNL** *are not recursively enumerable.*

The corollary means that the decision problems of an autoepistemic logic based on a full first-order calculus are not even semi-decidable. A first-order calculus is semi-decidable, i.e. the set of theorems (valid sentences) is recursively enumerable. Thus the autoepistemic layer turns the logic from semi-decidable to undecidable. This is to be expected as there is an implicit appeal to consistency in the definition of stable expansions (2.6). The theorem implies also that in the propositional case the decision problems are **NP**-hard. This is quite anticipated also in the light of Marek and Truszczyński's [87] result showing that in a subset of propositional autoepistemic logic the **SE** problem is **NP**-complete.

Corollary 6.20 *If the underlying logic* CL *is the classical propositional calculus, then* **DERc**, **DERb**, **SE**, **DERc**(E), **DERb**(E), **SE**(E), **DERc**(LE), **DERb**(LE), **SE**(LE), **FULL**, *and* **CNL** *are all* **NP**-*hard.*

We would like to show that autoepistemic reasoning is decidable exactly when the underlying consequence relation $\models_{\mathrm{ae}}$ is decidable. One direction is implied by Theorems 6.12 and 6.13. By Theorem 6.18 to prove the other direction it would be enough to reduce the underlying consequence relation $\models_{\mathrm{ae}}$ to satisfiability in CL. But this result requires the underlying logic CL to satisfy an additional condition. There should be formulae in CL corresponding to $L\phi$ formulae which behave essentially like atomic formulae in the propositional calculus, i.e. have the same truth value in a CL-structure regardless of the context in which they appear. This requirement is satisfied by any first-order calculus but, e.g., in the case of modal logics a detailed analysis is needed.

Lemma 6.21 *If the underlying logic* CL *is a first-order calculus, then* **CN** *is polynomial time Turing reducible to* **SAT**.

Proof. By the definition of the consequence relation $\models_{\mathrm{ae}}$, $\Sigma \models_{\mathrm{ae}} \psi$ holds if and only if there is no AE-structure in which $(\Sigma \cup \{\neg\psi\})$ is true. Let us denote the conjunction of the elements in $\Sigma \cup \{\neg\psi\}$ by ϕ. The satisfiability in autoepistemic logic can be reduced to satisfiability in CL by replacing every formula in $\mathrm{Sf}^{\mathrm{qL}}(\phi)$ by a new atomic sentence not appearing in ϕ. We denote this sentence by $\widehat{\phi}$. Now we show that for every $\phi \in \mathcal{L}_{\mathrm{ae}}$, ϕ is AE-satisfiable if and only if $\widehat{\phi}$ is CL-satisfiable. Let ϕ be AE-satisfiable. So there is a AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \phi$. Let I' be a CL-structure which is the same as I except that for every new atomic sentence χ which replaced a quasi-atomic sentence $L\chi'$ in ϕ, $I' \models_{\mathrm{cl}} \chi$ if and only if $\chi' \in T$. It can be shown that $I' \models_{\mathrm{cl}} \widehat{\phi}$ holds. On the other hand, let $\widehat{\phi}$ be CL-satisfiable. So there is a CL-structure I such that $I \models_{\mathrm{cl}} \widehat{\phi}$. Let T be a set of sentences such that $\chi' \in T$ if and only if $I \models_{\mathrm{cl}} \chi$ for every quasi-atomic sentence $L\chi'$ which was replaced by a new atomic sentence χ. It is straightforward to show that $\langle I, T\rangle \models_{\mathrm{ae}} \phi$.

Let TM be a deterministic oracle machine which first asks the oracle **SAT** whether $\widehat{\phi} \in \mathbf{SAT}$ holds where ϕ is the conjunction of the elements of the set $\Sigma \cup \{\neg\psi\}$ and accepts if this is not the case and otherwise rejects. Then TM accepts **CN** in polynomial time. □

Now it is quite easy to show that in the first-order case all the problems discussed above are decidable exactly when the underlying consequence relation is decidable.

Theorem 6.22 *If the underlying logic* CL *is a first-order calculus, then for each of the problems* **DERc**, **DERb**, **SE**, **DERc**(E), **DERb**(E), **SE**(E), **DERc**(LE), **DERb**(LE), **SE**(LE), **FULL**, *and* **CNL** *holds that the problem is decidable if and only if* **CN** *is decidable.*

Proof. ($\Rightarrow$) By Lemma 6.21 **CN** is polynomial time Turing reducible to **SAT** which is polynomial time reducible to each of the problems by Theorem 6.18. Thus the decision procedure for each of the problems can be used as a decision procedure for **CN** as follows. For example, assume that **SE** is decidable. We are interested in deciding whether $\langle \Sigma, \psi\rangle \in \mathbf{CN}$ holds. First apply the reduction of Lemma 6.21 to $\langle \Sigma, \psi\rangle$ and obtain $\widehat{\phi}$ where ϕ is the conjunction of the elements in $\Sigma \cup \{\neg\psi\}$. Then apply the reduction of Lemma 6.14 to $\widehat{\phi}$ and finally use the decision procedure for **SE** to decide whether $\{L\neg L\neg\widehat{\phi}\} \in \mathbf{SE}$. By Lemmata 6.21 and 6.14 $\langle \Sigma, \psi\rangle \in \mathbf{CN}$ if and only if $\{L\neg L\neg\widehat{\phi}\} \notin \mathbf{SE}$.

($\Leftarrow$) The other direction is implied by Theorems 6.12 and 6.13. □

The following theorem is a direct consequence of Lemmata 6.21, 6.7–6.9 and 6.11.

Theorem 6.23 *If the underlying logic is a first-order calculus, then* **SE**, $\overline{\mathbf{DERc}}$, **DERb**, **SE**(E), $\overline{\mathbf{DERc}(\mathrm{E})}$, **DERb**(E), **SE**(LE), $\overline{\mathbf{DERc}(\mathrm{LE})}$, *and* **DERb**(LE) *are polynomial time nondeterministic Turing reducible to* **SAT**. *Furthermore,* **FULL** *and* **CNL** *are polynomial time (deterministic) Turing reducible to* **SAT**.

As an immediate corollary of Theorem 6.23 we can relate the problems to the polynomial time hierarchy (cf. [3]) as follows:

Corollary 6.24 *If the underlying logic* CL *is the propositional calculus, then*

1. $\mathbf{DERc} \in \Pi_2^p$ $\quad \mathbf{DERc}(\mathrm{E}) \in \Pi_2^p$ $\quad \mathbf{DERc}(\mathrm{LE}) \in \Pi_2^p$
2. $\mathbf{DERb} \in \Sigma_2^p$ $\quad \mathbf{DERb}(\mathrm{E}) \in \Sigma_2^p$ $\quad \mathbf{DERb}(\mathrm{LE}) \in \Sigma_2^p$
3. $\mathbf{SE} \in \Sigma_2^p$ $\quad \mathbf{SE}(\mathrm{E}) \in \Sigma_2^p$ $\quad \mathbf{SE}(\mathrm{LE}) \in \Sigma_2^p$
4. $\mathbf{CNL} \in \Delta_2^p$
5. $\mathbf{FULL} \in \Delta_2^p$

Some results on Moore style stable expansions in Corollary 6.24 ($\mathbf{DERc} \in \Pi_2^p$, $\mathbf{DERb} \in \Sigma_2^p$, $\mathbf{SE} \in \Sigma_2^p$, $\mathbf{CNL} \in \Delta_2^p$, and $\mathbf{FULL} \in \Delta_2^p$) have appeared previously [105, 107]. Gottlob [39, 40] extends these results. He shows that **DERc** is Π_2^p-complete and **DERb** and **SE** are Σ_2^p-complete. He also establishes that corresponding decision problems related to **N**-expansions are on the second level of the polynomial time hierarchy. We show in Chapter 5 that, in the propositional case, iterative expansions and enumeration-based expansions coincide. Marek and Truszczyński [87] note that iterative expansions are the same as **N**-expansions. This implies immediately $\mathbf{DERc}(\mathrm{E}) \in \Pi_2^p$, $\mathbf{DERb}(\mathrm{E}) \in \Sigma_2^p$, and $\mathbf{SE}(\mathrm{E}) \in \Sigma_2^p$. Gottlob [40] and Stillman [139] have independently established that similar decision problems in Reiter's default logic are Π_2^p and Σ_2^p-complete. Default logic can be seen as a special case of autoepistemic logic based on **N**-expansions [86]. Gottlob [40] uses this result in establishing that the decision problems related to **N**-expansions are complete problems with respect to the second level of the polynomial time hierarchy. As **N**-expansions and enumeration-based expansions coincide, this implies directly that **DERc**(E) is Π_2^p-complete and **DERb**(E) as well as **SE**(E) are Σ_2^p-complete. Default logic can also be seen as a special case of autoepistemic logic based on L-hierarchic expansions as will be discussed in Chapter 7. This together with the completeness results for default logic shows that **DERc**(LE) is Π_2^p-hard and **DERb**(LE) and **SE**(LE) are Σ_2^p-hard. But then by the results of Corollary 6.24 ($\mathbf{DERc}(\mathrm{LE}) \in \Pi_2^p$, $\mathbf{DERb}(\mathrm{LE}) \in \Sigma_2^p$, $\mathbf{SE}(\mathrm{LE}) \in \Sigma_2^p$) **DERc**(LE) is Π_2^p-complete and **DERb**(LE) and **SE**(LE) are Σ_2^p-complete.

Hence completeness results can be established for the decision problems in autoepistemic logic based on Moore style, enumeration-based, and L-hierarchic expansions. However, for enumeration-based and L-hierarchic expansions the results are obtained indirectly through default logic. Here we want to offer a more direct and uniform proof. We start by showing that the problem of deciding whether there exists an expansion is Σ_2^p-complete. To establish Σ_2^p-hardness, we employ a well known Σ_2^p-complete problem, the validity problem of a subclass of quantified Boolean formulae [28]. The same problem is used by Gottlob [40] and Stillman [139] to establish corresponding hardness results for Moore's autoepistemic logic and default logic. The mapping from quantified Boolean formulae to sets of autoepistemic formulae used in the proof is somewhat different from that used by Gottlob [40]. Our mapping enables a uniform Σ_2^p-hardness proof for the expansion existence problem with respect to Moore style expansions, enumeration-based expansions, and L-hierarchic expansions.

Theorem 6.25 *If the underlying logic* CL *is the propositional calculus, then* **SE**, **SE**(E), *and* **SE**(LE) *are* Σ_2^p*-complete.*

Proof. We show first that **SE** is Σ_2^p-complete. As $\mathbf{SE} \in \Sigma_2^p$, it remains to be shown that **SE** is Σ_2^p-hard which can be proved by showing that some Σ_2^p-complete problem is polynomial time reducible to **SE**. The problem $\mathbf{QBF}^2$ of deciding the validity of quantified Boolean formulae of a particular restricted type is a suitable Σ_2^p-complete problem [28].

$\mathbf{QBF}^2$ is the set of valid quantified Boolean formulae of the form

$$\exists p_1 \ldots \exists p_n \forall q_1 \ldots \forall q_m E \tag{6.4}$$

where $n, m \geq 0$ and E is a well formed Boolean expression over the set of variables $\{p_1, \ldots, p_n,\ q_1, \ldots, q_m\}$. Boolean expressions are built in the usual way from variables, constants $(0, 1)$, negation $(\neg)$, conjunction $(\wedge)$, and disjunction $(\vee)$. A Boolean assignment is a mapping from the set of variables to truth values $\{T, F\}$. An assignment V can be extended to quantified Boolean formulae in the usual way.

1. $V(0) = F, V(1) = T$.
2. $V(\neg E) = T$ if $V(E) = F$ otherwise $V(\neg E) = F$.
3. $V(E' \wedge E'') = T$ if $V(E') = V(E'') = T$ otherwise $V(E' \wedge E'') = F$.
4. $V(E' \vee E'') = F$ if $V(E') = V(E'') = F$ otherwise $V(E' \vee E'') = T$.
5. $V(\exists p E) = V(E\mid_{p \leftarrow 0} \vee E\mid_{p \leftarrow 1})$.
6. $V(\forall p E) = V(E\mid_{p \leftarrow 0} \wedge E\mid_{p \leftarrow 1})$.

Here $\phi\mid_{p \leftarrow a}$ denotes the formula obtained by substituting a for every free occurrence of p in ϕ. An assignment V satisfies a quantified Boolean formula ϕ if $V(\phi) = T$ and finally a quantified Boolean formula ϕ is valid if it is satisfied by every assignment.

For formulae of the form (6.4) we define a mapping $\mathrm{QBF}^2_{\mathrm{AE}}$ which maps a quantified Boolean formula ϕ to a set of propositional autoepistemic formulae

$\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \subseteq \mathcal{L}_{\mathrm{ae}}$. The mapping consists of first giving each quantified variable in ϕ a unique name. Then every constant 1 in E is replaced by $a \vee \neg a$ and every constant 0 by $a \wedge \neg a$ where a is a new atomic sentence because constants 0 and 1 do not belong to the language of autoepistemic logic. Finally the resulting quantified Boolean formula $\phi' = \exists p_1 \ldots \exists p_n \forall q_1 \ldots \forall q_m E$ is mapped to the set

$$\begin{aligned} &\{\neg L\neg p_1 \to p_1, \neg Lp_1 \to \neg p_1, \ldots, \\ &\neg L\neg p_n \to p_n, \neg Lp_n \to \neg p_n, \neg LE \to (a \wedge \neg a)\}. \end{aligned} \tag{6.5}$$

We use the same Σ^p_2-complete problem $\mathbf{QBF}^2$ as that used by Gottlob [40] to prove that **SE** is Σ^p_2-hard but our mapping from quantified Boolean formulae to sets of autoepistemic formulae differs slightly from that of Gottlob. Our mapping gives a uniform proof for **SE**, **SE**(E), and **SE**(LE). This is because the resulting set of autoepistemic formulae is of the special form given in Theorem 5.13. To emphasize this we write $\neg LE \to (a \wedge \neg a)$ instead of LE. By Theorem 5.13 every stable expansion of $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$ is an L-hierarchic expansion and thus an enumeration-based expansion. By Theorem 5.4 enumeration-based expansions and L-hierarchic expansions are stable expansions. Hence for sets of premises of the form resulting from the mapping $\mathrm{QBF}^2_{\mathrm{AE}}$, Moore style stable expansions, enumeration-based expansions, and L-hierarchic expansions coincide. This means that the reduction of $\mathbf{QBF}^2$ to **SE** implies also a reduction to **SE**(E) and **SE**(LE). Observe that it is essential to assign each quantified variable a unique name before mapping a quantified Boolean formula to a set of autoepistemic formulae. For example, the quantified Boolean formula $\exists p \forall p\, p$ is not valid but $\Sigma = \{\neg L\neg p \to p, \neg Lp \to \neg p, \neg Lp \to (a \wedge \neg a)\}$ has a stable expansion $\mathrm{SE}_\Sigma(\{\neg L\neg p, Lp\})$.

Clearly, ϕ' is satisfied by the same assignments as ϕ and thus ϕ is valid if and only if ϕ' is. The mapping $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$ can be computed in polynomial time. So it remains to be shown that $\phi \in \mathbf{QBF}^2$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \in \mathbf{SE}$.

($\Rightarrow$) Let $\phi \in \mathbf{QBF}^2$. Thus $\phi' = \exists p_1 \ldots \exists p_n \forall q_1 \ldots \forall q_m E$ is valid and there is a sequence $S = p_1 \leftarrow \mathrm{TV}(p_1), \ldots, p_n \leftarrow \mathrm{TV}(p_n)$ of substitutions where each $\mathrm{TV}(p_i)$ is 0 or 1 such that for every assignment V, $V(\forall q_1 \ldots \forall q_m E \mid_S) = T$. We show that

$$\begin{aligned} \Lambda \;=\; &\{\neg L\neg p_i \mid \mathrm{TV}(p_i) = 1, 1 \le i \le n\} \cup \\ &\{L\neg p_i \mid \mathrm{TV}(p_i) = 0, 1 \le i \le n\} \cup \\ &\{\neg Lp_i \mid \mathrm{TV}(p_i) = 0, 1 \le i \le n\} \cup \\ &\{Lp_i \mid \mathrm{TV}(p_i) = 1, 1 \le i \le n\} \cup \{LE\} \end{aligned} \tag{6.6}$$

is $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$-full which implies $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \in \mathbf{SE}$ by Theorem 3.8.

By Proposition 5.12 for all $\chi \in \mathcal{L}$,

$$\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \models_{\mathrm{ae}} \chi \text{ if and only if } \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda) \models_{\mathrm{cl}} \chi \tag{6.7}$$

where $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda) = \{\psi' \mid \neg L\psi \rightarrow \psi' \in \mathrm{QBF}^2_{\mathrm{AE}}(\phi) \text{ and } \neg L\psi \in \Lambda\}$.

As $p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda)$ if and only if $\neg L \neg p_i \in \Lambda$ if and only if $\mathrm{TV}(p_i) = 1$ and $\neg p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda)$ if and only if $\neg L p_i \in \Lambda$ if and only if $\mathrm{TV}(p_i) = 0$,

$$\begin{aligned}&\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda) = \\ &\{p_i \mid \mathrm{TV}(p_i) = 1, 1 \leq i \leq n\} \cup \{\neg p_i \mid \mathrm{TV}(p_i) = 0, 1 \leq i \leq n\}.\end{aligned} \tag{6.8}$$

Using this result and Proposition 5.12 it is easy to show that for all $i = 1, \ldots, n$,

1. $\neg L \neg p_i \in \Lambda$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \not\models_{\mathrm{ae}} \neg p_i$,
2. $L \neg p_i \in \Lambda$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \models_{\mathrm{ae}} \neg p_i$,
3. $\neg L p_i \in \Lambda$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \not\models_{\mathrm{ae}} p_i$ and
4. $L p_i \in \Lambda$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \models_{\mathrm{ae}} p_i$.

(1.) If $\neg L \neg p_i \in \Lambda$, then $\mathrm{TV}(p_i) = 1$. By (6.8) $p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda)$ and $\neg p_i \notin \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda)$. Hence $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda) \not\models_{\mathrm{cl}} \neg p_i$ and by Proposition 5.12 $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \not\models_{\mathrm{ae}} \neg p_i$. If $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \not\models_{\mathrm{ae}} \neg p_i$, then by Proposition 5.12 $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda) \not\models_{\mathrm{cl}} \neg p_i$. Thus $\neg p_i \notin \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda)$ which implies $\mathrm{TV}(p_i) = 1$ by (6.8). Hence $\neg L \neg p_i \in \Lambda$. Cases 2.–4. can be shown similarly.

For all $L\phi \in \mathrm{Sf}^{\mathrm{qL}}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi))$, $\phi \in \mathcal{L}$ and thus $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \models_{\mathrm{ae}} \phi$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \models_{\mathrm{L}} \phi$. Hence to prove that Λ is $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$-full we still have to show that $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \models_{\mathrm{ae}} E$ because $LE \in \Lambda$. Assume that $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup \Lambda \not\models_{\mathrm{ae}} E$. Then by Proposition 5.12 $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda) \not\models_{\mathrm{cl}} E$. So there is a propositional interpretation I such that $I \models_{\mathrm{cl}} \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), \Lambda)$ but $I \not\models_{\mathrm{cl}} E$. For a propositional interpretation, there is a corresponding sequence of substitutions such that a Boolean expression is true in the interpretation if and only if the expression after the substitutions is valid. The assertion (S) gives the correspondence.

(S) *Let E be a Boolean formula without any constants $0, 1$ or quantifiers. Let I be a propositional interpretation and $S = y_1 \leftarrow \mathrm{TV}(y_1), \ldots, y_m \leftarrow \mathrm{TV}(y_m)$ a sequence of substitutions where for all $i = 1, \ldots, m$, $\mathrm{TV}(y_i) = 1$ if $I \models_{\mathrm{cl}} y_i$ and otherwise $\mathrm{TV}(y_i) = 0$ and there is a substitution in S for each variable in E. Then for every assignment V, $V(E\mid_S) = T$ if and only if $I \models_{\mathrm{cl}} E$.*

We prove (S) by structural induction. Let E be a variable. Then there is a substitution $E \leftarrow \mathrm{TV}(E)$ in S and $\mathrm{TV}(E) = 1$ if and only if $I \models_{\mathrm{cl}} E$. Thus $V(E \mid_S) = T$ if and only if $\mathrm{TV}(E) = 1$ if and only if $I \models_{\mathrm{cl}} E$.

Let $E = \neg E'$. Then $V(\neg E' \mid_S) = T$ if and only if $V(E' \mid_S) = F$. By the inductive hypothesis this is the case if and only if $I \not\models_{\mathrm{cl}} E'$ which holds exactly when $I \models_{\mathrm{cl}} \neg E'$.

Let $E = E' \wedge E''$. Then $(E' \wedge E'') \mid_S = E' \mid_S \wedge E'' \mid_S$ and $V((E' \wedge E'') \mid_S) = T$ if and only if $V(E' \mid_S) = T$ and $V(E'' \mid_S) = T$. By the inductive hypothesis this holds exactly when $I \models_{\mathrm{cl}} E'$ and $I \models_{\mathrm{cl}} E''$ which is the case if and only if $I \models_{\mathrm{cl}} E' \wedge E''$.

The case $E = E' \vee E''$ can be shown similarly. Thus (S) holds.

Let $S' = p_1 \leftarrow \mathrm{TV}(p_1), \ldots, p_n \leftarrow \mathrm{TV}(p_n), q_1 \leftarrow \mathrm{TV}(q_1), \ldots, q_m \leftarrow \mathrm{TV}(q_m)$, $a \leftarrow \mathrm{TV}(a)$ be a sequence of substitutions corresponding to I as stated in (S). By (S) as $I \not\models_{\mathrm{cl}} E$, $V(E \mid_{S'}) = F$ for every assignment V. For all $i = 1, \ldots, n$,

$$\begin{aligned} p_i \leftarrow 1 \text{ is a substitution in } S' \quad & \text{if and only if} \quad I \models_{\mathrm{cl}} p_i \\ & \text{if and only if} \quad p_i \in \mathrm{Hds}(\Sigma, \Lambda) \\ & \text{if and only if} \quad p_i \leftarrow 1 \text{ is a substitution in } S. \end{aligned}$$

Similarly, $p_i \leftarrow 0$ is a substitution in S' if and only if $p_i \leftarrow 0$ is a substitution in S. Hence each p_i has the same substitution in S and S' which implies

$$E \mid_{S'} = E \mid_S\mid_{q_1 \leftarrow \mathrm{TV}(q_1), \ldots, q_m \leftarrow \mathrm{TV}(q_m), a \leftarrow \mathrm{TV}(a)} .$$

But then there is an assignment V with $V(a) = T$ if $\mathrm{TV}(a) = 1$ and otherwise $V(a) = F$ such that $V(E \mid_S\mid_{q_1 \leftarrow \mathrm{TV}(q_1), \ldots, q_m \leftarrow \mathrm{TV}(q_m)}) = F$, which is a contradiction as for all assignments V, $V(\forall q_1 \ldots \forall q_m E \mid_S) = T$. Thus $\Sigma \cup \Lambda \models_{\mathrm{ae}} E$ and Λ is $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$-full. By Theorem 3.8 $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \in \mathbf{SE}$.

($\Leftarrow$) Let $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \in \mathbf{SE}$ and Δ be a stable expansion of $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$. Assume $\Delta = \mathcal{L}_{\mathrm{ae}}$. Then for any CL-structure I, $\langle I, \mathcal{L}_{\mathrm{ae}} \rangle \models_{\mathrm{ae}} \mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup L\Delta \cup \neg L\overline{\Delta}$. But then $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup L\Delta \cup \neg L\overline{\Delta}$ is not inconsistent although $\Delta = \mathrm{Cn}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup L\Delta \cup \neg L\overline{\Delta})$ is, a contradiction. Thus $\Delta \neq \mathcal{L}_{\mathrm{ae}}$. This implies $E \in \Delta$ because otherwise Δ would be inconsistent. So $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} E$. By Proposition 5.12 $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta}) \models_{\mathrm{cl}} E$.

$$\begin{aligned} &\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta}) = \\ &\{p_i \mid \neg L\neg p_i \to p_i \in \mathrm{QBF}^2_{\mathrm{AE}}(\phi), \neg L\neg p_i \in \neg L\overline{\Delta}, 1 \leq i \leq n\} \cup \\ &\{\neg p_i \mid \neg L p_i \to \neg p_i \in \mathrm{QBF}^2_{\mathrm{AE}}(\phi), \neg L p_i \in \neg L\overline{\Delta}, 1 \leq i \leq n\} \end{aligned} \tag{6.9}$$

We show that for all $i = 1, \ldots, n$, $p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta})$ if and only if $p_i \in \Delta$. If $p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta})$, then $p_i \in \Delta$, as Δ is a stable expansion and $\neg L\neg p_i \rightarrow p_i \in \mathrm{QBF}^2_{\mathrm{AE}}(\phi) \subseteq \Delta$ and $\neg L\neg p_i \in \neg L\overline{\Delta} \subseteq \Delta$. If $p_i \in \Delta$, then $\neg p_i \notin \Delta$ as Δ is not inconsistent. Thus $\neg L\neg p_i \in \neg L\overline{\Delta}$ and $p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta})$. Similarly, it can be shown that $\neg p_i \in \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta})$ if and only if $\neg p_i \in \Delta$. These two results imply

$$\begin{aligned}&\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta}) = \\ &\{p_i \in \Delta \mid 1 \leq i \leq n\} \cup \{\neg p_i \in \Delta \mid 1 \leq i \leq n\}.\end{aligned} \tag{6.10}$$

Let $S = p_1 \leftarrow \mathrm{TV}(p_1), \ldots, p_n \leftarrow \mathrm{TV}(p_n)$ be a sequence of substitutions where for $j = 1, \ldots, n$, $TV(p_j) = 1$ if $p_j \in \Delta$ and otherwise $TV(p_j) = 0$. The quantified Boolean formula ϕ' is valid if $V(\forall q_1 \ldots \forall q_m E \mid_S) = T$ for all assignments V. We show that $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta}) \models_{\mathrm{cl}} E$ implies this. Assume $V(\forall q_1 \ldots \forall q_m E \mid_S) = F$ for some assignment V. So S can be extended to a sequence of substitutions $S' = p_1 \leftarrow \mathrm{TV}(p_1), \ldots, p_n \leftarrow \mathrm{TV}(p_n)$, $q_1 \leftarrow \mathrm{TV}(q_1), \ldots, q_m \leftarrow \mathrm{TV}(q_m)$, $a \leftarrow \mathrm{TV}(a)$ where $\mathrm{TV}(q_i)$ is 1 or 0 and $TV(a) = 1$ if $V(a) = T$ and otherwise $TV(a) = 0$ such that $V(E \mid_{S'}) = F$. Let $I = \{p_i \mid \mathrm{TV}(p_i) = 1, 1 \leq i \leq n\} \cup \{q_i \mid \mathrm{TV}(q_i) = 1, 1 \leq i \leq m\} \cup \{a \mid \mathrm{TV}(a) = 1\}$. As the quantified variables have unique names, S' is a sequence of substitutions corresponding to I as given in (S) and thus $I \not\models_{\mathrm{cl}} E$. But as by (6.10) $I \models_{\mathrm{cl}} \mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta})$, $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta}) \not\models_{\mathrm{cl}} E$ holds. Thus $\mathrm{Hds}(\mathrm{QBF}^2_{\mathrm{AE}}(\phi), L\Delta \cup \neg L\overline{\Delta}) \models_{\mathrm{cl}} E$ implies that for every assignment V, $V(\forall q_1 \ldots \forall q_m E \mid_S) = T$ holds. Hence ϕ' is valid. So we have shown that $\phi \in \mathbf{QBF}^2$ if and only if $\mathrm{QBF}^2_{\mathrm{AE}}(\phi) \in \mathbf{SE}$.

By Theorems 5.4 and 5.13 Moore style, enumeration-based, and L-hierarchic expansions coincide for sets of premises resulting from the mapping $\mathrm{QBF}^2_{\mathrm{AE}}$. Hence the Σ^p_2-hardness proof for $\mathbf{SE}$ shows that $\mathbf{SE}(\mathrm{E})$ and $\mathbf{SE}(\mathrm{LE})$ are Σ^p_2-hard. Then by Corollary 6.24 $\mathbf{SE}(\mathrm{E})$ and $\mathbf{SE}(\mathrm{LE})$ are Σ^p_2-complete. □

The following corollary is a direct consequence of the previous theorem and Proposition 6.6 together with Corollary 6.24.

Corollary 6.26 *If the underlying logic* CL *is the propositional calculus, then* $\mathbf{DERb}$, $\mathbf{DERb}(\mathrm{E})$, *and* $\mathbf{DERb}(\mathrm{LE})$ *are* Σ^p_2*-complete.*

Theorem 6.27 *If the underlying logic* CL *is the propositional calculus, then* $\mathbf{DERc}$, $\mathbf{DERc}(\mathrm{E})$, *and* $\mathbf{DERc}(\mathrm{LE})$ *are* Π^p_2*-complete.*

Proof. As all the problems are in Π^p_2 by Corollary 6.24, it is sufficient to show that they are Π^p_2-hard. We prove that $\overline{\mathbf{DERc}}$, $\overline{\mathbf{DERc}(\mathrm{E})}$, and $\overline{\mathbf{DERc}(\mathrm{LE})}$ are Σ^p_2-hard by showing that $\mathbf{QBF}^2$ is polynomial time reducible to each of the problems. This implies that $\mathbf{DERc}$, $\mathbf{DERc}(\mathrm{E})$, and $\mathbf{DERc}(\mathrm{LE})$ are Π^p_2-hard. The reduction is easily obtained from that used in the proof of Theorem 6.25. We show that $\phi \in \mathbf{QBF}^2$ if and only if $\langle \mathrm{QBF}^2_{\mathrm{AE}}(\phi), b \wedge \neg b \rangle \in \overline{\mathbf{DERc}}$. If

$\phi \in \mathbf{QBF}^2$, then $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$ has a consistent stable expansion as shown in the proof of Theorem 6.25. Thus the stable expansion does not contain $b \wedge \neg b$ and so $\langle \mathrm{QBF}^2_{\mathrm{AE}}(\phi), b \wedge \neg b\rangle \in \overline{\mathbf{DERc}}$. If $\langle \mathrm{QBF}^2_{\mathrm{AE}}(\phi), b \wedge \neg b\rangle \in \overline{\mathbf{DERc}}$, then $\mathrm{QBF}^2_{\mathrm{AE}}(\phi)$ has a stable expansion and thus $\phi \in \mathbf{QBF}^2$. Hence $\overline{\mathbf{DERc}}$ is Σ^p_2-hard. The Σ^p_2-hardness of $\overline{\mathbf{DERc}(\mathrm{E})}$ and $\overline{\mathbf{DERc}(\mathrm{LE})}$ is implied by the fact that the three types of expansions coincide for sets of premises of the form resulting from the mapping $\mathrm{QBF}^2_{\mathrm{AE}}$. □

The Σ^p_2 and Π^p_2-completeness results strongly suggest that propositional autoepistemic reasoning is strictly harder than reasoning in the classical propositional calculus because otherwise the polynomial time hierarchy will collapse which is considered unlikely by the experts [48]. It is perhaps illuminating to see how the collapse of the polynomial time hierarchy comes about.

We have shown that the decision problems in propositional autoepistemic logic are **NP**-hard. The decision problems in propositional reasoning are the satisfiability problem (**SAT**) which is **NP**-complete and the validity problem which is **co-NP**-complete. Consider the satisfiability problem. If autoepistemic reasoning is not strictly harder than propositional satisfiability, then the decision problems in autoepistemic reasoning must be in **NP**. The decision problems in autoepistemic reasoning are either Σ^p_2 or Π^p_2-complete. First we note that if a Π^p_2-hard problem belongs to **NP**, then it is also Σ^p_2-hard.

> Given a Π^p_2-hard problem $P2 \in \mathbf{NP}$, any problem $Q \in \Sigma^p_2$ is polynomial time reducible to $P2$ which can be shown as follows. If $Q \in \Sigma^p_2$, then $\overline{Q}$, the complement of Q, is in Π^p_2. As $P2$ is Π^p_2-hard, $\overline{Q}$ is polynomial time reducible to $P2$. As $P2 \in \mathbf{NP} = \Sigma^p_1$, $\overline{Q} \in \Sigma^p_1$. Hence $Q \in \Pi^p_1 \subseteq \Delta^p_2 \subseteq \Pi^p_2$. By the Π^p_2-hardness of $P2$, Q is polynomial time reducible to $P2$. Thus $P2$ is Σ^p_2-hard.

So if a decision problem in propositional autoepistemic logic is in **NP**, then it is a Σ^p_2-hard set belonging to **NP**. It can be shown that this implies $\Sigma^p_1 = \Pi^p_1$.

> Let $S2$ be a Σ^p_2-hard set such that $S2 \in \mathbf{NP} = \Sigma^p_1$. Let $Q \in \Sigma^p_1$. As $\Sigma^p_1 \subseteq \Delta^p_2 \subseteq \Pi^p_2$, $Q \in \Pi^p_2$. Hence $\overline{Q}$, the complement of Q, is in Σ^p_2. As $S2$ is Σ^p_2-hard, $\overline{Q}$ is polynomial time reducible to $S2$. Thus $\overline{Q} \in \Sigma^p_1$ as $S2 \in \Sigma^p_1$ and this implies $Q \in \Pi^p_1$. Hence $\Sigma^p_1 \subseteq \Pi^p_1$.
>
> Let $Q \in \Pi^p_1$. Hence $\overline{Q} \in \Sigma^p_1$. As $\Sigma^p_1 \subseteq \Delta^p_2 \subseteq \Pi^p_2$, $\overline{Q} \in \Pi^p_2$ and thus $Q \in \Sigma^p_2$. As $S2$ is Σ^p_2-hard, Q is polynomial time reducible to $S2$. Thus $Q \in \Sigma^p_1$ as $S2 \in \Sigma^p_1$. Hence $\Pi^p_1 \subseteq \Sigma^p_1$ and $\Sigma^p_1 = \Pi^p_1$.

This leads to the collapse of the polynomial time hierarchy at level 1 because the polynomial time hierarchy has the property that if for some $k \geq 1$ $\Sigma^p_k = \Pi^p_k$, then $\Sigma^p_k = \Sigma^p_j = \Pi^p_j$ for all $j \geq k$ [3]. Hence $\Sigma^p_1 = \Sigma^p_j = \Pi^p_j$ for all $j \geq 1$.

The decision procedures presented in this chapter clearly indicate that there are two orthogonal sources of complexity in autoepistemic reasoning. One is required classical monotonic reasoning, i.e. deciding **CN**. The other is determining an acceptable order of introspection (or in the case of Moore style expansions choosing an appropriate subset of $L\phi$ quasi-subformulae of the premises). The independence of the two sources of complexity explains why the decision problems reside on the second level of the polynomial time hierarchy: although an oracle is provided for deciding **CN**, an additional set of nondeterministic choices is needed to supply the required sequence of introspection steps (or subset of $L\phi$ quasi-subformulae) to ensure that an accepting computation can be accomplished in polynomial time.

From the practical point of view it would be interesting to identify tractable subclasses of autoepistemic reasoning, i.e. subclasses for which a polynomial time decision procedure can be developed. The complexity results presented in this chapter imply that both sources of complexity have to be addressed if tractable reasoning is to be achieved. The computational complexity of required monotonic reasoning can be reduced by simply restricting the form of sentences allowed in the premises. However, it is not uncomplicated to devise restrictions that reduced the computational complexity related to finding an acceptable order of introspection. In particular, the straightforward approach of limiting the form of sentences in the premises does not yield expected results. Marek and Truszczyński [87] show that for formulae of the form

$$Lp_1 \wedge \cdots \wedge Lp_r \wedge \neg Lq_1 \wedge \cdots \wedge \neg Lq_s \rightarrow u \tag{6.11}$$

where p_i, q_j, and u are all atomic, the problem of deciding whether there exists a stable expansion is **NP**-complete. So even in this very limited case determining whether there exists some introspection sequence (or a subset of $L\phi$ quasi-subformulae of the premises) inducing an expansion seems to lead inevitably to searching through exponentially many possible alternatives.

In autoepistemic reasoning it is not enough to determine the existence of an expansion of the premises but the central problem is to find an expansion satisfying some interesting property, e.g. an expansion containing a given sentence or not containing a given sentence. It turns out that although the existence of an expansion is guaranteed, determining whether there is an expansion fulfilling a required property may call for exponential computational resources. The intuitive explanation is that already in very limited cases there can be exponentially many expansions and there seems to be no other way than to go systematically through every expansion until an expansion satisfying the required property is found. As an example of this, we consider a very simple and regular set of premises.

Example 6.1 Consider a finite set of atomic formulae P. Let

$$\Sigma(P) = \{\neg Lp' \rightarrow p \mid p \in P\} \cup \{\neg Lp \rightarrow p' \mid p \in P\}. \tag{6.12}$$

By Theorem 5.13 Moore style, enumeration-based, and L-hierarchic expansions coincide for $\Sigma(P)$. For a set $\Sigma(P)$ induced by some finite set of atomic formulae P, it is straightforward to prove that the following two properties hold.

(P1) For every $S \subseteq P$,

$$\begin{aligned}\Lambda(S) \quad = \quad & \{Lp \mid p \in S\} \cup \{\neg Lp' \mid p \in S\} \cup \\ & \{\neg Lp \mid p \in P - S\} \cup \{Lp' \mid p \in P - S\}\end{aligned} \tag{6.13}$$

is $\Sigma(P)$-full and thus $\Delta = \mathrm{SE}_{\Sigma(P)}(\Lambda(S))$ is a stable expansion of $\Sigma(P)$.

(P2) For every stable expansion Δ of $\Sigma(P)$, $\Delta = \mathrm{SE}_{\Sigma(P)}(\Lambda(P \cap \Delta))$.

So for every subset of P, there is an expansion and $\Sigma(P)$ has 2^n expansions where n is the cardinality of P. Every expansion contains either p or p' for each $p \in P$. Hence an expansion Δ can be viewed as a propositional interpretation where $p \in \Delta$ means that an atom p is true in the interpretation and $p' \in \Delta$ means that an atom p is false in the interpretation. This leads to a situation where although the set of premises is very restricted so that required monotonic reasoning is almost trivial, it seems that finding an expansion satisfying some simple property may require exponential resources. We consider the problem of deciding whether there is an expansion containing a propositional formula in conjunctive normal form but containing no negation symbols. We call such formulae NF-formulae. For NF-formulae the membership in an expansion Δ of $\Sigma(P)$ can be tested very easily in polynomial time given the set of atoms A in Δ:

- $\phi \in \Delta$ if and only if $\phi \in A$ for an atom ϕ;
- $\phi' \wedge \phi'' \in \Delta$ if and only if $\phi' \in \Delta$ and $\phi'' \in \Delta$;
- $\phi' \vee \phi'' \in \Delta$ if and only if $\phi' \in \Delta$ or $\phi'' \in \Delta$.

For each expansion Δ, the set of atoms A in Δ is a finite set $A = (P \cap \Delta) \cup \{p' \mid p \in P - \Delta\}$. Although the requirement that the expansion contains an NF-formula is simple and can be tested efficiently, determining whether such an expansion exists seems to require the examination of exponentially many alternatives. This is because, e.g., satisfiability of propositional formulae in conjunctive normal form, which is an **NP**-complete problem, can be reduced to this problem. Let ϕ be a propositional formula in conjunctive normal form. Let ϕ' be the formula obtained from ϕ by replacing every negated variable $\neg p$ by p'. Then ϕ is satisfiable if and only if ϕ' belongs to some expansion of $\Sigma(P)$ where P is the set of atoms in ϕ. ■

A characteristic feature of sets of premises for which autoepistemic reasoning is **NP**-hard is that the premises contain formulae which establish a connection between the truth value of a formula and the agent's disbelief of the same formula. The sets of premises used in the proof of Theorem 6.25 and in Example 6.1 are simple instances of such a structure. This indicates that these kinds of structures are responsible for the fact that limiting the form of formulae in the premises to make required monotonic reasoning efficient does not lead to tractable autoepistemic reasoning.

It is interesting to study in what way disallowing such structures affects the complexity of autoepistemic reasoning. Niemelä and Rintanen [112] have studied an interesting subclass of autoepistemic reasoning where such structures are not permitted. This subclass is based on the notion of stratification [87] originally introduced for logic programs. A stratified set of premises can be partioned into a sequence of layers such that the truth value of a formula on a layer depends on beliefs and disbeliefs of formulae whose truth values hinge only on formulae appearing in the previous layers in the sequence. Such a restriction forces a unique stable expansion. For a stratified set of premises, brave and cautious autoepistemic reasoning coincide: a formula is an autoepistemic conclusion from a stratified set of premises if and only if it belongs to the unique stable expansion of the premises. A sequence of layers can be computed from a stratified set of premises and it provides an acceptable sequence of introspection steps inducing the corresponding full set. By exploiting these observations a simple iterative decision procedure has been developed for stratified autoepistemic reasoning [112]. The complexity of the procedure is dominated by required monotonic reasoning. Thus tractable subclasses of autoepistemic reasoning can be obtained by further restricting the form of formulae in the premises so that required monotonic reasoning can be implemented efficiently. As an example of the technique, Niemelä and Rintanen [112] have identified two sublasses of autoepistemic reasoning with linear time and quadratic time decision procedures, respectively. These results are shown to be applicable to other forms of nonmonotonic reasoning such as stratified default logic, logic programs, truth maintenance systems, and McDermott and Doyle style nonmonotonic modal logics [112].

Chapter 7

Autoepistemic Logic and Other Forms of Nonmonotonic Reasoning

In this chapter we study relationships between autoepistemic logic and other forms of nonmonotonic reasoning. As indicated in Chapter 1, Moore style autoepistemic logic captures a large range of forms of nonmonotonic reasoning. We build on these results. Our aim is to show that the results can be extended considerably when employing the tightly grounded class of L-hierarchic expansions instead of Moore style expansions. There is a large body of results to draw on because the relationships between more tightly grounded forms of autoepistemic reasoning and other types of nonmonotonic reasoning have been investigated by several researchers. In particular, default logic with its close connections to logic programs and deductive databases has drawn considerable interest [61, 84, 86, 145, 144].

In this chapter, we show that default logic, important forms of circumscription, answer set semantics of logic programs with classical negation, nonmonotonic truth maintenance systems, and forms of abduction can be seen as special cases of autoepistemic logic based on L-hierarchic expansions using very simple and intuitive embeddings. Eliminating weakly grounded expansions seems to bring autoepistemic logic closer to other important forms of nonmonotonic reasoning. For example, when embedding default logic into Moore style autoepistemic logic extra conditions or complicated embeddings are needed to compensate the weak groundedness of Moore style expansions [61, 84, 41]. When considering only L-hierarchic expansions, the embedding is straightforward and no additional conditions are required. The results provide a unified view of different forms of nonmonotonic reasoning through autoepistemic reasoning and indicate that autoepistemic logic based on L-hierarchic expansions offers an interesting theoretical basis for nonmonotonic reasoning.

Autoepistemic reasoning is a form of symbolic nonmonotonic reasoning. This means that numeric factors describing probability, possibility, degree of belief, or fuzziness are not directly within the reach of autoepistemic logic. Nonmonotonicity of autoepistemic reasoning is brought about by the negative introspection rule. This rule is very simple and straightforward. It does not contain any involved principles of nonmonotonic reasoning. For example, in inheritance reasoning a specificity notion is used to resolve conflicts among competing properties inherited in a hierarchy. The specificity condition says informally that more specific facts prevail over those less specific and leads to a principle of reasoning in inheritance hierarchies which states roughly that information about subclasses is more reliable than that about superclasses [142]. However, formalization of this kind of a specificity notion is not uncomplicated [143]. There is no particular notion of specificity in autoepistemic reasoning and thus autoepistemic conclusions from a set of premises do not directly satisfy any definite specificity requirement. However, this does not rule out the possibility of enforcing a desired notion of specificity. Gelfond and Przymusinska [35] show how to embed inheritance reasoning into autoepistemic logic and how to incorporate required specificity notions to autoepistemic reasoning. A similar phenomenon emerges in reasoning about actions. Hanks and McDermott [43] come to the conclusion that standard nonmonotonic logics such as circumscription and default logic are inherently incapable of representing certain kind of default reasoning after examining a simple problem of reasoning about actions (the Yale shooting problem). They attempt to capture reasoning about effects of actions by applying circumscription and default logic but discover that their straightforward formalization leads to counterintuitive results. The claim of Hanks and McDermott led to a sequence of papers which proposed solutions to the Yale shooting problem [30, 57, 73, 72, 102, 134]. A popular group of solutions (e.g. [57, 72, 134]) is based on the idea that default assumptions are applied in temporal order. Of course, autoepistemic logic does not support directly such a mode of reasoning. However, it seems that temporal order is not essential. Gelfond [30] shows how to formalize the Yale shooting problem in autoepistemic logic by utilizing the negative introspection rule in a straightforward and natural way. Baker [2] presents how to formalize the problem using circumscription so that no temporal order is needed either.

7.1 Default Logic

Reiter's [122] default logic is one of the leading formalizations of nonmonotonic reasoning. It attempts to formalize default reasoning of the form "in the absence of any information to the contrary, assume ...". This is achieved by extending classical predicate logic with default rules of the form

$$\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma} \tag{7.1}$$

where $\alpha, \beta_1, \ldots, \beta_n, \gamma$ are all closed formulae of predicate logic. The rule can be read as if the *prerequisite* α holds and it is consistent to assume all the *justifications* $\beta_1, \ldots, \beta_n$, then the *conclusion* γ holds.

Statements of the type "birds typically fly" are formalized in default logic using a default rule stating that a bird for which it is consistent to assume that it flies also flies:

$$\frac{b : f}{f}.$$

In default logic knowledge is represented by a pair (D, W) where D is a set of default rules and W is a set of ordinary first-order sentences. A set of correct nonmonotonic inferences drawn from a *default theory* (D, W) is called an *extension* of (D, W). When presenting Reiter's definition of an extension, we use an operator $\text{apply}_{D,E}(E')$ which applies the default rules from D whose prerequisites are known (are in E') and whose justifications are consistent with E.

$$\begin{aligned} &\text{apply}_{D,E}(E') = \\ &\qquad \{\gamma \mid \frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma} \in D, \alpha \in E', \neg\beta_1 \notin E, \ldots, \neg\beta_n \notin E\} \end{aligned} \tag{7.2}$$

Extensions are fixed points of an operator $\Gamma_{(D,W)}$ defined as follows.

For a set of first-order sentences E, $\Gamma_{(D,W)}(E)$ is the smallest set satisfying the following properties.

1. $W \subseteq \Gamma_{(D,W)}(E)$,
2. $\text{Cn}_{\text{cl}}(\Gamma_{(D,W)}(E)) = \Gamma_{(D,W)}(E)$ and
3. $\text{apply}_{D,E}(\Gamma_{(D,W)}(E)) \subseteq \Gamma_{(D,W)}(E)$.

A set $E \subseteq \mathcal{L}$ is an extension of a default theory (D, W) if and only if $E = \Gamma_{(D,W)}(E)$. Thus an extension E of (D, W) is a minimal set containing W

which is closed under first-order consequence and for which none of the default rules can be applied to obtain a conclusion not already in E.

Reiter [122, Theorem 2.1] shows that an extension can be defined in an alternative way. Given a default theory (D,W) and $E \subseteq \mathcal{L}$, let

$$E_0 = W \tag{7.3}$$

and for all $i = 0, 1, \ldots,$

$$E_{i+1} = \mathrm{Cn}_{\mathrm{cl}}(E_i) \cup \mathrm{apply}_{D,E}(E_i). \tag{7.4}$$

Then E is an extension of (D, W) if and only if

$$E = \bigcup_{i=0}^{\infty} E_i. \tag{7.5}$$

Reiter [122] points out that $\bigcup_{i=0}^{\infty} E_i$ enjoys the following properties.

1. $W \subseteq \bigcup_{i=0}^{\infty} E_i$,
2. $\mathrm{Cn}_{\mathrm{cl}}(\bigcup_{i=0}^{\infty} E_i) = \bigcup_{i=0}^{\infty} E_i$ and
3. $\mathrm{apply}_{D,E}(\bigcup_{i=0}^{\infty} E_i) \subseteq \bigcup_{i=0}^{\infty} E_i$.

Furthermore, Reiter [122, Theorem 2.5] shows that if E is an extension of (D, W), then

$$E = \mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E)). \tag{7.6}$$

Default logic is one of the most prominent formalizations of nonmonotonic reasoning. The books of Besnard [6] and Etherington [26] are good sources on default logic. There is a body of results indicating that default logic can capture a large number of different forms of nonmonotonic reasoning. Default logic is especially closely related to reasoning in logic programs and deductive databases [8, 9, 33]. Default logic has been applied in formalizing inheritance reasoning and truth maintenance systems [24, 49]. Connections have been established between default logic and autoepistemic logic and McDermott and Doyle style nonmonotonic modal logics [61, 84, 86, 89, 144, 145], circumscription [25, 26], diagnosis [123], and abductive reasoning [113]. So default logic can be put forward as an attractive candidate for a unified framework for a large part of nonmonotonic reasoning. Default logic has not a uniform syntax but two categories of syntactical objects, first-order sentences and default rules. Hence representing defaults about defaults is not quite straightforward, for example. Also expressions saying that something is believed or disbelieved prove to be useful in particular for queries and integrity constraints [124] but are not available in default logic.

In this section we show that default logic can be seen as a special case of autoepistemic logic based on L-hierarchic expansions where default rules correspond to a particular subclass of autoepistemic sentences and extensions correspond to L free parts of L-hierarchic expansions. We start by a brief survey of proposals for embedding default logic into systems based on modal logics.

Although default logic seems to be quite far from autoepistemic logic, the two are very closely related. Konolige [61, 62] was the first to establish a correspondence between default logic and autoepistemic logic. Konolige [61] proposes the following translation of a default theory into a set of autoepistemic sentences.

$$\begin{aligned} \mathrm{tr}_{\mathrm{K}}(\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}) &= (L\alpha \wedge \neg L \neg \beta_1 \wedge \cdots \wedge \neg L \neg \beta_n) \rightarrow \gamma \\ \mathrm{tr}_{\mathrm{K}}(D, W) &= W \cup \{\mathrm{tr}_{\mathrm{K}}(d) \mid d \in D\} \end{aligned} \tag{7.7}$$

He shows that under this translation extensions of a default theory correspond to stable expansions of the translated defaults satisfying a special groundedness condition.

Konolige [61] shows that the sentences in a normal form

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \gamma \tag{7.8}$$

where $\alpha, \beta_1, \ldots, \beta_n, \gamma$ are all sentences without the belief operator can be translated into default rules using the same idea.

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \gamma \quad \mapsto \quad \frac{\alpha : \neg\beta_1, \ldots, \neg\beta_n}{\gamma} \tag{7.9}$$

Stable expansions of a set of premises in the normal form satisfying the special groundedness condition correspond to extensions of a default theory obtained by translating the premises using the transformation above.

The work of Konolige is extended by Marek and Truszczyński [84] who establish a multilevel correspondence between propositional autoepistemic logic and propositional default logic using the Konolige translation. There are several groundedness conditions characterizing the subclass of stable expansions corresponding to extensions in default logic under the translation (7.7). Strongly grounded expansions suggested by Konolige and robust and strongly iterative expansions proposed by Marek and Truszczyński are discussed in Chapter 4. The conditions are all defined only for sentences in the normal form resulting from the translation (7.7) and suffer from the dependence on the syntactic representation of the premises as discussed in Chapter 4.

Marek and Truszczyński [85, 86] propose an embedding of default logic into autoepistemic logic based on a more satisfying class of grounded expansions. They introduce the following translation of default rules

$$\begin{aligned} \mathrm{tr}_{\mathrm{MT}}\left(\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}\right) &= (L\alpha \wedge \neg LL\neg\beta_1 \wedge \cdots \wedge \neg LL\neg\beta_n) \rightarrow \gamma \quad (7.10) \\ \mathrm{tr}_{\mathrm{MT}}(D, W) &= W \cup \{\mathrm{tr}_{\mathrm{MT}}(d) \mid d \in D\} \end{aligned}$$

and show that an extension of a default theory is the L free part of an iterative expansion of the translated default theory. The advantage of this approach is that iterative expansions are not dependent on the syntactic representation of the premises but the translation of default rules lacks a natural, intuitive justification.

Truszczyński [144] introduces a very natural translation of default rules

$$\begin{aligned} \mathrm{tr}_{\mathrm{T}}\left(\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}\right) &= (L\alpha \wedge L\neg L\neg\beta_1 \wedge \cdots \wedge L\neg L\neg\beta_n) \rightarrow \gamma \quad (7.11) \\ \mathrm{tr}_{\mathrm{T}}(D, W) &= W \cup \{\mathrm{tr}_{\mathrm{T}}(d) \mid d \in D\} \end{aligned}$$

and shows that under this translation extensions in propositional default logic correspond to expansions in McDermott and Doyle style nonmonotonic modal logics with a wide range of underlying monotonic modal logics from **N** to **S4F**. Truszczyński [144] also shows that under the Konolige translation (7.7) extensions correspond to S-expansions satisfying a special groundedness condition for a wide range of McDermott and Doyle style nonmonotonic logics S. Truszczyński [145] introduces yet another embedding of default logic into McDermott and Doyle style nonmonotonic logics.

$$\begin{aligned} \mathrm{tr}_{\mathrm{T'}}\left(\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}\right) &= (L\alpha \wedge L\neg L\neg\beta_1 \wedge \cdots \wedge L\neg L\neg\beta_n) \rightarrow L\gamma \quad (7.12) \\ \mathrm{tr}_{\mathrm{T'}}(D, W) &= \{L\phi \mid \phi \in W\} \cup \{\mathrm{tr}_{\mathrm{T'}}(d) \mid d \in D\} \end{aligned}$$

Here each sentences in the classical part of the default theory is suffixed by a belief operator. The advantage of this approach is that also disjunctive default rules proposed by Gelfond et al. [34] can be captured but, on the other hand, the underlying logic is a modal calculus and not a first-order calculus as in the case of default logic. Lin and Shoham [77, 78] and Lifschitz [75] show how to embed default logic into a system based on two modal operators.

Gottlob [41] establishes a rather surprising result that default logic can also be embedded into standard Moore style autoepistemic logic. The translation is polynomial time but rather involved. The key idea is that weakly grounded expansions not corresponding to extensions in default logic can be eliminated by adding new premises, grounding conditions.

In this section our aim is to show that default logic can be seen as a special case of autoepistemic logic based on L-hierarchic expansions. In fact there are many alternative ways to embed default logic into autoepistemic logic based on L-hierarchic expansions. One possibility is to apply the approach of Marek and Truszczyński [86] (7.10) because the translated default theory is a set of T-clauses and Theorem 5.16 shows that for T-clauses iterative and L-hierarchic expansions coincide.

Theorem 7.1 *Let (D, W) be a default theory where D is finite. Then $E \subseteq \mathcal{L}$ is an extension of (D, W) if and only if $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr}_{\mathrm{T}}(D, W)) \cap \mathcal{L}$ for a $\mathrm{tr}_{\mathrm{MT}}(D, W)$-acceptable L-hierarchic enumeration ε.*

Proof. ($\Rightarrow$) Let $E \subseteq \mathcal{L}$ be an extension of (D, W). Then by [85, Theorem 4.2] $E = S \cap \mathcal{L}$ for an iterative expansion S of $\mathrm{tr}_{\mathrm{T}}(D, W)$. As D is finite, $\mathrm{Sf}^{\mathrm{L}}(\mathrm{tr}_{\mathrm{T}}(D, W))$ is a finite set and by Theorem 5.16 there is a $\mathrm{tr}_{\mathrm{T}}(D, W)$-acceptable L-hierarchic enumeration ε such that $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr}_{\mathrm{T}}(D, W)) \cap \mathcal{L}$.

($\Leftarrow$) Let $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr}_{\mathrm{T}}(D, W)) \cap \mathcal{L}$ for a $\mathrm{tr}_{\mathrm{T}}(D, W)$-acceptable L-hierarchic enumeration ε. By Theorem 5.14 $\mathbf{SE}^{\varepsilon}(\mathrm{tr}_{\mathrm{T}}(D, W))$ is an iterative expansion of $\mathrm{tr}_{\mathrm{T}}(D, W)$ and by [85, Theorem 4.2] E is an extension of (D, W). $\square$

This embedding is not completely satisfying as it lacks a natural and intuitive justification. Next we show that Reiter's default logic can be embedded into autoepistemic logic based on L-hierarchic expansions using the more appealing translation of default rules into autoepistemic sentences (7.11) proposed by Truszczyński [144]. The idea in this translation is that default rules are inference rules stating that

> if α and $\neg L \neg \beta_1, \ldots, \neg L \neg \beta_n$ are believed, then γ holds.

This differs from the Konolige translation (7.7) where a default rule is interpreted as

> if α is believed and $\neg \beta_1, \ldots, \neg \beta_n$ are not believed, then γ holds.

The Konolige translation requires that the justification are possible (not believed not to hold). The Truszczyński translation strengthens this and demands that they have to be believed to be possible. This seems to capture the essence of default rules which are inference rules rather than implications. In the Konolige translation default rules are mapped into implications which can be used together through case-analysis. Clearly, default rules cannot interact in such a manner. This is why extra conditions are needed to eliminate unwanted interactions.

Example 7.1 Consider a default theory (D, W) where

$$D = \{\frac{\neg\neg p :}{p}, \frac{: \neg p}{p}\}, W = \emptyset. \tag{7.13}$$

Then $\mathrm{tr}_{\mathrm{K}}(D, W) = \{L\neg\neg p \rightarrow p, \neg L\neg\neg p \rightarrow p\}$ has a unique stable expansion $\mathrm{SE}_{\mathrm{tr}_{\mathrm{K}}(D,W)}(\{L\neg\neg p\})$ containing p but (D, W) has no extension. To achieve a match between expansions of $\mathrm{tr}_{\mathrm{K}}(D, W)$ and extensions of (D, W), the interaction of implications in $\mathrm{tr}_{\mathrm{K}}(D, W)$ is eliminated by an extra groundedness condition. But $\mathrm{tr}_{\mathrm{K}}(D, W)$ is propositionally equivalent to p which has the same unique expansion and there is no unwanted interaction of implications. So the extra groundedness condition is syntax dependent and as discussed in Chapter 4 this leads to an unsatisfying situation where classically equivalent sets of premises do not have the same set of grounded expansions. ■

Although the Truszczyński embedding (7.11) of default logic into McDermott and Doyle style nonmonotonic modal logics works for many underlying modal logic S, it is valid only for consistent S-expansions.

Example 7.2 Let

$$W = \{a\}, D = \{\frac{\neg b :}{\neg a}, \frac{: b}{\neg b}\} \tag{7.14}$$

where a and b are atomic. Then $\mathrm{tr}_{\mathrm{T}}(D, W) = \{a, L\neg b \rightarrow \neg a, L\neg L\neg b \rightarrow \neg b\}$ has an inconsistent N-expansion because $\neg a$ follows from $\mathrm{tr}_{\mathrm{T}}(D, W)$ in the following way using only propositional reasoning and the necessitation rule (4.16).

1.	a	(premise)
2.	$L\neg b \rightarrow \neg a$	(premise)
3.	$L\neg L\neg b \rightarrow \neg b$	(premise)
4.	$\neg L\neg b$	1, 2 (propositional reasoning)
5.	$L\neg L\neg b$	4 (necessitation rule)
6.	$\neg b$	3, 5 (propositional reasoning)
7.	$L\neg b$	6 (necessitation rule)
8.	$\neg a$	2, 7 (propositional reasoning)

This implies that given any reasonable McDermott and Doyle style nonmonotonic modal logic S containing the necessitation rule, $\mathrm{tr}_{\mathrm{T}}(D, W)$ has an inconsistent S-expansion. However, (D, W) has not an inconsistent extension and neither has $\mathrm{tr}_{\mathrm{T}}(D, W)$ an inconsistent L-hierarchic expansion. Thus weak groundedness of expansions in McDermott and Doyle style nonmonotonic modal logics prevents a perfect match between modal expansions and extensions in default logic under the Truszczyński translation (7.11). ■

The main result of the section states that $E \subseteq \mathcal{L}$ is an extension of a default theory (D, W) if and only if $E = \Delta \cap \mathcal{L}$ for an L-hierarchic expansion Δ of $\mathrm{tr}_{\mathrm{T}}(D, W)$. This result is significant in two respects. Firstly, it shows that L-hierarchic expansions capture extensions faithfully and there is no anomaly in the inconsistent case. Secondly, the result indicates that extensions can be understood in an alternative way as a product of a sequence of introspection steps.

We start by establishing some preliminary results. First we prove a proposition about the alternative definition of extensions (7.5).

Proposition 7.2 *Let (D, W) be a default theory and E a set of sentences. For all $i = 1, 2, \ldots$, $\mathrm{Cn}_{\mathrm{cl}}(E_i) = \mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E_{i-1}))$*

Proof. Let $i = 1$. Then $E_1 = \mathrm{Cn}_{\mathrm{cl}}(W) \cup \mathrm{apply}_{D,E}(W)$. Thus $E_1 \models_{\mathrm{cl}} \phi$ if and only if $W \cup \mathrm{apply}_{D,E}(E_0) \models_{\mathrm{cl}} \phi$.

Assume that the proposition holds for every $j < i$.

$$\mathrm{Cn}_{\mathrm{cl}}(E_i) = \mathrm{Cn}_{\mathrm{cl}}(\mathrm{Cn}_{\mathrm{cl}}(E_{i-1}) \cup \mathrm{apply}_{D,E}(E_{i-1})). \tag{7.15}$$

Thus by the inductive hypothesis

$$\mathrm{Cn}_{\mathrm{cl}}(E_i) = \mathrm{Cn}_{\mathrm{cl}}(\mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E_{i-2})) \cup \mathrm{apply}_{D,E}(E_{i-1})). \tag{7.16}$$

It is quite straightforward to show that this implies

$$\mathrm{Cn}_{\mathrm{cl}}(E_i) = \mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E_{i-2}) \cup \mathrm{apply}_{D,E}(E_{i-1})). \tag{7.17}$$

Hence

$$\mathrm{Cn}_{\mathrm{cl}}(E_i) = \mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E_{i-1})) \tag{7.18}$$

as $E_{i-2} \subseteq E_{i-1}$ which implies that $\mathrm{apply}_{D,E}(E_{i-2}) \subseteq \mathrm{apply}_{D,E}(E_{i-1})$). $\square$

The sentences resulting from the Truszczyński translation are of a simple type for which there is a close connection between the consequence relations $\models_{\mathrm{ae}}$ and $\models_{\mathrm{cl}}$. This connection is utilized frequently in the proofs.

Proposition 7.3 *Let the underlying logic be a first-order calculus. Let $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ be a set of sentences of the form $(Lb_1 \wedge \cdots \wedge Lb_n) \rightarrow c$ where $c \in \mathcal{L}$ and $n \geq 0$ and let Λ be a set of quasi-literals containing no formula and its negation simultaneously. Then for all $\phi \in \mathcal{L}$,*

$$\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi \textit{ if and only if } \mathrm{Hds}(\Sigma, \Lambda) \models_{\mathrm{cl}} \phi$$

where $\mathrm{Hds}(\Sigma, \Lambda) = \{c \mid (Lb_1 \wedge \cdots \wedge Lb_n) \rightarrow c \in \Sigma$ *and* $Lb_i \in \Lambda$ *for all* $i = 1, \ldots, n\}$.

Proof. ($\Rightarrow$) Assume that $\mathrm{Hds}(\Sigma, \Lambda) \not\models_{\mathrm{cl}} \phi$ holds. Then there exists a CL-structure I such that $I \models_{\mathrm{cl}} \mathrm{Hds}(\Sigma, \Lambda)$ but $I \not\models_{\mathrm{ae}} \phi$. Let $T = \{\psi \mid L\psi \in \Lambda\}$. Then $\langle I, T\rangle \models_{\mathrm{ae}} \Lambda$ as Λ contains no quasi-atom and its negation and $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$. Let $\psi = (Lb_1 \wedge \cdots \wedge Lb_n) \to c \in \Sigma$. Assume that $\langle I, T\rangle \not\models_{\mathrm{ae}} c$. Then $I \not\models_{\mathrm{cl}} c$ and $c \notin \mathrm{Hds}(\Sigma, \Lambda)$. So there is some i, $1 \leq i \leq n$ such that $Lb_i \notin \Lambda$. Then $b_i \notin T$ which implies $\langle I, T\rangle \not\models_{\mathrm{ae}} Lb_1 \wedge \cdots \wedge Lb_n$. Thus $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma$. Hence $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda$ but $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$, i.e. $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} c$.

($\Leftarrow$) Assume that $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} c$ holds. Then there exists an AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \Lambda$ but $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$. Let $c \in \mathrm{Hds}(\Sigma, \Lambda)$. Then $(Lb_1 \wedge \cdots \wedge Lb_n) \to c \in \Sigma$ and for all $i = 1, \ldots, n$, $Lb_i \in \Lambda$. As $\langle I, T\rangle \models_{\mathrm{ae}} \Lambda \cup \{(Lb_1 \wedge \cdots \wedge Lb_n) \to c\}$, $\langle I, T\rangle \models_{\mathrm{ae}} c$ holds. Thus $I \models_{\mathrm{cl}} c$. We have shown that $I \models_{\mathrm{cl}} \mathrm{Hds}(\Sigma, \Lambda)$. As $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$, $I \not\models_{\mathrm{cl}} \phi$ holds. Thus $\mathrm{Hds}(\Sigma, \Lambda) \not\models_{\mathrm{cl}} \phi$. □

Now we show that an extension of a default theory is the L free part of an L-hierarchic expansion of the translated theory.

Lemma 7.4 *Let E be an extension of a default theory (D, W) where D is finite. Then $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr}_{\mathrm{T}}(D, W)) \cap \mathcal{L}$ for some L-hierarchic $\mathrm{tr}_{\mathrm{T}}(D, W)$-acceptable enumeration ε of $\mathcal{L}_{\mathrm{ae}}$.*

Proof. By (7.5) $E = \bigcup_{i=0}^{\infty} E_i$. To show the existence of an L-hierarchic $\mathrm{tr}_{\mathrm{T}}(D, W)$-acceptable enumeration of $\mathcal{L}_{\mathrm{ae}}$, by Theorem 5.10 it is sufficient to construct an L-hierarchic $\mathrm{tr}_{\mathrm{T}}(D, W)$-acceptable enumeration of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\mathrm{tr}_{\mathrm{T}}(D, W))\}$. Let $\Sigma = \mathrm{tr}_{\mathrm{T}}(D, W)$. Let $\varepsilon_n = \psi_1, \ldots, \psi_n$ be an enumeration of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that

(C1) *$\psi_1, \ldots, \psi_{m_1}$ are all the sentences ψ_i such that $\psi_i \in \mathcal{L}$ and $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ but $\psi_i \notin E$;*

(C2) *$\psi_{m_1+1}, \ldots, \psi_{m_1+m_2}$ are all the sentences ψ_i such that $L\psi_i = L\neg L\psi_j \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ for some j, $1 \leq j \leq m_1$;*

(C3) *$\psi_{m_1+m_2+1}, \ldots, \psi_{m_1+m_2+m_3}$ are all the sentences ψ_i such that $\psi_i \in E$ and $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and the sentences are in the order they appear in $\bigcup_{i=0}^{\infty} E_i$, i.e. if $\psi_i, \psi_j \in E$, $L\psi_i, L\psi_j \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and there exists some $k \geq 0$ such that $E_k \models_{\mathrm{cl}} \psi_i$ but $E_k \not\models_{\mathrm{cl}} \psi_j$, then $i < j$;*

(C4) *$\psi_{m_1+m_2+m_3+1}, \ldots, \psi_{m_1+m_2+m_3+m_4}$ are all the sentences ψ_i such that $L\psi_i = L\neg L\psi_j \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ for some j, $m_1 + m_2 + 1 \leq j \leq m_1 + m_2 + m_3$.*

Let $n = m_1 + m_2 + m_3 + m_4$. Conditions (C1)–(C4) guarantee that the enumeration ε_n is L-hierarchic. We show that ε_n is a Σ-acceptable enumeration of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. First, we show that ε_n covers every sentence in $\{\phi \mid$

$L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. If $L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, then either $\phi \in \mathcal{L}$ or $\phi = \neg L\phi'$ and $\phi' \in \mathcal{L}$. If $\phi \in \mathcal{L}$, then $\phi \notin E$ implies by (C1) that ϕ is some ψ_i and $i \leq m_1$ and $\phi \in E$ implies by (C3) that ϕ is some ψ_i and $m_1+m_2 < i \leq m_1+m_2+m_3$. Otherwise, $\phi = \neg L\phi'$ and $\phi' \in \mathcal{L}$. Then $\phi' \notin E$ implies by (C1) that ϕ' is some ψ_i and $i \leq m_1$ and by (C2) ϕ is some ψ_j and $m_1 < j \leq m_1 + m_2$. On the other hand, $\phi' \in E$ implies by (C3) that ϕ' is some ψ_i and $m_1 + m_2 < i \leq m_1 + m_2 + m_3$ and thus by (C2) ϕ is some ψ_j and $m_1 + m_2 + m_3 < j \leq n$. Next we prove the following facts about the enumeration:

(F1) *For all* $i = 1, \ldots, m_1 : \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$.

Assume that there is some $i \leq m_1$ such that $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$. Let i be smallest such index. By Proposition 7.3 $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) \models_{\mathrm{cl}} \psi_i$. Assume that there exists some $c \in \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) - W$. So there is a sentence $(La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n) \rightarrow c \in \Sigma$ and $La \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$ and $L\neg L\neg b_i \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$ for all $i = 1, \ldots, n$. This is impossible because for all $i = 1, \ldots, i-1$, $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. So $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) \subseteq W$ and thus $\psi_i \in E$ which contradicts (C1).

(F2) *For all* $i = m_1 + 1, \ldots, m_1 + m_2 : \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$.

This is implied by (F1) as if $m_1 + 1 \leq i \leq m_1 + m_2$ then $\psi_i = \neg L\psi_j$ where $j \leq m_1$ and by (F1) $\psi_i = \neg L\psi_j \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$.

(F3) *For all* $i = m_1 + m_2 + 1, \ldots, m_1 + m_2 + m_3 : \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$.

Assume that there is some i such that $m_1 + m_2 < i \leq m_1 + m_2 + m_3$ and $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. Let i be smallest such index. We show that $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) \models_{\mathrm{cl}} \psi_i$ which leads to a contradiction by Proposition 7.3. As $\psi_i \in E$ by (C3), there exists a minimal $k \geq 0$ such that $E_k \models_{\mathrm{cl}} \psi_i$. If $k = 0$, then $W \models_{\mathrm{cl}} \psi_i$ and thus $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) \models_{\mathrm{cl}} \psi_i$. By Proposition 7.2 for all $k = 1, 2, \ldots$

$$E_k \models_{\mathrm{cl}} \phi \text{ if and only if } W \cup \mathrm{apply}_{D,E}(E_{k-1}) \models_{\mathrm{cl}} \phi. \tag{7.19}$$

Hence, $W \cup \mathrm{apply}_{D,E}(E_{k-1}) \models_{\mathrm{cl}} \psi_i$ holds. Let $c \in \mathrm{apply}_{D,E}(E_{k-1})$. Then there is a default $(a : b_1, \ldots, b_n/c) \in D$ such that $a \in E_{k-1}$ and $\neg b_l \notin E$ for all $l = 1, \ldots, n$. By (C3) $a = \psi_j$ such that $m_1 + m_2 < j \leq m_1 + m_2 + m_3$ and $j < i$. So $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{j-1}(\Sigma) \models_{\mathrm{ae}} \psi_j$ and thus $L\psi_j \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$. By (C1) $\neg b_l = \psi_j$ such that $j \leq m_1$. Thus by (F1) and (F2) $L\neg L\psi_j \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$. So $c \in \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma))$. Hence $\mathrm{apply}_{D,E}(E_{k-1}) \subseteq \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma))$. As $W \subseteq \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma))$, $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)) \models_{\mathrm{cl}} \psi_i$ by the monotonicity of $\models_{\mathrm{cl}}$, a contradiction. Hence (F3) holds.

(F4) *If for some* AE-*structure* $\langle I, T\rangle$, $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$ *holds, then there is a set* T' *such that* $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma)$ *holds.*

We show that if $m_1 + m_2 + m_3 < i \leq n$, then $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$ implies $\langle I, T - \{\psi_i\}\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i}(\Sigma)$. This shows that for $T' = T - \{\psi_i \mid m_1 + m_2 + m_3 < i \leq n\}$, $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma)$ holds and, hence, that (F4) holds.

Let there be an AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$. By (C4) $\psi_i = \neg L\psi_j$ such that $m_1 + m_2 < j \leq m_1 + m_2 + m_3$. By (F3) $L\psi_j \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$ and thus $\psi_j \in T$. So $\langle I, T\rangle \not\models_{\mathrm{ae}} \neg L\psi_j$ holds which implies $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. Hence $\mathbf{B}^{\varepsilon_n}_{i}(\Sigma) = \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \cup \{\neg L\psi_i\}$. Let $T' = T - \{\psi_i\}$. Then $\langle I, T'\rangle \models_{\mathrm{ae}} \mathbf{B}^{\varepsilon_n}_{i}(\Sigma)$. Let $\psi = (La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n) \rightarrow c \in \Sigma$. If $\langle I, T'\rangle \not\models_{\mathrm{ae}} c$, then $\langle I, T\rangle \not\models_{\mathrm{ae}} c$. As $\langle I, T\rangle \models_{\mathrm{ae}} \psi$, $\langle I, T\rangle \not\models_{\mathrm{ae}} La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n$ holds. So either $a \notin T$ or there is some l, $1 \leq l \leq n$ such that $\neg L\neg b_l \notin T$. As $T' \subseteq T$, either $a \notin T'$ or for some l, $\neg L\neg b_i \notin T'$. This implies $\langle I, T'\rangle \not\models_{\mathrm{ae}} La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n$. So $\langle I, T'\rangle \models_{\mathrm{ae}} \psi$. Hence $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i}(\Sigma)$.

So we have established the facts (F1)–(F4). Next we show that for all $i = 1, \ldots, m_1 + m_2 + m_3$,

$$\text{if } \psi_i \in \mathcal{L},\ \ L\psi_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \text{ if and only if } \psi_i \in E \text{ and} \qquad (7.20)$$
$$L\neg L\psi_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \text{ if and only if } \psi_i \notin E.$$

If $L\psi_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$, then $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$. By (F1) $i > m_1$ and by (C2) $i > m_1 + m_2$. As $i \leq m_1 + m_2 + m_3$, by (C3) $\psi_i \in E$. If $\psi_i \in E$, then by (C3) $i > m_1 + m_2$ and by (F3) $L\psi_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$.

Let $L\neg L\psi_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$ hold. Then $\neg L\psi_i = \psi_{i'}$ for some i' such that $i' \leq m_1 + m_2 + m_3$. As $\psi_{i'} \notin \mathcal{L}$, $m_1 < i' \leq m_1 + m_2$. Then by (C2) $i \leq m_1$ and by (C1) $\psi_i \notin E$. On the other hand, let $\psi_i \notin E$. Then by (C1) $i \leq m_1$ and by (C2) $\neg L\psi_i = \psi_{i'}$ and $m_1 < i' \leq m_1 + m_2$. Thus by (F2) $L\neg L\psi_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$.

Using this we can prove

$$\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)) = W \cup \mathrm{apply}_{D,E}(E). \qquad (7.21)$$

($\subseteq$) Let $c \in \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma))$. If $c \notin W$, there is a sentence $(La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n) \rightarrow c \in \Sigma$ such that $La \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$ and $L\neg L\neg b_l \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$ for all $l = 1, \ldots, n$. So there is a default $(a : b_1, \ldots, b_n/c) \in D$. By (C1) and (C3) $a = \psi_j$ for some $j \leq m_1+m_2+m_3$ and by (7.20) $a \in E$. Similarly, by (C1) and (C3) $\neg b_l = \psi_j$ for some $j \leq m_1+m_2+m_3$ and by (7.20) $\neg b_l \notin E$. Thus $c \in W \cup \mathrm{apply}_{D,E}(E)$.

($\supseteq$) Obviously, $W \subseteq \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma))$ holds. Let $c \in \mathrm{apply}_{D,E}(E)$. So there is a default $(a : b_1, \ldots, b_n/c) \in D$ such that $a \in E$ and $\neg b_i \notin E$. Then there is a sentence $(La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n) \rightarrow c \in \Sigma$. By (C1) and (C3) $a = \psi_j$ for some $j \leq m_1 + m_2 + m_3$ and by (7.20) $La \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$. Similarly, by (C1) and (C3) $\neg b_i = \psi_j$ for some $j \leq m_1 + m_2 + m_3$ and by (7.20) $L\neg L\neg b_i \in \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$. Thus $c \in \mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma))$ and (7.21) holds.

Next we show that for all $\phi \in \mathcal{L}_{\mathrm{ae}}$,

$$\Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \not\models_{\mathrm{ae}} \phi \text{ implies } \Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma) \not\models_{\mathrm{ae}} \phi. \tag{7.22}$$

Let $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \not\models_{\mathrm{ae}} \phi$ holds. Then there is an AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$ but $\langle I, T\rangle \not\models_{\mathrm{ae}} \phi$. By (F4) there is a set T' such that $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma)$. As $\phi \in \mathcal{L}$, $\langle I, T'\rangle \not\models_{\mathrm{ae}} \phi$. Thus $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma) \not\models_{\mathrm{ae}} \phi$. Hence (7.22) holds.

Then we show that for all $\neg L\psi_i \in \mathbf{B}^{\varepsilon_n}_{n}(\Sigma)$, $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. This implies by Proposition 5.8 that ε_n is Σ-acceptable. Let $\neg L\psi_i \in \mathbf{B}^{\varepsilon_n}_{n}(\Sigma)$. Thus $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. If $1 \leq i \leq m_1$, then by (C1) $\psi_i \notin E$. Reiter [122, Theorem 2.5] shows that $E = \mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E))$. Thus $W \cup \mathrm{apply}_{D,E}(E) \not\models_{\mathrm{cl}} \psi_i$. By (7.21) $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)) \not\models_{\mathrm{cl}} \psi_i$. By Proposition 7.3 $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. By (7.22) $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$. If $i > m_1$, then by (F2) and (F3) $i > m_1 + m_2 + m_3$. As $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ holds, there is an AE-structure $\langle I, T\rangle$ such that $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$. Since $i > m_1 + m_2 + m_3$, $\langle I, T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)$ holds. By (F4) there is a set T' such that $\langle I, T'\rangle \models_{\mathrm{ae}} \Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma)$. By (C4) $\psi_i = \neg L\psi_j$ and $m_1 + m_2 < j \leq m_1 + m_2 + m_3$. By (F3) $L\psi_j \in \mathbf{B}^{\varepsilon_n}_{i-1}(\Sigma)$ and thus $\psi_j \in T'$. So $\langle I, T'\rangle \not\models_{\mathrm{ae}} \psi_i (= \neg L\psi_j)$. Hence $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma) \not\models_{\mathrm{ae}} \psi_i$ holds.

We have shown that ε_n is a Σ-acceptable L-hierarchic enumeration of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. Then by Theorem 5.10 there is a Σ-acceptable L-hierarchic enumeration ε such that $\mathbf{SE}^{\varepsilon}(\Sigma) = \mathrm{SE}_{\Sigma}(\mathbf{B}^{\varepsilon_n}_{n}(\Sigma))$. It remains to be shown that $E = \mathbf{SE}^{\varepsilon}(\Sigma) \cap \mathcal{L}$ holds.

($\subseteq$) Let $\phi \in E$. Then by [122, Theorem 2.5] $W \cup \mathrm{apply}_{D,E}(E) \models_{\mathrm{cl}} \phi$. This implies by (7.21) $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)) \models_{\mathrm{cl}} \phi$. By Proposition 7.3 $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \models_{\mathrm{ae}} \phi$ holds. Hence $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma) \cap \mathcal{L}$.

($\supseteq$) Let $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma) \cap \mathcal{L}$. So $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{n}(\Sigma) \models_{\mathrm{ae}} \phi$. Then by (7.22) $\Sigma \cup \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma) \models_{\mathrm{ae}} \phi$ holds. By Proposition 7.3 $\mathrm{Hds}(\Sigma, \mathbf{B}^{\varepsilon_n}_{m_1+m_2+m_3}(\Sigma)) \models_{\mathrm{cl}} \phi$. Then by (7.21) $W \cup \mathrm{apply}_{D,E}(E) \models_{\mathrm{cl}} \phi$. By [122, Theorem 2.5] $E = \mathrm{Cn}_{\mathrm{cl}}(W \cup \mathrm{apply}_{D,E}(E))$. So $\phi \in E$. $\square$

Then we show that for a default theory, the L free part of an L-hierarchic expansion of the translated theory is an extension of the theory.

Lemma 7.5 *Let (D, W) be a default theory. If $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr}_{\mathrm{T}}(D, W)) \cap \mathcal{L}$ for* *some* $\mathrm{tr_T}(D,W)$*-acceptable L-hierarchic enumeration* ε*, then* E *is an extension of* (D,W)*.*

Proof. We show $E = \bigcup_{i=0}^{\infty} E_i$ which by [122, Theorem 2.1] proves that E is an extension of (D,W). Let $\Sigma = \mathrm{tr_T}(D,W)$.

($\supseteq$) First we show that for all $i = 0,1,\ldots, E_i \subseteq E$ by induction on i. This implies that $\bigcup_{i=0}^{\infty} E_i \subseteq E$ holds. Let $i = 0$. Then $E_i = W \subseteq \Sigma$ and thus $E_i \subseteq E$. Assume that $E_{i-1} \subseteq E$. Let $\phi \in E_i$. So either $E_{i-1} \models_{\mathrm{cl}} \phi$ or $\phi \in \mathrm{apply}_{D,E}(E_{i-1})$. If $E_{i-1} \models_{\mathrm{cl}} \phi$, then by the inductive hypothesis $E \models_{\mathrm{cl}} \phi$ and thus $\phi \in E$ because E is closed under $\models_{\mathrm{cl}}$. If $\phi \in \mathrm{apply}_{D,E}(E_{i-1})$, then there is a default $(a : b_1,\ldots,b_n/\phi) \in D$ such that $a \in E_{i-1}$ and $\neg b_l \notin E$ for all $l = 1,\ldots,n$. Then there is a sentence $(La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n) \rightarrow c \in \Sigma$. By the inductive hypothesis $a \in E$ and thus $La \in \mathbf{SE}^{\varepsilon}(\Sigma)$. As $\neg b_l \notin E$ for all $l = 1,\ldots,n$, then $\neg L\neg b_l \in \mathbf{SE}^{\varepsilon}(\Sigma)$ and thus $L\neg L\neg b_l \in \mathbf{SE}^{\varepsilon}(\Sigma)$ for all $l = 1,\ldots,n$. So $\phi \in \mathbf{SE}^{\varepsilon}(\Sigma) \cap \mathcal{L} = E$.

($\subseteq$) We show that for all $i = 0,1,\ldots$, if $\phi \in \mathcal{L}$ and $\Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$, then $\phi \in \bigcup_{i=0}^{\infty} E_i$ by induction on i. This implies that $E \subseteq \bigcup_{i=0}^{\infty} E_i$ holds. Let $i = 0$. Assume $\Sigma \models_{\mathrm{ae}} \phi$. By Proposition 7.3 $\mathrm{Hds}(\Sigma,\emptyset) \models_{\mathrm{cl}} \phi$ and thus $W \models_{\mathrm{cl}} \phi$. So $\phi \in E_1$. Assume that for all $\phi \in \mathcal{L}$ such that $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$, $\phi \in \bigcup_{i=0}^{\infty} E_i$. Let $\Sigma \cup \mathbf{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$ hold. If $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$ holds, then by the inductive hypothesis $\phi \in \bigcup_{i=0}^{\infty} E_i$. Hence the case where $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \phi$ holds remains to be considered. By Proposition 7.3 $\mathrm{Hds}(\Sigma, \mathbf{B}_i^{\varepsilon}(\Sigma)) \models_{\mathrm{ae}} \phi$. We show that

$$\mathrm{Hds}(\Sigma, \mathbf{B}_i^{\varepsilon}(\Sigma)) \subseteq \bigcup_{i=0}^{\infty} E_i \tag{7.23}$$

holds which implies $\phi \in \bigcup_{i=0}^{\infty} E_i$ as $\bigcup_{i=0}^{\infty} E_i$ is closed under $\models_{\mathrm{cl}}$. Let $c \in \mathrm{Hds}(\Sigma, \mathbf{B}_i^{\varepsilon}(\Sigma))$. If $c \in W = E_0$, then $c \in \bigcup_{i=0}^{\infty} E_i$. Otherwise, there is a sentence $(La \wedge L\neg L\neg b_1 \wedge \cdots \wedge L\neg L\neg b_n) \rightarrow c \in \Sigma$ and $La \in \mathbf{B}_i^{\varepsilon}(\Sigma)$ and $L\neg L\neg b_l \in \mathbf{B}_i^{\varepsilon}(\Sigma)$ for all $l = 1,\ldots,n$. So there is a default rule $(a : b_1,\ldots,b_n/c) \in D$. We show that $a \in \bigcup_{i=0}^{\infty} E_i$ and that $\neg b_l \notin E$ for all $l = 1,\ldots,n$ which shows that $c \in \mathrm{apply}_{D,E}(\bigcup_{i=0}^{\infty} E_i) \subseteq \bigcup_{i=0}^{\infty} E_i$. If $La \in \mathbf{B}_i^{\varepsilon}(\Sigma)$, then $a = \psi_j$ for some $j \leq i$ and $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j$ holds. By the inductive hypothesis $a = \psi_j \in \bigcup_{i=0}^{\infty} E_i$. If $L\neg L\neg b_l \in \mathbf{B}_i^{\varepsilon}(\Sigma)$, then $\neg L\neg b_l = \psi_j$, $j \leq i$ and $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j$. Here the fact that the expansion is induced by an L-hierarchic enumeration becomes crucial. As ε is L-hierarchic, $\neg b_l = \psi_{j'}$ such that $j' < j \leq i$. Assume $\Sigma \cup \mathbf{B}_{j'-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_{j'}$. Then $L\psi_{j'} \in \mathbf{B}_{j-1}^{\varepsilon}(\Sigma)$. This implies that $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon}(\Sigma)$ is inconsistent because $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j (= \neg L\neg b_l)$ but $(L\neg b_l =) L\psi_{j'} \in \mathbf{B}_{j-1}^{\varepsilon}(\Sigma)$. Thus $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma)$ would be inconsistent contradicting our assumption that $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \phi$ holds. Thus it must be the case that $\Sigma \cup \mathbf{B}_{j'-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_{j'}$. But then as ε is Σ-acceptable, $\psi_j' = b_l \notin E$. Thus $c \in \bigcup_{i=0}^{\infty} E_i$. So (7.23) holds which implies that $\phi \in \bigcup_{i=0}^{\infty} E_i$ as $\bigcup_{i=0}^{\infty} E_i$ is closed under $\models_{\mathrm{cl}}$. $\square$

With the aid of the two lemmata, the proof of the main result is straightforward.

Theorem 7.6 *Let (D, W) be a default theory where D is a finite set. Then $E \subseteq \mathcal{L}$ is an extension of (D, W) if and only if $E = \Delta \cap \mathcal{L}$ for some L-hierarchic expansion Δ of* $\mathrm{tr_T}(D, W)$.

Proof. ($\Rightarrow$) Let $E \subseteq \mathcal{L}$ be an extension of (D, W). Then as D is finite, by Lemma 7.4 there is a $\mathrm{tr_T}(D, W)$-acceptable L-hierarchic enumeration ε such that $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr_T}(D, W)) \cap \mathcal{L}$.

($\Leftarrow$) Let $E = \mathbf{SE}^{\varepsilon}(\mathrm{tr_T}(D, W)) \cap \mathcal{L}$ for a $\mathrm{tr_T}(D, W)$-acceptable L-hierarchic enumeration ε. By Lemma 7.5 E is an extension of (D, W). $\square$

The theorem expresses a close correspondence between autoepistemic logic and default logic. The correspondence can be utilized, e.g., to carry complexity and decidability results from one formalism to another. As in the case of autoepistemic logic, the interesting decision problems in default logic are the extension existence problem, brave default reasoning (is there some extension of a default theory containing a given sentence) and cautious default reasoning (is a given sentence in every extension of a default theory). By Theorem 7.6 these problems can be reduced to corresponding decision problems in autoepistemic logic based on L-hierarchic expansions. By Theorem 6.22 each of the problems is decidable if and only if the underlying first-order consequence relation is decidable. Theorem 7.6 also implies that a decision method for autoepistemic logic based on L-hierarchic expansions can be used as a decision method for default logic and that a decision method for default logic can be used as a decision method for a fragment of autoepistemic logic.

As default reasoning can be reduced to autoepistemic reasoning, by Corollary 6.24 the extension existence problem and brave default reasoning are in Σ_2^p and cautious default reasoning is in Π_2^p in the propositional case. Respective completeness results can be obtained using a similar reduction as that employed in the proof of Theorem 6.25. As discussed in Chapter 6, the completeness results have already been established by Gottlob [40] and Stillman [139]. Kautz and Selman [59], Stillman [138], and Ben-Eliyahu and Dechter [4, 5] have obtained further complexity results on restricted subclasses of default theories. These results are directly applicable to autoepistemic logic because the subclasses can be seen as subclasses of autoepistemic premises under the translation (7.11). Several decision methods for (decidable fragments of) default logic have been proposed [122, 26, 50, 132, 150, 4, 5].

By Theorem 7.6 we obtain an alternative view of default reasoning as a sequence of introspection steps. As in the case of autoepistemic reasoning, two independent sources of complexity in default reasoning can be identified: required classical reasoning and determining a suitable order of introspection. Complexity studies on restricted subclasses of default reasoning show that even though the form of default rules and sentences in a default theory is highly limited, the decision problems in default reasoning remain **NP**-hard. This is a consequence of the fact that restricting the form of expressions in a default theory addresses the source of complexity related to required classical reasoning. The notion of stratification offers a restriction which can be shown to have a direct impact on the complexity of finding an appropriate sequence of introspection steps. This notion can be applied straightforwardly to default theories. The iterative decision method developed for stratified autoepistemic reasoning [112] can be employed in stratified default reasoning. If the form of expressions in a stratified default theory is restricted further so that required classical reasoning can be done in polynomial time, tractable subclasses of stratified default reasoning can be obtained [112].

7.2 Circumscription

Circumscription is based on the idea of considering minimal models. Circumscriptive conclusions from the set of premises are the sentences true in every minimal model of the original premises. There is obviously many different notions of a minimal model. Consequently, there is a large number of variants of circumscription starting from the work of McCarthy [92, 93]. Etherington's [26] and Brewka's [11] books are good sources on circumscription.

Statements of the type "birds typically fly" can be formalized using circumscription in terms of abnormality. A bird flies unless it is abnormal in some sense:

$$\forall x((b(x) \land \neg ab(x)) \rightarrow f(x)).$$

The idea is to minimize abnormality to only those objects which are known to be abnormal.

The relationships between default logic and circumscription as well as between autoepistemic logic and circumscription have been investigated by a number of researchers [25, 26, 47, 74, 63]. Imielinski [47] points out that default rules cannot not be translated modularly into circumscription. This due to the fact that circumscription cannot capture the notion of provability or logical consequence required to translate a default rule.

We show that Imielinski's [47] argument applies also to autoepistemic logic and thus circumscription cannot modularly capture autoepistemic reasoning because circumscription cannot distinguish between the idea of believing a sentence and the truth of the sentence. We consider the propositional case. We show that autoepistemic reasoning cannot be modularly translated into order semantics. In order semantics we have a (reflexive and transitive) relation $\leq$ on the propositional interpretations. Order semantics is a general form of circumscription where the relation $\leq$ gives the notion of minimality: if S is a set of propositional interpretations, then $I \in S$ is a $\leq$-minimal interpretation in S if for all $I' \in S$, $I' \leq I$ implies $I \leq I'$. The corresponding consequence relation $\models_{\leq}$ can be defined in the following way. For a set of sentences $\Sigma \subseteq \mathcal{L}$,

$$\Sigma \models_{\leq} \phi \text{ if and only if } \phi \text{ is true in every } \leq\text{-minimal model of } \Sigma. \tag{7.24}$$

For representing autoepistemic reasoning in order semantics, we are looking for a translation of sentences Σ containing L operators into a set of propositional sentences $T(\Sigma)$ and an order relation $\leq$ on the propositional interpretations such that a propositional sentence ϕ is contained in every stable expansion exactly when it is true in every $\leq$-minimal model of $T(\Sigma)$. The intuition is that the set of propositional sentences and the order relation would capture the beliefs of the agent. We say that a set of sentences $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ is *modularly representable by order semantics* if and only if there is a set of sentences $T(\Sigma) \subseteq \mathcal{L}$ and an order relation $\leq$ such that for every set $O \subseteq \mathcal{L}$ and for every sentence $\phi \in \mathcal{L}$,

$$O \cup \Sigma \mathrel{|\!\sim_{\mathrm{c}}} \phi \text{ if and only if } O \cup T(\Sigma) \models_{\leq} \phi. \tag{7.25}$$

Modularity means that if the premises are extended by sentences not containing L operators, the translation need not to be changed.

Theorem 7.7 *There are sets of sentences in propositional autoepistemic logic which are not modularly representable by order semantics.*

Proof. Consider the set $B = \{L\phi \rightarrow \psi\}$ where ϕ and ψ are atomic. Assume that B is modularly representable by order semantics. Thus there is a translation $T(B)$ and an order relation $\leq$ such that (7.25) holds. The set B has exactly one stable expansion $\mathrm{SE}_B(\{\neg L\phi\})$. Hence $B \mathrel{|\!\not\sim_{\mathrm{c}}} \phi \rightarrow \psi$ and thus $T(B) \not\models_{\leq} \phi \rightarrow \psi$. So there is a minimal model I' of $T(B)$ such that $I' \models_{\mathrm{cl}} \phi$ but $I' \not\models_{\mathrm{cl}} \psi$. As I' is minimal, for all models I of $T(B)$, $I \leq I'$ implies $I' \leq I$. Let $O = \{\phi\}$. Then I' is a model of $O \cup T(B)$. Let I be a model of $O \cup T(B)$ such that $I \leq I'$. Then I is a model of $T(B)$ and as I' is a minimal model of $T(B)$, $I' \leq I$ holds. Hence for all models I of $O \cup T(B)$, if $I \leq I'$, then $I' \leq I$. This implies that I' is a minimal model of $O \cup T(B)$. As $I' \not\models_{\mathrm{cl}} \psi$, $O \cup T(B) \not\models_{\leq} \psi$ although $O \cup B \mathrel{|\!\sim_{\mathrm{c}}} \psi$ and we are in contradiction with (7.25). So $B = \{L\phi \rightarrow \psi\}$ is not modularly representable by semantics based on minimal models. In fact we have shown that $\{\phi\} \cup T(\{L\phi \rightarrow \psi\}) \models_{\leq} \psi$ implies $T(\{L\phi \rightarrow \psi\}) \models_{\leq} \phi \rightarrow \psi$, a kind of a deduction theorem. This property is an immediate consequence of the minimal model approach of circumscription. It indicates that the minimal model approach cannot distinguish between the case where ϕ is provable and the case where ϕ is true in a model. □

There is nothing special in the counter example in the proof of the previous theorem. The same argument can be repeated for any reasonable variant of autoepistemic logic. Hence, autoepistemic logic based on enumeration-based or L-hierarchic expansions is not modularly representable by order semantics. On the other hand, in autoepistemic logic it is possible to capture minimal models of a set of premises by adding new autoepistemic sentences to that effect. We consider the propositional case and a notion of minimal models where the models are minimized with respect to a given set P of atomic sentences keeping another set F of atomic sentences fixed and letting all the other atomic sentences vary during the minimization. Propositional interpretations can be seen as sets of atomic sentences and the corresponding order relation $\leq$ on the propositional interpretations can be defined as follows given the sets P and F of atoms. For any propositional interpretations I and I', $I \leq I'$ if and only if $I \cap F = I' \cap F$ and $I \cap P \subseteq I' \cap P$. So a model I of Σ is $\langle P, F\rangle$-minimal if there is no model I' of Σ containing the same fixed atomic sentences from F but for which the atomic sentences in P are properly included in the atomic sentences from P in I. This notion of minimality is a propositional version of the minimality in parallel circumscription [71] where the atomic sentences P are minimized and the atomic sentences not in F or P are allowed to vary. Furthermore, Gelfond and et al. [36] show that parallel circumscription corresponds to a particular form of the negation as failure rule (the extended closed world assumption). We use the following notations. The complement $\overline{\phi}$ of a formula ϕ of the form $\neg\phi'$ is ϕ' and for other formulae it is $\neg\phi$. For a set of formulae P we define

$$\begin{aligned} \mathrm{MIN}(P) &= \{\neg L\phi \rightarrow \overline{\phi} \mid \phi \in P\}, && (7.26) \\ \mathrm{FIX}(P) &= \mathrm{MIN}(P) \cup \mathrm{MIN}(\neg P). && (7.27) \end{aligned}$$

Definition 7.8 *Let P and F be sets of atomic sentences and $\Sigma \subseteq \mathcal{L}$. A model I of $\Sigma \subseteq \mathcal{L}$ is $\langle P, F\rangle$-minimal if and only if there is no model I' of Σ such that $I \cap F = I' \cap F$ and $I' \cap P \subset I \cap P$.*

First we show that for every $\langle P, F\rangle$-minimal model I of Σ, there is a corresponding stable expansion.

Proposition 7.9 *Let P and F be sets of atomic sentences and $\Sigma \subseteq \mathcal{L}$. For every $\langle P, F\rangle$-minimal model I of Σ, there is a stable expansion Δ of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ such that $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$.*

Proof. Let I be a $\langle P, F\rangle$-minimal model of Σ. We show first that

$$\begin{aligned} \Lambda &= \{L\phi \mid \phi \in P \cup F \cup \neg F, I \models_{\mathrm{cl}} \phi\} \cup \\ &\quad \{\neg L\phi \mid \phi \in P \cup F \cup \neg F, I \not\models_{\mathrm{cl}} \phi\} \end{aligned} \tag{7.28}$$

is $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$-full. We start by establishing the following fact. Let $\Lambda^+ = \{\psi \mid L\psi \in \Lambda\}$. Then

$$\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda. \tag{7.29}$$

As Λ contains no quasi-atom and its negation, $\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \Lambda$. As $\Sigma \subseteq \mathcal{L}$ and $I \models_{\mathrm{cl}} \Sigma$, $\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \Sigma$. Let $\neg L\psi \to \neg\psi \in \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ where $\psi \in P \cup F$. Then $\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \neg L\psi \to \neg\psi$ because if $\langle I, \Lambda^+\rangle \not\models_{\mathrm{ae}} \neg\psi$, then $I \models_{\mathrm{cl}} \psi$ which implies that $\psi \in \Lambda^+$ and thus $\langle I, \Lambda^+\rangle \not\models_{\mathrm{ae}} \neg L\psi$. Let $\neg L\neg\psi \to \psi \in \mathrm{FIX}(F)$ where $\psi \in F$. Then $\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \neg L\neg\psi \to \psi$ because if $\langle I, \Lambda^+\rangle \not\models_{\mathrm{ae}} \psi$, then $I \models_{\mathrm{cl}} \neg\psi$ which implies $\neg\psi \in \Lambda^+$ and thus $\langle I, \Lambda^+\rangle \not\models_{\mathrm{ae}} \neg L\neg\psi$. Hence (7.29) holds.

Now we show that for all $L\phi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F))$,

$$\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{ae}} \phi \quad \text{if and only if} \quad L\phi \in \Lambda \text{ and} \tag{7.30}$$

$$\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \not\models_{\mathrm{ae}} \phi \quad \text{if and only if} \quad \neg L\phi \in \Lambda. \tag{7.31}$$

This implies that Λ is $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$-full because for all $L\phi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F))$, $\phi \in \mathcal{L}$ and thus $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{ae}} \phi$ if and only if $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{L}} \phi$. Let $L\phi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F))$. Then $\phi \in P \cup F \cup \neg F$ and thus either $L\phi \in \Lambda$ or $\neg L\phi \in \Lambda$. First we establish (7.30). ($\Rightarrow$) Let $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{ae}} \phi$ hold. Then by (7.29) $\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \phi$ which implies $I \models_{\mathrm{cl}} \phi$. Thus $L\phi \in \Lambda$. ($\Leftarrow$) Let $L\phi \in \Lambda$. Then $I \models_{\mathrm{cl}} \phi$ and $\phi \in P \cup F \cup \neg F$. If $\phi \in F \cup \neg F$, then $\neg L\overline{\phi} \in \Lambda$ and thus $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{ae}} \phi$ as $\neg L\overline{\phi} \to \phi \in \mathrm{FIX}(F)$. Let $\phi \in P$. We show that $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \not\models_{\mathrm{ae}} \phi$ implies that there is a model I' of Σ such that $I' \cap F = I \cap F$ and $I' \cap P \subset I \cap P$. Assume that $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \not\models_{\mathrm{ae}} \phi$. Then there is an AE-structure $\langle I', T\rangle$ such that $\langle I', T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda$ but $\langle I', T\rangle \not\models_{\mathrm{ae}} \phi$. So I' is a model of Σ. Let $\psi \in I' \cap (P \cup F)$. Assume that $\psi \notin I$. Then $\neg L\psi \in \Lambda$. But as $\langle I', T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda$ and $\neg L\psi \to \neg\psi \in \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$, $\langle I', T\rangle \models_{\mathrm{ae}} \neg\psi$ and thus $\psi \notin I'$, a contradiction. Hence $\psi \in I' \cap (P \cup F)$ implies $\psi \in I$. Thus $I' \cap F \subseteq I \cap F$ and $I' \cap P \subseteq I \cap P$. Let $\psi \in I \cap F$. Then $\neg L\neg\psi \in \Lambda$. But as $\langle I', T\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda$ and $\neg L\neg\psi \to \psi \in \mathrm{FIX}(F)$, $\langle I', T\rangle \models_{\mathrm{ae}} \psi$ and thus $\psi \in I'$. Hence $I \cap F = I' \cap F$. Furthermore, $\langle I', T\rangle \not\models_{\mathrm{ae}} \phi$ implies $\phi \in (I - I') \cap P$. Hence, $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \not\models_{\mathrm{ae}} \phi$ implies that there is a model I' of Σ such that $I' \cap F = I \cap F$ and $I' \cap P \subset I \cap P$. As I is a $\langle P, F\rangle$-minimal model of Σ, $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{ae}} \phi$ must hold. So we have established (7.30).

Then we show that (7.31) holds. If $\neg L\phi \in \Lambda$, then $I \not\models_{\mathrm{cl}} \phi$. By (7.29) $\langle I, \Lambda^+\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda$ but $\langle I, \Lambda^+\rangle \not\models_{\mathrm{ae}} \phi$. Thus $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \not\models_{\mathrm{ae}} \phi$ holds. Let $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \not\models_{\mathrm{ae}} \phi$ hold. Assume that $\neg L\phi \notin \Lambda$. Then $L\phi \in \Lambda$ which implies that $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \cup \Lambda \models_{\mathrm{ae}} \phi$ holds, a contradiction. Thus (7.31) holds.

Hence Λ is a $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$-full set. By Lemma 3.7 there is a stable expansion Δ of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Let $\neg L\psi \rightarrow \neg\psi \in \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ where $\psi \in P \cup F$. Then $\langle I, \Delta\rangle \models_{\mathrm{ae}} \neg L\psi \rightarrow \neg\psi$ because if $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \neg\psi$, then $I \models_{\mathrm{cl}} \psi$ and thus $L\psi \in \Lambda$ which implies $\psi \in \Delta$ and hence $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \neg L\psi$. Let $\neg L\neg\psi \rightarrow \psi \in \mathrm{FIX}(F)$. Then $\langle I, \Delta\rangle \models_{\mathrm{ae}} \neg L\neg\psi \rightarrow \psi$ because if $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \psi$, then $I \models_{\mathrm{cl}} \neg\psi$ and thus $L\neg\psi \in \Lambda$ which implies $\neg\psi \in \Delta$ and hence $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \neg L\neg\psi$. Thus $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$. □

It turns out that the propositional part of a model of a stable expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ is always $\langle P, F\rangle$-minimal.

Proposition 7.10 *Let P and F be sets of atomic sentences and $\Sigma \subseteq \mathcal{L}$. For every stable expansion Δ of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$, if $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$, then I is a $\langle P, F\rangle$-minimal model of Σ.*

Proof. Let Δ be a stable expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ and let $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ hold. Then I is a model of Σ. Assume that there exists a model I' of Σ such that $I \cap F = I' \cap F$ and $I' \cap P \subset I \cap P$. So there exists $\phi \in P$ such that $\phi \in I - I'$. Clearly, $\langle I', \Delta\rangle \models_{\mathrm{ae}} \Sigma$. Let $\neg L\psi \rightarrow \neg\psi \in \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ where $\psi \in P \cup F$. Then $\langle I', \Delta\rangle \models_{\mathrm{ae}} \neg L\psi \rightarrow \neg\psi$ because if $\langle I', \Delta\rangle \not\models_{\mathrm{ae}} \neg\psi$, then $\psi \in I'$ which implies $\psi \in I$. But then $\langle I, \Delta\rangle \models_{\mathrm{ae}} \psi$ and as $\langle I, \Delta\rangle \models_{\mathrm{ae}} \neg L\psi \rightarrow \neg\psi$, $\psi \in \Delta$ and hence $\langle I', \Delta\rangle \not\models_{\mathrm{ae}} \neg L\psi$. Let $\neg L\neg\psi \rightarrow \psi \in \mathrm{FIX}(F)$. Then $\langle I', \Delta\rangle \models_{\mathrm{ae}} \neg L\neg\psi \rightarrow \psi$ because if $\langle I', \Delta\rangle \not\models_{\mathrm{ae}} \psi$, then $\psi \notin I' \cap F = I \cap F$. But then $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \psi$ and as $\langle I, \Delta\rangle \models_{\mathrm{ae}} \neg L\neg\psi \rightarrow \psi$, $\neg\psi \in \Delta$ and hence $\langle I', \Delta\rangle \not\models_{\mathrm{ae}} \neg L\neg\psi$. Thus $\langle I', \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ but $\langle I', \Delta\rangle \not\models_{\mathrm{ae}} \phi$ as $\phi \notin I'$. So $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \not\models_{\Delta} \phi$ and by Proposition 2.6 $\phi \notin \Delta$. This leads to a contradiction because $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ holds. Then as $\neg L\phi \rightarrow \neg\phi \in \mathrm{MIN}(P)$, $\langle I, \Delta\rangle \models_{\mathrm{ae}} \neg L\phi \rightarrow \neg\phi$. But as $\phi \in I \cap P$, $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \neg\phi$. Thus $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \neg L\phi$ and $\phi \in \Delta$, a contradiction. Hence I is a $\langle P, F\rangle$-minimal model of Σ. □

Using these results it is not hard to prove that the sentences true in every minimal model of the premises Σ are exactly those sentences contained in every expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$.

Theorem 7.11 *Let P and F be sets of atomic sentences and $\Sigma \subseteq \mathcal{L}$. For every sentence $\phi \in \mathcal{L}$, ϕ is true in every $\langle P, F\rangle$-minimal model of Σ if and only if ϕ is in every stable expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$.*

Proof. ($\Rightarrow$) Let ϕ be true in every $\langle P, F\rangle$-minimal model of Σ. Assume there is a stable expansion Δ of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ such that $\phi \notin \Delta$. Thus by Proposition 2.6 $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \not\models_{\Delta} \phi$ and there is a propositional interpretation I such that $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ and $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \phi$. By Proposition 7.10 I is $\langle P, F\rangle$-minimal model of Σ and $I \not\models_{\mathrm{cl}} \phi$, a contradiction. Thus ϕ is in every stable expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$.

($\Leftarrow$) Let ϕ be in every stable expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$. Assume that there is a $\langle P, F\rangle$-minimal model I of Σ such that $I \not\models_{\mathrm{cl}} \phi$. By Proposition 7.9 there is a stable expansion Δ of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ such that $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$. Then $\langle I, \Delta\rangle \models_{\mathrm{ae}} \Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ and $\langle I, \Delta\rangle \not\models_{\mathrm{ae}} \phi$ as $\phi \in \mathcal{L}$. So $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F) \not\models_{\Delta} \phi$ and thus $\phi \notin \Delta$, a contradiction. So ϕ is true in every $\langle P, F\rangle$-minimal model of Σ. □

By Theorem 5.13 Moore style stable expansions and L-hierarchic expansions of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$ coincide if $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F))$ is finite. This implies that (propositional parallel) circumscription can be captured in autoepistemic logic based on L-hierarchic expansions.

Corollary 7.12 *Let P and F be finite sets of atomic sentences and $\Sigma \subseteq \mathcal{L}$. For every sentence $\phi \in \mathcal{L}$ holds that ϕ is true in every $\langle P, F\rangle$-minimal model of Σ if and only if ϕ belongs to every L-hierarchic expansion of $\Sigma \cup \mathrm{MIN}(P) \cup \mathrm{FIX}(F)$.*

Etherington [24] reports similar results of minimization using default logic in the first-order case but without any fixed predicates and with a finite, fixed domain. Konolige [63] shows how to achieve minimization in the case of non-finite domains using a variant of autoepistemic logic which allows quantification into the scope of L operators. He employs the schemes (7.26) and (7.27) to minimize and fix predicates.

As the results above show, autoepistemic logic can capture the idea of minimizing the models of a set of premises which is the basis of circumscription. We have demonstrated that model minimization approaches cannot distinguish between the idea of believing a sentence and the truth of the sentence in a modular fashion. Furthermore, the minimal model approaches are *cautious* forms of reasoning: only those sentences which are common to all stable expansions are inferred. Autoepistemic reasoning is brave: there might be many competing expansions. Autoepistemic logic can make nonmonotonic conclusions about the equality predicate but in circumscription this is not straightforwardly possible [25, 119]. Thus it seems that the minimal model approaches such as circumscription are not powerful enough to capture autoepistemic reasoning.

Theorem 7.11 shows that parallel circumscription reduces to cautious autoepistemic reasoning. Hence, the problem of deciding whether a propositional sentence is a circumscriptive consequence of a set of propositional sentence is in Π_2^p by Corollary 6.24. Eiter and Gottlob [20] have shown that in fact the problem is Π_2^p-complete. By Theorem 7.11 a decision method for cautious autoepistemic reasoning can be used for circumscriptive reasoning and a circumscriptive theorem prover can be employed as a decision method for cautious autoepistemic reasoning for a subclass of autoepistemic logic. By Corollary 7.12 this is true also for cautious autoepistemic reasoning based on L-hierarchic expansions. As cautious autoepistemic reasoning and circumscriptive reasoning are both Π_2^p-complete, there are polynomial time reductions between the problems. However, the mapping from autoepistemic reasoning to circumscriptive reasoning is not modular as shown by Theorem 7.7. Cadoli and Lenzerini [12] establish complexity results on limited subclasses of circumscriptive reasoning including some tractability results. General circumscription is highly uncomputable [126]. Several methods for automating circumscriptive reasoning in some limited cases have been proposed [71, 115, 38, 44].

7.3 McDermott and Doyle Style Nonmonotonic Modal Logics

McDermott and Doyle's [96] nonmonotonic logic was one of the first logics proposed as a formalization of nonmonotonic reasoning. Autoepistemic logic was introduced as a reconstruction of this logic to avoid some deficiences of the logic. Thus McDermott and Doyle's [96] nonmonotonic logic can be regarded an origin of autoepistemic logic in the form proposed by Moore. It is illuminating to have a closer look at the basic motivation of the logic. McDermott and Doyle augmented first-order logic with a modal operator M. The intended meaning of a formula $M\phi$ was that ϕ is consistent with everything else that is known. They would have liked to capture nonmonotonic reasoning with an inference rule of the form:

$$\text{if } \Sigma \nvdash \neg\phi, \text{ then } \Sigma \vdash M\phi. \tag{7.32}$$

The rule states that if $\neg\phi$ is not derivable from Σ, then $M\phi$ is. But the rule (7.32) is circular. On one hand, the derivability relation $\vdash$ has to be known completely in order to determine that the rule is applicable, i.e. to determine that $\neg\phi$ is not derivable. But on the other hand, the derivability relation is defined partly by the rule itself. The informal semantics of the operator M also has this flavour of circularity: "everything else that is known" may well depend on what is consistent. In order to devise a well-founded definition, McDermott and Doyle formalized correct nonmonotonic conclusions as a fixed point of the operator NM_Σ which can be written in the following form:

$$NM_\Sigma(\Delta) = \{\phi \mid \Sigma \cup \{M\phi \mid \neg\phi \notin \Delta\} \vdash \phi\} \tag{7.33}$$

where $\vdash$ is the first-order consequence relation treating the new $M\phi$ formulae as atomic formulae. Thus Δ is a possible set of correct nonmonotonic conclusions from Σ if it is the set of first-order consequences of Σ and those $M\phi$ formulae for which $\neg\phi$ is not a nonmonotonic conclusion from Σ.

The logic of McDermott and Doyle does not capture nonmonotonic reasoning in a satisfactory way [96, 100]. For example, the set of premises

$$\Sigma = \{\neg p, Mp\} \tag{7.34}$$

has a consistent fixed point. We can write the fixed point equation defining the correct nonmonotonic conclusions in McDermott and Doyle's nonmonotonic logic in the language of autoepistemic logic where M denotes $\neg L \neg$ as follows.

$$\Delta = \{\phi \mid \Sigma \cup \neg L\overline{\Delta} \vdash \phi\} \tag{7.35}$$

Thus the correct conclusions from Σ are the first-order consequences of Σ and the disbeliefs obtained by negative introspection from the set of correct conclusions. McDermott and Doyle's nonmonotonic logic does not contain any form of positive introspection. This leads to peculiarities. For example, in the case of premises (7.34) as $L\neg p$ cannot be inferred from $\neg p$, the fixed point remains consistent. McDermott [95] proposed that missing positive introspection could be achieved by adding some new axioms and inference rules, i.e. by strengthening the underlying derivability relation $\vdash$ used in the fixed point equation (7.35). McDermott [95] considered modal logics **T**, **S4**, and **S5** as the basis of nonmonotonic reasoning. (For example, Chellas [14] and Hughes and Cresswell [46] are good sources on modal logics.) The work was extented first by Shvarts [135] and then by Marek et al. [80]. Using Equation (7.35), nonmonotonic modal logics based on monotonic modal logics are defined in the following way as discussed already in Chapter 4. Given a classical monotonic modal logic S, a solution to the fixed point equation (7.35) where the derivability relation $\vdash$ is that of the logic S is called an *S-expansion* of the premises Σ and the nonmonotonic consequences of Σ can be defined, for example, as the intersection of all S-expansions of Σ.

Note that the difference between this approach and Moore's autoepistemic logic is that Moore modified the fixed point equation (7.35) by adding a direct rule for positive introspection saying that if ϕ is belief, so is $L\phi$. In Moore's fixed point equation

$$\Delta = \{\phi \mid \Sigma \cup L\Delta \cup \neg L\overline{\Delta} \vdash \phi\} \tag{7.36}$$

it is sufficient to consider only first-order consequences so that the derivability relation $\vdash$ treats $L\phi$ formulae like atomic formulae. In McDermott and Doyle style nonmonotonic logics positive introspection is achieved by using appropriate modal axioms and the necessitation rule (4.16).

Marek et al. [80] offer a unified treatment of this approach in the propositional case. They show that the underlying monotonic modal logic plays a secondary role in the definition of nonmonotonic reasoning based on the fixed point equation (7.35): several monotonic modal logics induce the same nonmonotonic system, i.e. expansions coincide for a range of underlying modal logics. For example, for every modal logic S between **5** and **KD45**, an S-expansion is a stable expansion.

Theorem 7.13 (Marek et al. [80]) *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$. For every modal logic S such that $\mathbf{5} \subseteq S \subseteq \mathbf{KD45}$, a consistent set of sentences Δ is an S-expansion of Σ if and only if Δ is a stable expansion of Σ.*

This correspondence does not hold for inconsistent stable expansions. For example, let $\Sigma = \{\neg L\phi\}$, where ϕ is atomic. Then $\mathcal{L}_{\mathrm{ae}}$ is a stable expansion of Σ but $\mathcal{L}_{\mathrm{ae}}$ does not satisfy the fixed point equation for an S-expansion. If the set of premises is taken from a restricted class of formulae, even larger ranges exist.

Theorem 7.14 (Marek et al. [80]) *Let Σ contain no L operators or be strongly stratified [80]. Then for every modal logic such that $\mathbf{N} \subseteq S$ and $S \subseteq \mathbf{KD45}$ or $S \subseteq \mathbf{S4}$, Δ is an S-expansion if and only if Δ is the unique stable expansion of Σ.*

These results imply that full sets can be used directly to provide finitary characterizations of S-expansions for modal logics S not containing the axiom schema **T**. For example, let S be any modal logic contained in **KD45**. Then an S-expansion is a stable expansion by Theorem 7.13 and by the fact that if S and T are two modal logics such that $S \subseteq T \subseteq \mathbf{S5}$, then each S-expansion is a T-expansion [80]. Hence for every S-expansion Δ of Σ there is a unique Σ-full set Λ such that $\Delta = \mathrm{SE}_{\Sigma}(\Lambda)$. For a modal logic S contained in **KD45** but which does not include the axiom schema **5**, there are full sets Λ such that the stable expansion $\mathrm{SE}_{\Sigma}(\Lambda)$ induced by Λ is not an S-expansion. This is the case, e.g., for the modal logic **N**. However, a finitary characterization for **N**-expansions can be obtained by tightening the groundedness condition for full sets. We say that a Σ-full set Λ is **N**-*grounded* if $\Sigma \cup \Lambda^- \vdash_{\mathbf{N}} \phi$ for every $L\phi \in \Lambda^+$ where $\Lambda^- = \{\neg L\chi \in \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma) \mid \Sigma \cup \Lambda \not\models_{\mathrm{L}} \chi\}$ and $\Lambda^+ = \{L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) \mid \Sigma \cup \Lambda \models_{\mathrm{L}} \chi\}$. It is straightforward to show that for each **N**-grounded Σ-full set Λ, $\mathrm{SE}_{\Sigma}(\Lambda)$ is the unique **N**-expansion Δ such that $\Lambda \subseteq L\Delta \cup \neg L\overline{\Delta}$. Gottlob [40] introduces a similar finitary characterization for **N**-expansions. Gottlob's characterization employs the earlier full set approach [105, 106, 107] and uses the same extra groundedness condition to eliminate full sets not corresponding to **N**-expansions. Marek et al. [80] present a finitary characterization for **N**-expansions using the approach of Shvarts [135] which captures consistent **N**-expansions. The new characterization offers a more compact characterizing set than those of Gottlob and Marek et al. if the set of premises contains nested modal operators. This is because full sets do not contain $L\phi$ subformulae of the premises which appear only in the scope of another L operator but in the other characterizing sets they are included. As **N**-expansions and enumeration-based expansions coincide, all these characterizations can be used for enumeration-based expansions but it is not clear how to refine them to capture L-hierarchic expansions.

When the underlying modal logic S contains the axiom schema **T** resulting S-expansions are not necessarily Moore style stable expansions. Consequently, there are S-expansions not corresponding to a full set. A straightforward idea would be to adjust the groundedness condition for full sets to reflect the change in the underlying logic and require that for Λ to be Σ-full, $\Sigma \cup \{\neg L\chi \in \Lambda\} \vdash_S \phi$ holds for every $L\phi \in \Lambda$. However, in the presence of the schema **T** consistency needs special attention. For example, consider the set $\Sigma = \{\neg p, Lp\}$ and a set $\Lambda = \{Lp\}$. Now $\Sigma \vdash_S p$ if S includes the schema **T** and Σ is (S-)inconsistent. However, $\Sigma \cup \Lambda$ is propositionally consistent. Shvarts [135] uses in his characterizing sets consistency conditions where also formulae ϕ for each $L\phi$ in the characterizing set are included. This handles the inconsistency problem and leads to finitary characterizations of S-expansions for several underlying modal logics S containing the axiom schema **T**.

Example 4.3 indicates that **N**-expansions are not necessarily stable set minimal. This feature is shared by every nonmonotonic modal logic based on a monotonic modal logic contained in the modal logic **S5** and containing the necessitation rule, e.g. by every nonmonotonic system discussed by Marek et al. [80]. This is because S-expansions have the property that if S and T are two modal logics such that $S \subseteq T \subseteq \mathbf{S5}$, then each S-expansion of Σ is a T-expansion of Σ [80].

Tiomkin and Kaminski [140, 141] propose a modification to the fixed point equation of McDermott and Doyle where negative introspection is restricted only to L free formulae

$$\Delta = \{\phi \mid \Sigma \cup \neg L(\overline{\Delta} \cap \mathcal{L}) \vdash_S \phi\}. \qquad (7.37)$$

A solution to this equation is called a ground S-expansion. Kaminski [56], Truszczyński [146], Konolige [65], and Schwarz [128] have continued the research. Truszczyński studies a still more general setting where negative introspective is restricted into a given set X not necessary the set of L free formulae. Konolige considers the properties of the modal system S and presents a minimal modal system such that the resulting ground expansions are stable sets containing the premises. Schwarz studies the effect of restricting introspection in McDermott and Doyle style logics to formulae of bounded depth of nesting of modal operators. Ground expansions have some properties not present in stable expansions in autoepistemic logic. For example, introducing a new name in the premises might change the set of ground expansions. Also Schwarz [128] discusses this feature.

Example 7.3 Let $\Sigma = \{L\phi \rightarrow \phi\}$ where ϕ is atomic. The premise Σ has a single ground **K45**-expansion not containing ϕ. Let $\Sigma' = \{L\phi \rightarrow \phi, L\phi \leftrightarrow \psi\}$ where also ψ is atomic. It turns out that ϕ and ψ follow from Σ' and a result of negative introspection $\neg L \neg \psi$ using only the necessitation rule (4.16) and the axiom schemata **K** (4.20) and **5**

$$\neg L\phi \rightarrow L\neg L\phi \tag{7.38}$$

in the following way.

1.	$L\phi \rightarrow \phi$	(premise)
2.	$L\phi \leftrightarrow \psi$	(premise)
3.	$\neg L \neg \psi$	(premise)
4.	$\neg L\phi \rightarrow \neg\psi$	2 (propositional reasoning)
5.	$L(\neg L\phi \rightarrow \neg\psi)$	4 (necessitation rule)
6.	$L(\neg L\phi \rightarrow \neg\psi) \rightarrow (L\neg L\phi \rightarrow L\neg\psi)$	(K)
7.	$L\neg L\phi \rightarrow L\neg\psi$	5, 6 (propositional reasoning)
8.	$\neg L\neg L\phi$	3, 7 (propositional reasoning)
9.	$\neg L\phi \rightarrow L\neg L\phi$	(5)
10.	$L\phi$	8, 9 (propositional reasoning)
11.	ϕ	1, 10 (propositional reasoning)
12.	ψ	2, 10 (propositional reasoning)

This indicates that Σ' has two ground **K45**-expansions: one containing ϕ and the other not containing ϕ. Hence, adding a definition $(L\phi \leftrightarrow \psi)$ of a new atom ψ changes the set of ground expansions. ■

The idea in ground expansions is to narrow the class of expansions to rule out some unwanted expansions. The discussion of the groundedness problems in Chapter 4 indicates that also ground expansions may contain circularly justified beliefs.

By Theorems 5.14 and 5.15 iterative expansions [84] and enumeration-based expansions coincide. In the propositional case, iterative expansions are **N**-expansions [86]. Thus, **N**-expansions coincide with enumeration-based expansions and, e.g., the complexity and decidability results obtained in Chapter 6 for enumeration-based expansions are valid for **N**-expansions. For example, decision problems related to N-expansions are complete problems on the second level of the polynomial time hierarchy. In fact Gottlob [40] has already established this result. Theorems 7.13 and 7.14 provide further cases where autoepistemic logic and McDermott and Doyle style nonmonotonic modal logic coincide and where results for one formalism are valid for the other. For example, by Theorem 7.14 the decision method developed for stratified autoepistemic reasoning [112] can be used for a wide range of McDermott and Doyle style logics in the stratified case.

7.4 Logic Programs and Deductive Databases

The declarative semantics of logic programs and deductive databases has been investigated intensively during the last years and several approaches have been introduced. Przymusinska and Przymusinski [114] and Bidoit [7] present surveys of the field.

We consider *extended logic programs* which are sets of rules of the form

$$l_0 \leftarrow l_1, \ldots, l_m, not\, l_{m+1}, \ldots, not\, l_n \tag{7.39}$$

where *not* is the negation as failure operator, $n \geq m \geq 0$ and each l_i is a literal, i.e. an atomic formula or a negation of an atomic formula. Extended logic programs are a generalization of *general logic programs* where each l_i is an atomic formula: classical negation is allowed in extended logic programs. Gelfond and Lifschitz [33] have proposed a semantics for extended logic programs which is a generalization of the stable model semantics of general logic programs [32]. The semantics is given in terms of answer sets which are sets of literals giving the answers to queries. An atomic ground query A is answered "yes", "no" or "unknown" depending whether the answer set in question contains A, $\neg A$ or neither. Gelfond and Lifschitz [33] show that the answer set semantics is closely related to Reiter's default logic. We exploit this connection and our results on embedding default logic into autoepistemic logic and show how answer sets correspond to L-hierarchic expansions under a simple and natural translation of extended logic programs into sets of autoepistemic sentences.

It is sufficient to study propositional logic programs as a rule with variables can be seen as a shorthand for the set of its ground instances when Herbrand interpretations are used.

An *answer set* of an extended logic P program is defined as follows [33]. The *reduct* P^S of P with respect to a set of literals S is obtained from P by deleting

1. each rule that has a formula $not\, l$ in its body with $l \in S$ and

2. all formulae of the form *not* l in the bodies of the remaining rules.

For an extended program P which does not contain *not* operators, we define the set $\alpha(P)$ as the smallest subset S of Lit (the set of ground literals in $\mathcal{L}$) such that

1. for all $l_0 \leftarrow l_1, \ldots, l_m \in P$, if $l_1, \ldots, l_m \in S$, then $l_0 \in S$ and
2. if S contains an atomic sentence and its negation, then $S = \mathrm{Lit}$.

A set of literals $S \subseteq \mathrm{Lit}$ is an answer set of an extended logic P program if and only if $S = \alpha(P^S)$.

Gelfond and Lifschitz [33] show that answer sets can be characterized using default logic by translating rules in an extended logic program in the following way.

$$\mathrm{LP}_{\mathrm{DL}}(l_0 \leftarrow l_1, \ldots, l_m, not\, l_{m+1}, \ldots, not\, l_n) = \frac{l_1 \wedge \cdots \wedge l_m : \overline{l_{m+1}}, \ldots, \overline{l_n}}{l_0} \tag{7.40}$$

where $\overline{l}$ is the literal complementary to l: $\overline{a} = \neg a$ and $\overline{\neg a} = a$. For a set of rules P, $\mathrm{LP}_{\mathrm{DL}}(P) = \{\mathrm{LP}_{\mathrm{DL}}(r) \mid r \in P\}$. The following theorem is quite straightforward to establish with the aid of two lemmata of Gelfond and Lifschitz [33].

Lemma 5: If S is an answer set of an extended logic program, then $S = \mathrm{Cn}_{\mathrm{cl}}(S) \cap \mathrm{Lit}$.

Lemma 6: For any extended logic program P and any set of sentences $E = \mathrm{Cn}_{\mathrm{cl}}(E)$, $\Gamma_{\mathrm{LP}_{\mathrm{DL}}(P)}(E) = \mathrm{Cn}_{\mathrm{cl}}(\alpha(P^{E\cap\mathrm{Lit}}))$.

Theorem 7.15 *Let P be an extended logic program. Then S is an answer set of P if and only if there exists an extension E of* $\mathrm{LP}_{\mathrm{DL}}(P)$ *such that* $S = E \cap \mathrm{Lit}$.

Proof. ($\Rightarrow$) Let S be an answer set of P and $E = \mathrm{Cn}_{\mathrm{cl}}(S)$. Then by Lemma 5 $E \cap \mathrm{Lit} = S$ and thus by Lemma 6 and the definition of an answer set,

$$\Gamma_{\mathrm{LP}_{\mathrm{DL}}(P)}(E) = \mathrm{Cn}_{\mathrm{cl}}(\alpha(P^{E\cap\mathrm{Lit}})) = \mathrm{Cn}_{\mathrm{cl}}(\alpha(P^S)) = \mathrm{Cn}_{\mathrm{cl}}(S) = E.$$

So E is an extension of $\mathrm{LP}_{\mathrm{DL}}(P)$ and $S = E \cap \mathrm{Lit}$.

($\Leftarrow$) Let E be an extension of $\mathrm{LP_{DL}}(P)$. We prove that $E \cap \mathrm{Lit}$ is an answer set of P by showing that $E \cap \mathrm{Lit} = \alpha(P^{E\cap\mathrm{Lit}})$. By Lemma 5 and 6 and as E is an extension of $\mathrm{LP_{DL}}(P)$,

$$\alpha(P^{E\cap\mathrm{Lit}}) = \mathrm{Cn_{cl}}(\alpha(P^{E\cap\mathrm{Lit}})) \cap \mathrm{Lit} = \Gamma_{\mathrm{LP_{DL}}(P)}(E) \cap \mathrm{Lit} = E \cap \mathrm{Lit}.$$

Thus $E \cap \mathrm{Lit}$ is an answer set of P. □

Default logic and the stable model semantics of general logic programs are closely related as shown by Bidoit and Froidevaux [9]. The relationship can be established also using the previous theorem because if P does not contain classical negation, i.e. P is a general logic program, then the answer sets of P are also its stable models.

Corollary 7.16 *Let P be a general logic program. Then S is a stable model of P if and only if there exists an extension E of* $\mathrm{LP_{DL}}(P)$ *such that $S = E \cap \mathrm{At}$ where* At *is the set of atomic sentences in $\mathcal{L}$.*

Proof. ($\Rightarrow$) Let S be a stable model of P. Then S is an answer set of P. Thus by Theorem 7.15 there exists an extension E of $\mathrm{LP_{DL}}(P)$ such that $S = E \cap \mathrm{Lit}$. But as S contains only atomic formulae, $S = E \cap \mathrm{At}$.

($\Leftarrow$) Let E be an extension of $\mathrm{LP_{DL}}(P)$ and $S = E \cap \mathrm{At}$. By Theorem 7.15 $S' = E \cap \mathrm{Lit}$ is a answer set of P and also a stable model of P. But as S' contains only atomic formulae, $S' = E \cap \mathrm{Lit} = E \cap \mathrm{At} = S$. □

Instead of a fixed point characterization, answer sets can be defined using L-hierarchic enumerations by translating the rules in an extended logic program in the following way.

$$\begin{aligned}&\mathrm{LP_{AE}}(l_0 \leftarrow l_1, \ldots, l_m, not\, l_{m+1}, \ldots, not\, l_n) = \\ &\qquad Ll_1 \wedge \cdots \wedge Ll_m \wedge L\neg Ll_{m+1} \wedge \cdots \wedge L\neg Ll_n \rightarrow l_0\end{aligned} \tag{7.41}$$

Here the idea is that negation as failure (not) is interpreted as disbelief ($\neg L$) and then it is realized that rules in a logic program are inference rules rather than implications. For example, we do not expect that from a logic program rule its contrapositive follows. This is captured in the translation by requiring that every premise of a logic program rule must be believed. This is the same translation as that proposed by Truszczyński [144] for embedding extended logic programs into McDermott and Doyle style nonmonotonic modal logics. Truszczyński shows that answer sets correspond to S-expansions for a wide range of modal logics S. As in the case of default logic, the correspondence holds only for consistent expansions. For example, the extended program

$$\begin{aligned}P \;=\; &\{a \leftarrow; \\ &\neg a \leftarrow \neg b; \\ &\neg b \leftarrow not\, b\}\end{aligned} \tag{7.42}$$

has no answer set but $\mathrm{LP}_{\mathrm{AE}}(P) = \{a, L\neg b \rightarrow \neg a, L\neg L\neg b \rightarrow \neg b\}$ has an inconsistent S-expansion for any modal logic S containing the necessitation rule as shown in Example 7.2. When considering L-hierarchic expansions, the anomaly in the inconsistent case does not exist. So autoepistemic logic based on L-hierarchic expansions provides a faithful characterization of the answer set semantics in terms of sequences of introspection steps.

Theorem 7.17 *Let P be a finite extended logic program. Then S is an answer set of P if and only if there exists an L-hierarchic expansion Δ of* $\mathrm{LP}_{\mathrm{AE}}(P)$ *such that $S = \Delta \cap \mathrm{Lit}$ where* Lit *is the set of literals in $\mathcal{L}$.*

Proof. ($\Rightarrow$) Let S be an answer set of P. By Theorem 7.15 there is an extension E of $\mathrm{LP}_{\mathrm{DL}}(P)$ such that $S = E \cap \mathrm{Lit}$. As $\mathrm{LP}_{\mathrm{DL}}(P)$ is finite by Theorem 7.6 there exists an L-hierarchic expansion Δ of $\mathrm{tr}_{\mathrm{T}}(\mathrm{LP}_{\mathrm{DL}}(P))$ such that $E \cap \mathrm{Lit} = \Delta \cap \mathcal{L} \cap \mathrm{Lit}$. Thus $S = \Delta \cap \mathrm{Lit}$. For each rule $l_0 \leftarrow l_1, \ldots, l_m, not\, l_{m+1}, \ldots, not\, l_n \in P$,

$$L(l_1 \wedge \cdots \wedge l_m) \wedge L\neg L\neg \overline{l_{m+1}} \wedge \cdots \wedge L\neg L\neg \overline{l_n} \rightarrow l_0 \in \mathrm{tr}_{\mathrm{T}}(\mathrm{LP}_{\mathrm{DL}}(P)) \text{ and}$$
$$Ll_1 \wedge \cdots \wedge Ll_m \wedge L\neg Ll_{m+1} \wedge \cdots \wedge L\neg Ll_n \rightarrow l_0 \in \mathrm{LP}_{\mathrm{AE}}(P).$$

It is quite straightforward to show that Δ is an L-hierarchic expansion of $\mathrm{LP}_{\mathrm{AE}}(P)$.

($\Leftarrow$) Let Δ be an L-hierarchic expansion of $\mathrm{LP}_{\mathrm{AE}}(P)$ and $S = \Delta \cap \mathrm{Lit}$. Then as above it can be shown that Δ is an L-hierarchic expansion of $\mathrm{tr}_{\mathrm{T}}(\mathrm{LP}_{\mathrm{DL}}(P))$. Thus by Theorem 7.6 there is an extension E of $\mathrm{LP}_{\mathrm{DL}}(P)$ such that $\Delta = E \cap \mathcal{L}$. Thus $S = \Delta \cap \mathrm{Lit} = E \cap \mathrm{Lit}$ is an answer set of P by Theorem 7.15. $\square$

As answer sets of a logic program without classical negation coincide with the stable models of the program, the theorem above implies that also stable models can be characterized using L-hierarchic enumerations.

Corollary 7.18 *Let P be a finite general logic program. Then S is a stable model of P if and only if there exists an L-hierarchic expansion Δ of* $\mathrm{LP}_{\mathrm{AE}}(P)$ *such that $S = \Delta \cap \mathrm{At}$ where* At *is the set of atoms in $\mathcal{L}$.*

Answers to general queries are not quite straightforward to define. When using the stable model semantics one possibility is take a stable model as a propositional interpretation giving answers to queries which are propositional sentences. In autoepistemic logic it is possible to capture truth in a model in terms of the corresponding expansion. It turns out that a sentence ϕ is true in a model I ($I \models_{\mathrm{cl}} \phi$) consisting of the atomic sentences of a consistent expansion exactly when $\mathrm{B}(\phi)$ is in the expansion where $\mathrm{B}(\phi)$ denotes the formula which is obtained by replacing every atomic subformulae χ of ϕ by $L\chi$.

Proposition 7.19 *Let Δ be a consistent stable expansion and let $I = \Delta \cap \mathrm{At}$. Then for every propositional sentence $\phi \in \mathcal{L}$, $I \models_{\mathrm{cl}} \phi$ if and only if $\mathrm{B}(\phi) \in \Delta$.*

Proof. The proof is valid for any consistent stable set Δ. Stable expansions are stable sets. Consider a sentence $\phi \in \mathcal{L}$. It can be transformed into a conjunctive normal form $\phi' = \bigwedge_{i=1}^{n} d_i$ where each d_i is a disjunction of the form

$$l_1 \vee \cdots \vee l_m \tag{7.43}$$

and each l_j is a literal. In addition, it holds that $I \models_{\mathrm{cl}} \phi$ if and only if $I \models_{\mathrm{cl}} \phi'$ for any propositional interpretation I. We show that for any stable set Δ, $\mathrm{B}(\phi) \in \Delta$ if and only if $\mathrm{B}(\phi') \in \Delta$ by proving the following result.

(B) *For any stable set Δ, if ϕ and ϕ' are two propositionally equivalent formulae, then $\mathrm{B}(\phi) \in \Delta$ if and only if $\mathrm{B}(\phi') \in \Delta$.*

This can be shown in the following way. Let $\mathrm{B}(\phi) \notin \Delta$. Then as Δ is stable, $\Delta \not\models_{\mathrm{ae}} \mathrm{B}(\phi)$ and thus there is an AE-structure $\langle I', T' \rangle$ such that $\langle I', T' \rangle \models_{\mathrm{ae}} \Delta$ but $\langle I', T' \rangle \not\models_{\mathrm{ae}} \mathrm{B}(\phi)$. But then $T' = \Delta$ which can be shown as follows. If $\chi \in T'$, then $\neg L\chi \notin \Delta$ as $\langle I', T' \rangle \models_{\mathrm{ae}} \Delta$. But as Δ is a stable set, then $\chi \in \Delta$. If $\chi \in \Delta$, then $L\chi \in \Delta$ as Δ is a stable set. Thus $\chi \in T'$ because $\langle I', T' \rangle \models_{\mathrm{ae}} \Delta$.

Let $I = \Delta \cap \mathrm{At}$. We show by structural induction that for any $\phi \in \mathcal{L}$,

$$I \models_{\mathrm{cl}} \phi \text{ if and only if } \langle I', \Delta \rangle \models_{\mathrm{ae}} \mathrm{B}(\phi). \tag{7.44}$$

If ϕ is an atomic sentence, then $\mathrm{B}(\phi) = L\phi$. Thus $I \models_{\mathrm{cl}} \phi$ if and only if $\phi \in I$ if and only if $\langle I', \Delta \rangle \models_{\mathrm{ae}} L\phi$.

Let $\phi = \neg\phi'$. Then $\mathrm{B}(\phi) = \neg\mathrm{B}(\phi')$ and $I \models_{\mathrm{cl}} \phi$ if and only if $I \not\models_{\mathrm{cl}} \phi'$. By the inductive hypothesis $I \not\models_{\mathrm{cl}} \phi'$ if and only if $\langle I', \Delta \rangle \not\models_{\mathrm{ae}} \mathrm{B}(\phi')$. Hence $I \models_{\mathrm{cl}} \neg\phi'$ if and only if $\langle I', \Delta \rangle \models_{\mathrm{ae}} \neg\mathrm{B}(\phi')$.

Let $\phi = \phi' \wedge \phi''$. Then $\mathrm{B}(\phi) = \mathrm{B}(\phi') \wedge \mathrm{B}(\phi'')$ and $I \models_{\mathrm{cl}} \phi$ if and only if $I \models_{\mathrm{cl}} \phi'$ and $I \models_{\mathrm{cl}} \phi''$. By the inductive hypothesis $I \models_{\mathrm{cl}} \phi'$ and $I \models_{\mathrm{cl}} \phi''$ if and only if $\langle I', \Delta \rangle \models_{\mathrm{ae}} \mathrm{B}(\phi')$ and $\langle I', \Delta \rangle \models_{\mathrm{ae}} \mathrm{B}(\phi'')$. Hence $I \models_{\mathrm{cl}} \phi$ if and only if $\langle I', \Delta \rangle \models_{\mathrm{ae}} \mathrm{B}(\phi') \wedge \mathrm{B}(\phi'')$. Hence (7.44) holds.

By (7.44) $I \not\models_{\mathrm{cl}} \phi$ holds as $\langle I', \Delta \rangle \not\models_{\mathrm{ae}} \mathrm{B}(\phi)$ holds. As ϕ and ϕ' are equivalent $I \not\models_{\mathrm{cl}} \phi'$. But then by (7.44) $\langle I', \Delta \rangle \not\models_{\mathrm{ae}} \mathrm{B}(\phi')$. This implies $\mathrm{B}(\phi') \notin \Delta$ as $\langle I', \Delta \rangle \models_{\mathrm{ae}} \Delta$. The other direction can be shown in a similar way. Hence (B) holds.

Then we establish the following result.

(D) *Let Δ be a stable set and let $I = \Delta \cap \mathrm{At}$. Then for any disjunction d of the form (7.43), $I \models_{\mathrm{cl}} d$ if and only if* $\mathrm{B}(d) \in \Delta$.

We use the following properties of consistent stable sets.

(1) $L\phi \vee \psi \in \Delta$ if and only if $\phi \in \Delta$ or $\psi \in \Delta$.

(2) $\neg L\phi \vee \psi \in \Delta$ if and only if $\phi \notin \Delta$ or $\psi \in \Delta$.

(3) $L\phi \in \Delta$ if and only if $\phi \in \Delta$.

(4) $\neg L\phi \in \Delta$ if and only if $\phi \notin \Delta$.

Properties (1) and (2) are properties (2.12) and (2.13) which hold for any stable set as shown in the proof of Proposition 2.19. Property (3) holds for every stable set Δ but Property (4) depends on the consistency of Δ. Let $\neg L\phi \in \Delta$. Assume that $\phi \in \Delta$. Then $L\phi \in \Delta$ and Δ is inconsistent, a contradiction. Thus $\neg L\phi \in \Delta$ implies $\phi \notin \Delta$. The other direction is immediate as Δ is stable. Hence (4) holds.

Let $d = a_1 \vee \cdots \vee a_m \vee \neg a_{m+1} \vee \cdots \vee \neg a_n$. Then $\mathrm{B}(d) = La_1 \vee \cdots \vee La_m \vee \neg La_{m+1} \vee \cdots \vee \neg La_n$. By Property (1) $\mathrm{B}(d) \in \Delta$ iff $a_1 \in \Delta$ or $\ldots$ or $a_m \in \Delta$ or $\neg La_{m+1} \vee \cdots \vee \neg La_n \in \Delta$

$$\begin{array}{lll}
& \text{iff } a_1 \in \Delta \text{ or } \ldots \text{ or } a_{n-1} \notin \Delta \text{ or } \neg La_n \in \Delta & \text{[Property (2)]} \\
& \text{iff } a_1 \in \Delta \text{ or } \ldots \text{ or } a_n \notin \Delta & \text{[Property (4)]} \\
& \text{iff } I \models_{\mathrm{cl}} a_1 \text{ or } \ldots \text{ or } I \not\models_{\mathrm{cl}} a_n & [I = \Delta \cap \mathrm{At}] \\
& \text{iff } I \models_{\mathrm{cl}} d. &
\end{array}$$

Hence (D) holds.

Now we can give the proof of Proposition 7.19. Let $\phi \in \mathcal{L}$. Then there is a formula $\phi' = \bigwedge_{i=1}^{n} d_i$ where each d_i is a disjunction of the form (7.43) such that $I \models_{\mathrm{cl}} \phi$ if and only if $I \models_{\mathrm{cl}} \phi'$. Then

$$\begin{array}{lll}
I \models_{\mathrm{cl}} \phi & \text{iff } I \models_{\mathrm{cl}} d_i \text{ for all } i = 1, \ldots, n & \\
& \text{iff } \mathrm{B}(d_i) \in \Delta \text{ for all } i = 1, \ldots, n & [\text{(D)}] \\
& \text{iff } \mathrm{B}(d_1) \wedge \cdots \wedge \mathrm{B}(d_n) \in \Delta & [\Delta \text{ is a stable set}] \\
& \text{iff } \mathrm{B}(\phi') \in \Delta & [\mathrm{B}(\phi') = \mathrm{B}(d_1) \wedge \cdots \wedge \mathrm{B}(d_n)] \\
& \text{iff } \mathrm{B}(\phi) \in \Delta. & [\text{(B)}]
\end{array}$$

□

As the expansions of translated general logic programs are consistent, Proposition 7.19 implies the following theorem which shows that not only the stable models but also answers to queries can be captured in the framework of autoepistemic logic.

Theorem 7.20 *Let P be a finite general logic program. For every sentence $\phi \in \mathcal{L}$, $S \models_{\mathrm{cl}} \phi$ for a stable model S of P if and only if there exists an L-hierarchic expansion Δ of* $\mathrm{LP}_{\mathrm{AE}}(P)$ *such that* $\mathrm{B}(\phi) \in \Delta$.

Proof. ($\Rightarrow$) Let S be a stable model of P such that $S \models_{\mathrm{cl}} \phi$. By Corollary 7.18 there exists an L-hierarchic expansion Δ of $\mathrm{LP}_{\mathrm{AE}}(P)$ such that $S = \Delta \cap \mathrm{At}$. Let $I = \{a \mid B \rightarrow a \in \mathrm{LP}_{\mathrm{AE}}(P)\}$. Now $\langle I, \Delta \rangle \models_{\mathrm{ae}} \mathrm{LP}_{\mathrm{AE}}(P) \cup L\Delta \cup \neg L\overline{\Delta}$. Thus Δ is consistent and by Proposition 7.19 $\mathrm{B}(\phi) \in \Delta$.

($\Leftarrow$) Let Δ be an L-hierarchic expansion of $\mathrm{LP}_{\mathrm{AE}}(P)$ such that $\mathrm{B}(\phi) \in \Delta$. By Corollary 7.18 there exists a stable model S of P such that $S = \Delta \cap \mathrm{At}$ and by Proposition 7.19 $S \models_{\mathrm{cl}} \phi$. □

Although stable models correspond to extensions in default logic, answers to queries cannot be captured similarly in default logic because default logic lacks explicit expressions meaning nonmembership in an extension. Here we see a point where the uniform syntax of autoepistemic logic with an explicit operator for belief proves very useful.

If a general logic program has a unique stable model, then answers to queries are defined by that model. If the program has several stable models, then the answer could be "yes" if the query is true in all the models, "no" if in none of the them and otherwise "maybe". In connection with the answer set semantics it is not immediately clear how to define semantics for general queries containing classical negation, negation as failure, and other connectives.

Example 7.4 Consider the logic program P of Example 1.2. It is given below with abbreviated predicate symbols.

$$\begin{aligned}
&p \leftarrow q, \mathit{not}\, h \\
&h \leftarrow r, \mathit{not}\, p \\
&q \leftarrow \\
&r \leftarrow
\end{aligned}$$

The corresponding set of autoepistemic sentences

$$\begin{aligned}
\mathrm{LP}_{\mathrm{AE}}(P) \;\;=\;\; &\{Lq \wedge L\neg Lh \rightarrow p, \\
&Lr \wedge L\neg Lp \rightarrow h, \\
&q, r\}
\end{aligned}$$

has two L-hierarchic expansions

$$\begin{aligned}
\Delta_1 &= \mathrm{SE}_{\mathrm{LP}_{\mathrm{AE}}(P)}(\{Lq, Lr, L\neg Lp, \neg L\neg Lh\}) \text{ and} \\
\Delta_2 &= \mathrm{SE}_{\mathrm{LP}_{\mathrm{AE}}(P)}(\{Lq, Lr, \neg L\neg Lp, L\neg Lh\}).
\end{aligned}$$

As P contains no classical negations, we consider the stable model semantics first. By Corollary 7.18 P has two stable models

$$\begin{aligned} S_1 &= \Delta_1 \cap \mathrm{At} = \{q, r, h\} \text{ and} \\ S_2 &= \Delta_2 \cap \mathrm{At} = \{q, r, p\}. \end{aligned}$$

When using the stable model semantics, queries can be answered by determining their truth value in the stable models seen as propositional interpretations. This leads to an interpretation of queries where negation is treated like negation as failure. This is captured in autoepistemic logic by interpreting queries as statements about beliefs.

- The query p is answered "maybe" as its autoepistemic interpretation $\mathrm{B}(p) = Lp$ is some but not in all expansions of the program.
- The query $\neg r$ is answered "no" as $\mathrm{B}(\neg r) = \neg Lr$ is none of the expansions.
- The query $\neg p \vee \neg h$ is answered "yes" as $\mathrm{B}(\neg p \vee \neg h) = \neg Lp \vee \neg Lh$ is in every expansion.

When using the answer set semantics, there are two kinds of negations: the classical negation and the negation as failure. Interpreting queries as statements about beliefs provides a method for distinguishing between the two types of negations. Given an answer set S, an atomic query a is answered "yes" if $a \in S$. This is captured by requiring that La belongs to the corresponding expansion. The query a is answered "no" if $\neg a \in S$. This is captured by requiring that $L\neg a$ belongs to the corresponding expansion. The query a is answered "unknown" if $a \notin S$ and $\neg a \notin S$. This is captured by requiring that $\neg La \wedge \neg L\neg a$ belongs to the corresponding expansion. This differs considerably from the stable model semantics. Considering the set S_1 as a stable model, the query p is answered "no" with respect to S_1 but according to the answer set semantics the query p is answered "unknown" with respect to S_1. Note that the corresponding L-hierarchic expansion is the same for the stable model semantics and the answer set semantics. The difference is in the way the queries are interpreted as beliefs. The notion of belief offers a promising approach to the problem of providing semantics for general queries containing classical negation, negation as failure and other connectives. The need of a clear conceptual framework becomes evident in particular in the presense of multiple answer sets. ■

It seems that an explicit belief operator is useful in queries even in the case where negation as failure is not used at all. Levesque [67] and Reiter [124] argue in favour of a generalized query language with an explicit belief operator.

They consider databases which are essentially sets of first-order sentences and claim that queries and also integrity constraints should be allowed to address besides issues related to the external world as represented by the database also aspects of what the database system believes on the basis of a given database. As an example of the difference between the two issues, consider a simple database $\{p \vee q\}$.

- The query p (i.e. is p true in the external world) is answered "unknown".
- The query Lp (i.e. is p believed on the basis of the given database) is answered "no".

Using the autoepistemic approach, it is possible to distinguish between "$p \vee q$ is believed" and "p is believed or q is believed" in queries and integrity constraints. Gelfond et al. [34] have introduced an extension of Reiter's default logic where such a distinction can be made also in the default rules. Their disjunctive default rules are Reiter style default rules whose conclusions are allowed to be expressions of the form $\phi \mid \psi$. The intuitive meaning of $\phi \mid \psi$ is that either ϕ is believed or ψ is believed. It is not clear how to express disjunctive default rules in the framework based on autoepistemic logic but these expressions can be approximated, for example in the following two different ways. In the other interpretation proposed by Gelfond et al. [34] the expression means $\{\neg L\phi \rightarrow \psi, \neg L\psi \rightarrow \phi\}$ and in the other suggested by Bidoit and Froidevaux [8] it means $\{\phi \vee \psi\} \cup \mathrm{MIN}(\{\phi, \psi\})$.

Other embeddings of logic programs into autoepistemic logic have been proposed besides (7.41). Gelfond [29] proposes an embedding where rules of a general logic program are translated by interpreting negation as failure as disbelief.

$$\begin{aligned} &\mathrm{tr_G}(a_0 \leftarrow a_1, \ldots, a_m, \mathit{not}\ a_{m+1}, \ldots, \mathit{not}\ a_n) = \\ &\qquad a_1 \wedge \cdots \wedge a_m \wedge \neg La_{m+1} \wedge \cdots \wedge \neg La_n \rightarrow a_0 \end{aligned} \tag{7.45}$$

Elkan [23] shows that a set of atoms is a stable model of a general logic program P if and only if it is the atoms in a stable expansion of $\mathrm{tr_G}(P)$. Hence Moore style expansions are adequate for capturing the stable model semantics of general logic programs. Gelfond's embedding (7.45) is not directly applicable to extended logic programs containing classical negations. This is because in the Gelfond translation, logic program rules are mapped into classical implications from which contrapositives follow. For example, consider an extended logic program $P = \{p \leftarrow q; \neg p\}$. It has a unique answer set $S = \{\neg p\}$. Its Gelfond translation $\mathrm{tr_G}(P) = \{q \rightarrow p, \neg p\}$ has a unique stable expansion which contains $\neg q$ besides $\neg p$ because $\neg q$ follows from $q \rightarrow p$

and $\neg p$. This problem can be circumvented by modifying the translation and introducing an L operator in front of every literal in the body of a rule.

$$\begin{aligned}&\mathrm{tr}_{\mathrm{K}}(a_0 \leftarrow a_1, \ldots, a_m, \mathit{not}\ a_{m+1}, \ldots, \mathit{not}\ a_n) = \\ &\qquad La_1 \wedge \cdots \wedge La_m \wedge \neg La_{m+1} \wedge \cdots \wedge \neg La_n \rightarrow a_0 \end{aligned} \tag{7.46}$$

This approach is similar to the Konolige embedding of default logic into autoepistemic logic. To achieve a match between logic program semantics and autoepistemic expansions, weak groundedness of Moore style expansions has to be compensated. For example, the logic program $\{p \leftarrow p\}$ has a unique stable model (answer set) $\{\}$ but its autoepistemic translation $\{Lp \rightarrow p\}$ has two stable expansions $\mathrm{SE}_{\{Lp\rightarrow p\}}(\{\neg Lp\})$ and $\mathrm{SE}_{\{Lp\rightarrow p\}}(\{Lp\})$. A more tightly grounded class of expansions needs to be identified so that, e.g, the latter unwanted, weakly grounded expansion is excluded. Here problems discussed in Chapter 4 related to dependence on the syntactic representation of the premises arise. Consider logic programs $P_1 = \{p \leftarrow\}$ and $P_2 = \{p \leftarrow p; p \leftarrow \mathit{not}\, p\}$. The program P_1 has a unique stable model (answer set) $\{p\}$ and P_2 has no stable model (answer set). However, the translations of P_1 and P_2, $\mathrm{tr}_{\mathrm{K}}(P_1) = \{p\}$ and $\mathrm{tr}_{\mathrm{K}}(P_2) = \{Lp \rightarrow p, \neg Lp \rightarrow p\}$ are classically equivalent. So when striving for capturing logic program semantics using the embedding (7.46), the groundedness condition would have to treat classically equivalent sets of premises such as $\{p\}$ and $\{Lp \rightarrow p, \neg Lp \rightarrow p\}$ differently.

By Theorem 7.17 and Corollary 7.18 a decision method for autoepistemic logic based on L-hierarchic expansions can serve as a decision method for the answer set semantics and the stable model semantics. On the other hand, logic programming based on the stable model semantics and the answer set semantics can be used to implement fragments of autoepistemic reasoning. Deciding whether a propositional general logic program has a stable model is an **NP**-complete problem [87]. As answer sets of a general logic program coincide with the stable models of the program, this implies immediately that the answer set existence problem is **NP**-hard. The problem is in **NP** and thus **NP**-complete because given a set of literal S, it can be tested in polynomial time whether S is an answer set of an extended logic program P by computing directly $\alpha(P^S)$ and verifying whether $S = \alpha(P^S)$ holds. Gelfond and Lifschitz [33] show how to reduce extended logic programs to general logic programs by renaming the negative literals.

The results relating autoepistemic reasoning to the stable model/answer set semantics of logic programs and the results [10, 55, 116, 118, 117] on the correspondence between autoepistemic logic and the well-founded semantics indicate that autoepistemic logic offers a promising logical foundation for logic programming. This kind of a foundation is useful in comparing alternative semantics for logic programs and in studying extensions of logic programs.

The development of more efficient implementation and optimization methods for logic programs might benefit from such a foundation. For example, Niemelä and Rintanen [112] have studied stratified general logic programs and developed a linear time algorithm for computing the unique stable model of a stratified program using a decision method for stratified autoepistemic reasoning.

7.5 Truth Maintenance Systems

Nonmonotonic truth maintenance systems were first developed by Doyle [19]. De Kleer [17] has introduced another variant, assumption-based truth maintenance systems, which are based on monotonic reasoning. Although truth maintenance systems are widely used in AI problem solving, they have been lacking a sufficiently well understood logical foundation. Autoepistemic logic has been offered as a logical basis for nonmonotonic truth maintenance systems by several researchers [120, 27, 23].

Elkan [23] presents a particularly elegant logical characterization of a nonmonotonic truth maintenance system (TMS). A TMS is a program that manages a set of nodes and a set of justifications. Each node represents some piece of information and each justification that a particular node can be believed if some other nodes are believed and some others are not believed. A TMS consists of a pair $\langle N, J\rangle$ where N is a finite set of nodes and J is a finite set of justifications, i.e. a set of expressions of the form

$$c \leftarrow a_1, \ldots, a_n, \mathit{not}\, b_1, \ldots, \mathit{not}\, b_m \tag{7.47}$$

where c, a_i, b_j are nodes. The task of a TMS is to choose a subset M of nodes N as currently believed. This subset is called a *grounded model* of the justifications. Elkan shows that justifications in a TMS can be seen as a propositional general logic program and the subset of nodes chosen by a TMS is a stable model of the corresponding logic program. He also establishes a correspondence between grounded models and Moore style stable expansions using the Gelfond translation (7.45). Elkan shows that a subset M of nodes N is a grounded model of justifications J if and only if M is the set of atoms in a stable expansion of $\mathrm{tr}_{\mathrm{G}}(J)$. By Theorem 5.13 L-hierarchic expansions and Moore style expansions coincide for premises of the form in $\mathrm{tr}_{\mathrm{G}}(J)$. Hence, the same result holds for L-hierarchic expansions. Corollary 7.18 implies another way of characterizing nonmonotonic truth maintenance systems using the translation (7.41).

Corollary 7.21 *Let $\langle N, J\rangle$ be a TMS. A subset M of N is a grounded model of J if and only if M is the set of atomic sentences in an L-hierarchic expansion of* $\mathrm{LP}_{\mathrm{AE}}(J)$.

Proof. By Theorem 3.8 [23] M is a grounded model of J if and only if M is a stable model of J. Then by Corollary 7.18 M is a grounded model of J if and only if M is the set of atomic sentences in an L-hierarchic expansion of $\mathrm{LP}_{\mathrm{AE}}(J)$. □

The Gelfond translation (7.45) could have been used in Corollary 7.21 instead of the translation (7.41). However, translations behave differently if nodes with classical negation or other connectives are considered. The situation is similar to that discussed at the end of Section 7.4. Complications related to contrapositives and weak groundedness can be avoided by choosing the translation (7.41).

By Corollary 7.21 a truth maintenance system can be understood as a rational agent performing introspection steps. A TMS can be used to compute the atomic part of a stable expansion of a restricted class autoepistemic theories. Junker and Konolige [50] show that decision problems in autoepistemic logic can be reduced to finding grounded models also in the general case. Complications arising in the reduction are discussed in the beginning of Chapter 8. The logical foundations provided by autoepistemic logic may help to develop new truth maintenance algorithms. Using the basis provided by autoepistemic logic Niemelä and Rintanen [112] have developed a linear time algorithm for computing the unique grounded model of a stratified set of justifications.

7.6 Abduction

Several forms of abduction are used in artificial intelligence. We consider forms of logic-based abduction and show how principles of abductive reasoning can be embedded into autoepistemic logic. Levesque [69] has presented a further generalized notion of knowledge-level abduction.

Poole [113] introduces a fairly general abductive framework. The framework consists of a pair $\langle F, A\rangle$ where the background theory F is a consistent set of first-order sentences and the set of allowable hypotheses A is a set of sentences. A sentence ϕ is explainable from $\langle F, A\rangle$ if there exists a subset D of A (an explanation of ϕ) such that

1. $F \cup D$ is consistent and
2. $F \cup D \models_{\mathrm{cl}} \phi$.

In applications, explanations might be required to fulfill some additional condition, e.g. only the simplest or minimal explanations are of interest. We will not discuss such further criteria in this work. Poole calls a subset D of A such that $F \cup D$ is consistent a scenario. An extension of $\langle F, A\rangle$ is the set of logical consequences of a maximal scenario of $\langle F, A\rangle$. Extensions are important because there is an explanation of a sentence ϕ if and only if ϕ belongs to an extension of $\langle F, A\rangle$. Poole [113] shows that the extensions of an abductive framework $\langle F, A\rangle$ are the extensions of a default theory where each hypothesis $h \in A$ is replaced by a default

$$\frac{: h}{h}. \tag{7.48}$$

This suggests that extensions correspond to expansions when negations of the allowable hypotheses are "minimized".

Theorem 7.22 *Let $\langle F, A\rangle$ be an abductive framework. Then $E \subseteq \mathcal{L}$ is an extension of $\langle F, A\rangle$ if and only if there is a stable expansion Δ of $F \cup \mathrm{MIN}(\neg A)$ such that $E = \Delta \cap \mathcal{L}$.*

Proof. ($\Rightarrow$) Let E be an extension and let D be the corresponding maximal subset of A such that $E = \mathrm{Cn}_{\mathrm{cl}}(F \cup D)$. Let

$$\Lambda = \{L\neg\phi \mid \neg\phi \in \neg A \cap E\} \cup \{\neg L\neg\phi \mid \neg\phi \in \neg A - E\}. \tag{7.49}$$

First we show

$$E = \mathrm{Cn}(F \cup \mathrm{MIN}(\neg A) \cup \Lambda) \cap \mathcal{L}. \tag{7.50}$$

By Proposition 5.12

$$\mathrm{Cn}(F \cup \mathrm{MIN}(\neg A) \cup \Lambda) \cap \mathcal{L} = \mathrm{Cn}_{\mathrm{cl}}(\mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), \Lambda)). \tag{7.51}$$

As $F \subseteq \mathcal{L}$, $\mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), \Lambda) = F \cup \mathrm{Hds}(\mathrm{MIN}(\neg A), \Lambda)$. Thus it remains to be shown that

$$D = \mathrm{Hds}(\mathrm{MIN}(\neg A), \Lambda) \tag{7.52}$$

holds because this implies $F \cup D = \mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), \Lambda)$ which implies (7.50) by (7.51). By definition $\mathrm{Hds}(\mathrm{MIN}(\neg A), \Lambda) = \{\psi \mid \neg L\neg\psi \rightarrow \psi \in \mathrm{MIN}(\neg A), \neg L\neg\psi \in \Lambda\} = \{\psi \mid \neg\psi \in \neg A - E\}$. If $\phi \in D$, then $\phi \in E$ and $\neg\phi \notin E$ as E is consistent. Then $\neg\phi \in \neg A - E$ and thus $\phi \in \mathrm{Hds}(\mathrm{MIN}(\neg A), \Lambda)$. If $\psi \in \mathrm{Hds}(\mathrm{MIN}(\neg A), \Lambda)$, then $\neg\psi \in \neg A - E$. Assume $\psi \notin D$. As D is a maximal subset of A such that $F \cup D$ is consistent, $F \cup D \cup \{\psi\}$ is not consistent and thus $F \cup D \models_{\mathrm{cl}} \neg\psi$. But then $\neg\psi \in E$, a contradiction. So $\psi \in D$. Hence (7.52) holds and thus (7.50) holds.

Now it is easy to show that Λ is a $F \cup \mathrm{MIN}(\neg A)$-full set. Clearly, $\Lambda \subseteq \mathrm{Sf}^{\mathrm{qL}}(F \cup \mathrm{MIN}(\neg A)) \cup \neg\mathrm{Sf}^{\mathrm{qL}}(F \cup \mathrm{MIN}(\neg A))$. For all $L\neg\psi \in \mathrm{Sf}^{\mathrm{qL}}(F \cup \mathrm{MIN}(\neg A))$,

$$\begin{aligned} L\neg\psi \in \Lambda \text{ iff } & \neg\psi \in E & [(7.49)] \\ \text{iff } & F \cup \mathrm{MIN}(\neg A) \cup \Lambda \models_{\mathrm{ae}} \neg\psi & [(7.50)] \\ \text{iff } & F \cup \mathrm{MIN}(\neg A) \cup \Lambda \models_{\mathrm{L}} \neg\psi & [\neg\psi \in \mathcal{L}] \end{aligned}$$

Similarly, $\neg L\neg\psi \in \Lambda$ if and only if $F \cup \mathrm{MIN}(\neg A) \cup \Lambda \not\models_{\mathrm{L}} \neg\psi$ holds. Thus Λ is $F \cup \mathrm{MIN}(\neg A)$-full and by Corollary 3.11 $\mathrm{SE}_{F\cup\mathrm{MIN}(\neg A)}(\Lambda)$ is a stable expansion of $F \cup \mathrm{MIN}(\neg A)$. By (7.50) $\mathrm{SE}_{F\cup\mathrm{MIN}(\neg A)}(\Lambda) \cap \mathcal{L} = \mathrm{Cn}(F \cup \mathrm{MIN}(\neg A) \cup \Lambda) \cap \mathcal{L} = E$.

($\Leftarrow$) Let Δ be a stable expansion of $F \cup \mathrm{MIN}(\neg A)$. Assume $\Delta = \mathrm{Cn}(F \cup \mathrm{MIN}(\neg A) \cup L\Delta \cup \neg L\overline{\Delta})$ is inconsistent. Thus $\Delta = \mathcal{L}_{\mathrm{ae}}$. As F is consistent, there is a CL-structure I such that $I \models_{\mathrm{cl}} F$. Let $\neg L\neg\psi \rightarrow \psi \in \mathrm{MIN}(\neg A)$. Then $\langle I, \Delta\rangle \models_{\mathrm{ae}} \neg L\neg\psi \rightarrow \psi$ as $\neg\psi \in \Delta$. Thus $\langle I, \Delta\rangle \models_{\mathrm{ae}} F \cup \mathrm{MIN}(\neg A) \cup L\Delta \cup \neg L\overline{\Delta}$ which contradicts the assumption that Δ is inconsistent. Hence Δ is consistent.

Let $D = A \cap \Delta$. Clearly, as Δ is consistent, so is $F \cup D \subseteq \Delta$. We show that

$$F \cup D = \mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta}) \tag{7.53}$$

holds. As $\mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta}) = F \cup \mathrm{Hds}(\mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta})$, it remains to be shown that $D = \mathrm{Hds}(\mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta})$. If $\psi \in D$, then $\neg\psi \notin \Delta$ as Δ is consistent. Thus $\neg L\neg\psi \in \neg L\overline{\Delta}$ and $\psi \in \mathrm{Hds}(\mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta})$ as $\neg L\neg\psi \rightarrow \psi \in \mathrm{MIN}(\neg A)$. If $\psi \in \mathrm{Hds}(\mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta})$, then $\neg L\neg\psi \rightarrow \psi \in \mathrm{MIN}(\neg A) \subseteq \Delta$ and $\neg L\neg\psi \in \neg L\overline{\Delta}$. This implies that $\psi \in \Delta \cap A = D$. Thus (7.53) holds.

Now we show that $D = A \cap \Delta$ is a maximal subset of A such that $F \cup D$ is consistent by showing that for every $\psi \in A - D$, $F \cup D \models_{\mathrm{cl}} \neg\psi$ which implies that $F \cup D \cup \{\psi\}$ is inconsistent. Let $\psi \in A - D$. Then $\psi \notin \Delta$. This implies that $\neg\psi \in \Delta$ because otherwise $\neg L\neg\psi \in \Delta$ and thus $\psi \in \Delta$ as $\neg L\neg\psi \rightarrow \psi \in \mathrm{MIN}(\neg A) \subseteq \Delta$. Thus $F \cup \mathrm{MIN}(\neg A) \cup L\Delta \cup \neg L\overline{\Delta} \models_{\mathrm{ae}} \neg\psi$. By Proposition 5.12 $\mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta}) \models_{\mathrm{cl}} \neg\psi$ holds and by (7.53) $F \cup D \models_{\mathrm{cl}} \neg\psi$ holds. Thus $D = A \cap \Delta$ is a maximal subset of A such that $F \cup D$ is consistent. Hence $E = \mathrm{Cn}_{\mathrm{cl}}(F \cup D)$ is an extension of $\langle F, A\rangle$ and

$$\begin{aligned} E &= \mathrm{Cn}_{\mathrm{cl}}(F \cup D) \\ &= \mathrm{Cn}_{\mathrm{cl}}(\mathrm{Hds}(F \cup \mathrm{MIN}(\neg A), L\Delta \cup \neg L\overline{\Delta})) && [(7.53)] \\ &= \mathrm{Cn}(F \cup \mathrm{MIN}(\neg A) \cup L\Delta \cup \neg L\overline{\Delta}) \cap \mathcal{L}. && [\text{Proposition 5.12}] \end{aligned}$$

Thus $E = \Delta \cap \mathcal{L}$ as Δ is a stable expansion of $F \cup \mathrm{MIN}(\neg A)$. □

Theorem 5.13 implies that Moore style stable expansions and L-hierarchic expansions coincide for every set of premises $F \cup \mathrm{MIN}(\neg A)$ if A is finite. Thus extensions in an abductive framework can be characterized using L-hierarchic expansions.

Corollary 7.23 *Let $\langle F, A\rangle$ be an abductive framework where A is finite. A set E is an extension of $\langle F, A\rangle$ if and only if there is an L-hierarchic expansion Δ of $F \cup \mathrm{MIN}(\neg A)$ such that $E = \Delta \cap \mathcal{L}$.*

Another interesting abductive framework is proposed by Kakas and Mancarella [52]. Their framework is a triple $\langle P, A, IC\rangle$ where P is a general logic program and A is a set atomic sentences and IC is a set of first-order sentences (integrity constraints). A sentence ϕ is explainable from the abductive framework if and only if there exists a subset D of hypotheses A that satisfies the condition that $P \cup D$ has a stable model M such that $M \models_{\mathrm{cl}} IC$ and $M \models_{\mathrm{cl}} \phi$.

As in the case of logic programs, we consider this framework at the propositional level. Thus a program rule in P which contains variables can be seen as a shorthand for the set of its ground instances and the integrity constraints in IC are propositional sentences. By Corollary 7.18 stable models correspond to L-hierarchic expansions and by Proposition 7.19 a propositional sentence ϕ is true in a stable model if and only if $\mathrm{B}(\phi)$ belongs to the corresponding expansion. Putting the pieces together explainability can be characterized in autoepistemic terms.

Theorem 7.24 *Let $\langle P, A, IC\rangle$ be an abductive framework where $P \cup A$ is finite. A propositional sentence ϕ is explainable from the framework if and only if there is a subset D of A such that $\mathrm{LP}_{\mathrm{AE}}(P \cup D)$ has an L-hierarchic expansion Δ for which $\mathrm{B}(IC) \subseteq \Delta$ and $\mathrm{B}(\phi) \in \Delta$.*

Proof. ($\Rightarrow$) Let ϕ be explainable from the framework. Then there exists a subset D of hypotheses A that satisfies the condition that $P \cup D$ has a stable model M such that $M \models_{\mathrm{cl}} IC$ and $M \models_{\mathrm{cl}} \phi$. As $P \cup A$ is finite, by Corollary 7.18 $\mathrm{LP}_{\mathrm{AE}}(P \cup D)$ has an L-hierarchic expansion Δ such that $M = \Delta \cap \mathrm{At}$. As Δ is consistent, by Proposition 7.19 $\mathrm{B}(IC) \subseteq \Delta$ and $\mathrm{B}(\phi) \in \Delta$.

($\Leftarrow$) Let D be a subset of A such that $\mathrm{LP}_{\mathrm{AE}}(P \cup D)$ has an L-hierarchic expansion Δ for which $\mathrm{B}(IC) \subseteq \Delta$ and $\mathrm{B}(\phi) \in \Delta$. Then by Corollary 7.18 $M = \Delta \cap \mathrm{At}$ is a stable model of $P \cup D$ and by Proposition 7.19 $M \models_{\mathrm{cl}} IC$ and $M \models_{\mathrm{cl}} \phi$. Hence ϕ is explainable from the framework. $\square$

The autoepistemic characterization given above is different from the approach of Kakas and Mancarella [52] where another autoepistemic interpretation of the logic program P is employed and where the integrity constraints are included in the premises generating the stable expansion corresponding to the stable model.

Theorem 7.24 illustrates the usefulness of the explicit belief operator of autoepistemic logic. The belief operator allows a simple and uniform treatment of both the integrity constraints and the explainable sentence. A similar treatment of this form of abduction is not possible, e.g., in default logic because default logic lacks means to express that a particular formula is not in an extension as a part of a query or an integrity constraint.

Chapter 8

Automating Nonmonotonic Reasoning

The results presented in Chapter 7 show that many important forms of nonmonotonic reasoning can be embedded into autoepistemic logic. Furthermore, L-hierarchic expansions play a central role in the embeddings. The results imply that decision problems related to these forms of nonmonotonic reasoning can be reduced to membership tests in L-hierarchic expansions. Hence a decision method for autoepistemic reasoning which performs a required membership test for the L-hierarchic expansions of a set of premises could serve as a unified decision method for a wide range of forms of nonmonotonic reasoning.

Several decision methods have been proposed for autoepistemic reasoning. Some of these [101, 137, 80] are straightforward applications of a particular finitary characterization of autoepistemic reasoning, some [103] use extensively a particular proof method and some [87, 50, 88] are based on the idea of mapping a decision problem in autoepistemic logic into another problem. Except [50, 80] the methods treat only original Moore style autoepistemic reasoning. In particular, none of them handles autoepistemic reasoning based on L-hierarchic expansions.

In this chapter we develop a decision method for autoepistemic logic in a general setting, i.e. for an autoepistemic logic based on a given classical logic. The novel method addresses directly tightly grounded autoepistemic reasoning. We do not want to commit to a particular proof method and we aim at a conceptually clear method where the sources of complexity and targets for optimization are clearly identifiable. Furthermore, we try to avoid exponential worst-case space requirements which seem to be inherent in the more advanced methods [103, 87, 50, 88].

We have proved that in the general setting autoepistemic reasoning is decidable if the underlying consequence relation $\models_{\text{ae}}$ is decidable. However, the decision procedures such as those given in Chapter 6 which are direct applications of a particular finitary characterization are not very practical because the procedures typically iterate over every subset of a set or every enumeration of a set. If the cardinality of the set is n, it has 2^n subsets and $n! = 1 * 2 * \cdots * n$ enumerations. Even for $n = 10$ this means 1024 (3628800) iterations each of which might be computationally costly, i.e. involve a number of calls to a theorem prover. The complexity results indicate that the worst case *time* complexity of an algorithm solving even a propositional autoepistemic reasoning problem will most probably be exponential. However, a practically oriented decision method should be able to take advantage of possible regularities in a set of premises. For example, consider a set of premises

$$\Sigma = \{p_0, Lp_0 \rightarrow p_1, Lp_1 \rightarrow p_2, \ldots, Lp_{n-1} \rightarrow p_n\}. \tag{8.1}$$

A straightforward approach would examine the 2^n subsets of $\{Lp_0, \ldots, Lp_{n-1}\}$ to find the unique Σ-full set $\Lambda = \{Lp_0, \ldots, Lp_{n-1}\}$. However, by carefully exploiting the monotonicity of the underlying consequence relation $\models_{\text{ae}}$ it is possible to find the full set without iterating over the subsets: for any Σ-full set Λ, $\Sigma \cup \Lambda \models_{\text{ae}} p_0$ holds and thus $Lp_0 \in \Lambda$ must hold which in turn implies $Lp_1 \in \Lambda$ and so on. The novel decision method aims to take full advantage of the monotonicity of $\models_{\text{ae}}$.

First we develop a decision method for autoepistemic reasoning based on L-hierarchic expansions. The decision method builds the Σ-full set[1] for every L-hierarchic expansion of Σ. The Σ-full sets can be found by examining the enumerations of the set $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. However, the enumerations are not important but the sets of beliefs generated by the enumerations. This observation cuts the combinatorial explosion considerably as typically there is a large number of enumerations inducing the same Σ-full set. The full sets are developed together as far as possible so that the same computation is not repeated by exploiting the monotonicity of the underlying consequence relation $\models_{\text{ae}}$. The novel decision method for autoepistemic reasoning needs only a theorem prover for the underlying monotonic consequence relation $\models_{\text{ae}}$ as a subroutine.

The approach differs from techniques where nonmonotonic reasoning is reduced to another related problem such as a truth maintenance problem [50, 51], a theorem proving problem [88], or a constraint satisfaction problem [4, 5]. These reductions provide valuable insights to the relationships between the

[1] In this chapter we use the term full somewhat loosely. Often a set $\Upsilon \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ satisfying the two conditions in Proposition 3.18 is called a Σ-full set instead of the actual Σ-full set $\{\phi \in \Upsilon \mid \text{q-atom}(\phi) \in \mathrm{Sf}^{\mathrm{qL}}(\Sigma)\}$.

corresponding problems and enable new techniques to be used for solving nonmonotonic reasoning problems. However, while there are interesting subclasses of nonmonotonic reasoning problems where the problem size increases only polynomially as a result of the reduction mapping (see, e.g. [51, 88, 4]), in the general case the reductions can lead to an exponential increase in the problem size. Hence the reductions turn also the worst case *space* complexity exponential. Furthermore, the reductions appear computationally quite complex even for restricted classes of nonmonotonic reasoning problems. The use of a normal form and the need to encode classical proofs can be seen as factors leading to the increase in the problem size and the high (time) complexity of the reductions. As an example we consider the approach of Junker and Konolige [50, 51] where the set of premises is required to be in a normal form where there are no nested L operators. In the worst case, if a set of premises is transformed into the normal form, its size might increase exponentially. A set of premises in the normal form is mapped to a set of justifications in a truth maintenance system. The size of the resulting set of justifications can be exponential with respect to the size of the premises in the normal form and, furthermore, the mapping seems to be computationally quite complex. This is because the set of justifications corresponding to a set of premises Σ includes all monotonic justifications of the form

$$\phi \leftarrow \omega_1, \ldots, \omega_n \tag{8.2}$$

where $L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and $\{\omega_1, \ldots, \omega_n\}$ is a minimal *proof* of ϕ from $H = \{\omega \mid \neg L\alpha \vee L\beta_1 \vee \ldots \vee L\beta_m \vee \omega \in \Sigma - \mathcal{L}\}$, i.e. a minimal subset of H such that $(\Sigma \cap \mathcal{L}) \cup \{\omega_1, \ldots, \omega_n\} \models_{\mathrm{cl}} \phi$ holds. For each $L\phi$ subformula of the premises, there can be a very large number of minimal proofs of ϕ. In addition, it seems that it is computationally quite hard to find every such minimal subset even if the logical consequence test, i.e. whether $(\Sigma \cap \mathcal{L}) \cup \{\omega_1, \ldots, \omega_n\} \models_{\mathrm{cl}} \phi$ holds, is computationally very inexpensive.

Finding a proof of ϕ from a hypothesis set H given a background theory Σ, i.e. finding a set of $H' \subseteq H$ such that $\Sigma \cup H' \models_{\mathrm{cl}} \phi$, is a computationally hard problem. It is closely related to logic-based abduction [113, 133, 21] where usually the proof H' is require to be consistent with the background theory Σ. If the background theory is a set of clauses, then deciding whether there is a consistent proof of an atom from a set of atoms is Σ_2^p-complete in the propositional case [21] which implies that the decision problem is strictly harder than propositional reasoning unless the polynomial time hierarchy collapses. Recall that propositional brave autoepistemic reasoning is also Σ_2^p-complete. Thus the search problem of finding a consistent proof of a single atom is at least as hard as brave autoepistemic reasoning. Consistent proofs are computationally hard to find even in very restricted cases. For example, if the background theory is a set of propositional Horn clauses and the hypothesis set H is a set of atomic formulae, it is **NP**-complete to decide whether there is a consistent proof of an atomic formula from H [133, 21]. Note that the corresponding reasoning problem, i.e. whether an atom is a logical consequence of a set of Horn clauses, can be solved in linear time [18].

It is even harder to find minimal proofs. It can be the case that it is easy to check whether a proof exists but a minimal proof is still hard to find. For example, the problem of deciding whether there is a proof of size K or less is an **NP**-complete problem even when the background theory is empty and the hypothesis set can contain only conjunctions of atoms and the goal is a conjunction of atoms. This can be proved by showing that the minimum cover problem, which is **NP**-complete [28], can be reduced to this minimum proof problem. In the minimum cover problem a collection C of subsets of a set S and a positive integer K are given and it is require to determine whether there is a subset of C with K or less elements such that the union of the elements equals S. It is straightforward to show that such a subset of C exists if and only if there is a proof of S_f of size K or less from C_f where S_f is a conjunction of the elements of S and C_f is the set consisting of a conjunction of elements for each set in the collection C.

In Junker and Konolige's approach an expansion corresponds to a suitable model of the justifications. For example, strongly grounded expansions correspond to grounded models but deciding whether there is a grounded model is **NP**-complete [23]. The exponential worst case space complexity and the high time complexity of the reductions even for quite restricted classes of nonmonotonic reasoning problems raise questions about the practical usability of the reduction approaches as general decision methods for nonmonotonic reasoning.

The following two notions are used in the correctness proofs of the new decision method. We say that a set of sentences B is *generated* from Σ by an enumeration $\varepsilon = \psi_1, \ldots, \psi_n$ if $B = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. An enumeration $\varepsilon = \psi_1, \ldots, \psi_n$ is said to be *fully L-hierarchic* if for every ψ_i in ε, $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\psi_i)$ implies that χ is some ψ_j in ε and $j < i$. Clearly fully L-hierarchic enumerations are L-hierarchic.

In the decision method the Σ-full sets are built using two sets B and F which contain the common part of the full sets to be considered. The set B contains beliefs generated by some fully L-hierarchic enumeration ε_m of a subset of $\{\phi \mid L\phi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$, i.e. $B = \mathbf{B}_m^{\varepsilon_m}(\Sigma)$. The set F contains $L\chi$ sentences which should be included in the full sets to be generated but are "frozen": a sentence $L\chi \in F$ is added to B only after B has been extended by either $L\chi'$ or $\neg L\chi'$ for all $L\chi'$ subformulae of χ and $\Sigma \cup B \models_{\mathrm{ae}} \chi$ holds. This ensures that B is generated from Σ by a fully L-hierarchic enumeration.

The decision method employs a function $\mathbf{is}_{\mathrm{LE}}$ presented in Figure 8.1. The function exploits the monotonicity of the underlying consequence relation $\models_{\mathrm{ae}}$ and performs positive introspection. The next proposition states the basic properties of $\mathbf{is}_{\mathrm{LE}}$.

Proposition 8.1 *Let* $\langle B', F'\rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$.

(1) $B \subseteq B'$ *and* $F \subseteq F'$ *and if* $\text{q-atom}(B) \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, *then* $\text{q-atom}(B') \cup F' \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$.

(2) If there is some fully L-hierarchic enumeration generating B *from* Σ, *then there is a fully L-hierarchic enumeration generating* B' *from* Σ.

(3) If $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ *for some* Σ*-acceptable enumeration* ε_n *of* $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$, *then* $B' \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$.

(4) For all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, *if* $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B')$ *and* $\Sigma \cup B' \models_{\mathrm{ae}} \chi$, *then* $L\chi \in \text{q-atom}(B')$.

Proof. (1) The assertion holds clearly as the function only adds $L\chi$ sentences to the sets B and F where each $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$.

(2) We show that there is a fully L-hierarchic enumeration generating B from Σ after each execution of the body of the FOR loop if there was some fully L-hierarchic enumeration generating B from Σ before the execution. This implies that there is a fully L-hierarchic enumeration generating B' from Σ. Assume that there is some fully L-hierarchic enumeration ε_m generating B from Σ. If $\Sigma \cup B \models_{\mathrm{ae}} \chi$ and $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B)$, then let ε_{m+1} be the enumeration ε_m extended by χ. So $B = \mathbf{B}_m^{\varepsilon_m}(\Sigma) = \mathbf{B}_m^{\varepsilon_{m+1}}(\Sigma)$ and $\Sigma \cup \mathbf{B}_m^{\varepsilon_{m+1}}(\Sigma) \models_{\mathrm{ae}} \chi$. Then $B \cup \{L\chi\} = \mathbf{B}_{m+1}^{\varepsilon_{m+1}}(\Sigma)$ and ε_{m+1} generates $B \cup \{L\chi\}$. It remains to be shown that ε_{m+1} is fully L-hierarchic. As ε_m is fully L-hierarchic, it is sufficient to verify that if $L\chi' \in \mathrm{Sf}^{\mathrm{L}}(\chi)$, then χ' appears in ε_{m+1} before χ, i.e. in ε_m. As $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B) = \text{q-atom}(\mathbf{B}_m^{\varepsilon_m}(\Sigma))$, $L\chi' \in \text{q-atom}(\mathbf{B}_m^{\varepsilon_m}(\Sigma))$ which implies that χ' appears in ε_m and thus ε_{m+1} is fully L-hierarchic.

(3) We show that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ holds for some Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ after each execution of the body of the FOR loop if $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ holds before the execution. This implies $B' \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. Assume that $L\chi$ is added to either B or F. Then $\Sigma \cup B \models_{\mathrm{ae}} \chi$ holds. As $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, χ is some ψ_i in ε_n. As $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. Because ε_n is Σ-acceptable, by Proposition 5.8 $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \chi$ and thus $L\chi \in \mathbf{B}_i^{\varepsilon_n}(\Sigma) \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. So $B \cup F \cup \{L\chi\} \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$.

```
FUNCTION is_LE(Σ, B, F)
INPUT: Finite sets of sentences Σ, B, and F.

  REPEAT
    Bold := B;
    FOR every Lχ ∈ Sf^L(Σ) − q-atom(Bold) DO
      IF Σ ∪ B ⊨_ae χ THEN
        IF Sf^L(χ) ⊆ q-atom(B) THEN
          B := B ∪ {Lχ}
        ELSE
          F := F ∪ {Lχ}
      ELSE
        skip
  UNTIL B = Bold;
  return ⟨B, F⟩.
```

Figure 8.1: The function $\mathbf{is}_{\mathrm{LE}}$ for positive introspection based on L-hierarchic enumerations.

(4) First we note that after each execution of the FOR loop it holds that for all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \text{q-atom}(Bold)$, if $\Sigma \cup Bold \models_{\mathrm{ae}} \chi$ and $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(Bold)$, then $L\chi \in B$. This is because at every step in the FOR loop $Bold \subseteq B$. Thus if $\Sigma \cup Bold \models_{\mathrm{ae}} \chi$, then by the monotonicity of $\models_{\mathrm{ae}}$ $\Sigma \cup B \models_{\mathrm{ae}} \chi$ holds and if $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(Bold)$, then $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B)$ holds as $\text{q-atom}(Bold) \subseteq \text{q-atom}(B)$ holds. Hence after the execution $L\chi \in B$. The set B is returned when after the execution of the FOR loop $B = Bold$ holds. Now for such a set B it holds for all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \text{q-atom}(B)$, if $\Sigma \cup B \models_{\mathrm{ae}} \chi$ and $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B)$, then $L\chi \in B$. Consider such a set $B (= B')$. Let $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B)$ and $\Sigma \cup B \models_{\mathrm{ae}} \chi$. Assume that $L\chi \notin \text{q-atom}(B)$. Then $L\chi \in B$ which implies $L\chi \in \text{q-atom}(B)$, a contradiction. Thus for all $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, if $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B)$ and $\Sigma \cup B \models_{\mathrm{ae}} \chi$, then $L\chi \in \text{q-atom}(B)$. □

Figure 8.2 presents the function $\mathbf{der}_{\mathrm{LE}}$ which is a skeleton for the different decision procedures. For that purpose it contains an unspecified function **test**. The function $\mathbf{der}_{\mathrm{LE}}$ takes as input a set of premises Σ, sets B and F which give the common part of the Σ-full sets to be considered and a sentence ϕ which is just passed as an argument to the function **test**. By changing the function **test** the various decision procedures are obtained.

The next lemma shows that the function $\mathbf{der}_{\mathrm{LE}}$ is sound, i.e. if it returns true, there is an acceptable enumeration for which **test** succeeds.

Lemma 8.2 *Let Σ be a finite set of sentences. If B and F are sets of sen-*

FUNCTION $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$
INPUT: Finite sets of sentences Σ, B, and F and a sentence ϕ.

```
⟨B, F⟩ := is_LE(Σ, B, F);
IF for some ¬Lχ ∈ B, Σ ∪ B ∪ F ⊨_ae χ THEN
    return false
ELSE IF Sf^L(Σ) ⊆ q-atom(B) ∪ F THEN
    IF F ⊆ B THEN
        return test(Σ, B, φ)
    ELSE
        return false
ELSE IF there is some Lχ ∈ Sf^L(Σ) − (q-atom(B) ∪ F)
        such that Sf^L(χ) ⊆ q-atom(B) THEN
    IF der_LE(Σ, B ∪ {¬Lχ}, F, φ) returns false THEN
        return der_LE(Σ, B, F ∪ {Lχ}, φ)
    ELSE
        return true
ELSE
    return false.
```

Figure 8.2: A skeleton for the decision procedures for autoepistemic reasoning based on L-hierarchic enumerations.

tences such that q-atom$(B) \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$*, there exists a fully L-hierarchic enumeration generating B from Σ and* $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ *returns true, then there exists an L-hierarchic Σ-acceptable enumeration ε_n of the set* $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ *such that* $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ *and* $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ *returns true.*

Proof. We prove the lemma by induction on the cardinality of the set

$$\mathrm{SL}(\Sigma, B, F) = \mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B) \cup F). \tag{8.3}$$

Let q-atom$(B) \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, let there exist a fully L-hierarchic enumeration generating B from Σ and let $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ return true.

The function starts by introspection. First we establish the following basic result which shows that if $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq$ q-atom$(B) \cup F$ after introspection, an acceptable enumeration exists.

(**B**) *If* $\mathrm{SL}(\Sigma, B', F') = \emptyset$ *where* $\langle B', F' \rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$*, then there exists an L-hierarchic Σ-acceptable enumeration ε_n of the set* $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ *such that* $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ *and* $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ *returns true.*

This can be shown in the following way. If $\mathrm{SL}(\Sigma, B', F') = \emptyset$, then $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq$ q-atom$(B') \cup F'$. As $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true, $F' \subseteq B'$ and for all $\neg L\chi \in B'$, $\Sigma \cup B' \not\models_{\mathrm{ae}} \chi$ and, furthermore, $\mathbf{test}(\Sigma, B', \phi)$ returns true.

As B is generated by a fully L-hierarchic enumeration, by Proposition 8.1 (2) there is a fully L-hierarchic enumeration ε_n generating B', i.e. $B' = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. By Proposition 8.1 (1) q-atom$(B') \cup F' \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. As $F' \subseteq B'$, q-atom$(B') \cup F' =$ q-atom(B'). Thus q-atom$(B') = \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. This implies that ε_n is an enumeration of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ because for every χ such that $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) =$ q-atom$(B') =$ q-atom$(\mathbf{B}_n^{\varepsilon_n}(\Sigma))$, χ is in ε_n. As for all $\neg L\chi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{ae}} \chi$, by Proposition 5.8 ε_n is Σ-acceptable. If $\phi \in B \cup F$, then by Proposition 8.1 (1) $\phi \in B' \cup F' = B' = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. Thus $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. Finally, $B' = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ implies that $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. So (**B**) holds.

Let $\mathrm{SL}(\Sigma, B, F) = \emptyset$. Then $\mathrm{SL}(\Sigma, B', F') = \emptyset$ where $\langle B', F' \rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$ as $B \subseteq B'$ and $F \subseteq F'$ by Proposition 8.1 (1). By (**B**) there exists an acceptable enumeration for which **test** succeeds.

Assume that for all B, F such that the cardinality of $\mathrm{SL}(\Sigma, B, F)$ is less than i if q-atom$(B) \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, there exists a fully L-hierarchic enumeration generating B from Σ and $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true, then there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true.

Let the cardinality of $\mathrm{SL}(\Sigma, B, F)$ be i. Let $\langle B', F'\rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$. If $\mathrm{SL}(\Sigma, B', F') = \emptyset$, then (**B**) implies that there exists an acceptable enumeration for which **test** succeeds.

Assume that $\mathrm{SL}(\Sigma, B', F') \neq \emptyset$. Then as $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true, there is $L\chi \in \mathrm{SL}(\Sigma, B', F')$ such that $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B')$ and either $\mathbf{der}_{\mathrm{LE}}(\Sigma, B' \cup \{\neg L\chi\}, F', \phi)$ returns true or $\mathbf{der}_{\mathrm{LE}}(\Sigma, B', F' \cup \{L\chi\}, \phi)$ returns true.

Assume that $\mathbf{der}_{\mathrm{LE}}(\Sigma, B', F' \cup \{L\chi\}, \phi)$ returns true. The cardinality of $\mathrm{SL}(\Sigma, B', F' \cup \{L\chi\})$ is less than i and $\text{q-atom}(B') \cup F' \cup \{L\chi\} \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. As B is generated by a fully L-hierarchic enumeration, by Proposition 8.1 (2) there is a fully L-hierarchic enumeration $\varepsilon_{m'}$ generating B'. Hence by the inductive hypothesis there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $B' \cup F' \cup \{L\chi\} \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. By Proposition 8.1 (1) $B \subseteq B'$ and $F \subseteq F'$ which implies $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$.

On the other hand, assume that $\mathbf{der}_{\mathrm{LE}}(\Sigma, B' \cup \{\neg L\chi\}, F', \phi)$ returns true. The cardinality of $\mathrm{SL}(\Sigma, B' \cup \{\neg L\chi\}, F')$ less than i and $\text{q-atom}(B' \cup \{\neg L\chi\}) \cup F' \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$. Let $\varepsilon_{m'+1}$ be $\varepsilon_{m'}$ extended by χ. Thus $B' = \mathbf{B}_{m'}^{\varepsilon_{m'}}(\Sigma) = \mathbf{B}_{m'}^{\varepsilon_{m'+1}}(\Sigma)$. As $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B') = \text{q-atom}(\mathbf{B}_{m'}^{\varepsilon_{m'}}(\Sigma))$ and as $\varepsilon_{m'}$ is fully L-hierarchic, $\varepsilon_{m'+1}$ is fully L-hierarchic. As $L\chi \notin \text{q-atom}(B')$, by Proposition 8.1 (4) $\Sigma \cup \mathbf{B}_{m'}^{\varepsilon_{m'+1}}(\Sigma) \not\models_{\mathrm{ae}} \chi$. Hence $B' \cup \{\neg L\chi\} = \mathbf{B}_{m'+1}^{\varepsilon_{m'+1}}(\Sigma)$. Thus $B' \cup \{\neg L\chi\}$ is generated by the fully L-hierarchic enumeration $\varepsilon_{m'+1}$. By the inductive hypothesis there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $B' \cup \{\neg L\chi\} \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. By Proposition 8.1 (1) $B \subseteq B'$ and $F \subseteq F'$ which implies $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. □

The next lemma shows that the function $\mathbf{der}_{\mathrm{LE}}$ is complete, i.e. if there is an acceptable enumeration for which the function **test** succeeds, then the function $\mathbf{der}_{\mathrm{LE}}$ returns true.

Lemma 8.3 *Let Σ be a finite set of sentences. Let ε_n be an L-hierarchic Σ-acceptable enumeration of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that* $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ *returns true. Then for every set B of sentences and F of $L\chi$ sentences such that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$,* $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ *returns true.*

Proof. We prove the lemma by induction on the cardinality of $\mathrm{SL}(\Sigma, B, F) = \mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B) \cup F)$. Let there exist an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. Let B be a set of sentences and let F be a set of $L\chi$ sentences such that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$.

We start by establishing a result showing that if $\mathrm{SL}(\Sigma, B, F) = \emptyset$ after introspection, the function $\mathbf{der}_{\mathrm{LE}}$ returns true.

(**B**) *Let* $\langle B', F'\rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$. *Then* $\mathrm{SL}(\Sigma, B', F') = \emptyset$ *implies that* $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ *returns true.*

This can be shown in the following way. By Proposition 8.1 (3) $B' \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ holds. We show that

$$B' \cup F' = \mathbf{B}_n^{\varepsilon_n}(\Sigma) \tag{8.4}$$

holds. Let $\chi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. Thus $\text{q-atom}(\chi) = L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ which implies that $\text{q-atom}(\chi) = L\psi_i \in \text{q-atom}(B') \cup F'$ as $\mathrm{SL}(\Sigma, B', F') = \emptyset$. If $\chi = \neg L\psi_i$, then $L\psi_i \notin \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and thus $L\psi_i \notin B' \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. As $L\psi_i \in \text{q-atom}(B') \cup F'$, $\neg L\psi_i \in B'$. Similarly, if $\chi = L\psi_i$, then $\neg L\psi_i \notin \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and thus $\neg L\psi_i \notin B' \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. As $L\psi_i \in \text{q-atom}(B') \cup F'$, $L\psi_i \in B' \cup F'$. Hence (8.4) holds.

Then we show that

$$F' \subseteq B' \tag{8.5}$$

holds. Assume that there is some $L\chi \in F' - B'$. By (8.4) χ is some ψ_i in ε_n. Let ψ_i be a sentence with the smallest index (in ε_n) such that $L\psi_i \in F'$ but $L\psi_i \notin B'$. By (8.4) $L\psi_i \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and thus $\neg L\psi_i \notin \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. Then by (8.4) $\neg L\psi_i \notin B'$. Thus $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \text{q-atom}(B')$. As $L\psi_i \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, $\Sigma \cup \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_i$ holds. We show next that $\mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \subseteq B'$ holds. If $L\psi_j \in \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma)$, then $j < i$ and by (8.4) $L\psi_j \in B' \cup F'$. As $j < i$, $L\psi_j \in B'$. If $\neg L\psi_j \in \mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma)$, then by (8.4) $\neg L\psi_j \in B' \cup F'$ and thus $\neg L\psi_j \in B'$. So $\mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma) \subseteq B'$ and thus $\Sigma \cup B' \models_{\mathrm{ae}} \psi_i$ holds by the monotonicity of $\models_{\mathrm{ae}}$. But as ε_n is L-hierarchic, $\mathrm{Sf}^{\mathrm{L}}(\psi_i) \subseteq \text{q-atom}(\mathbf{B}_{i-1}^{\varepsilon_n}(\Sigma)) \subseteq \text{q-atom}(B')$. Because $L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$, Proposition 8.1 (4) implies $L\psi_i \in \text{q-atom}(B')$, a contradiction. Thus (8.5) holds.

By (8.4) and (8.5) $B' = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ holds. By Proposition 5.8 as ε_n is Σ-acceptable, if $\neg L\chi \in B'$, then $\Sigma \cup B' \cup F' \not\models_{\mathrm{ae}} \chi$. Hence false is not returned and by (8.5) $\mathbf{test}(\Sigma, B', \phi)$ is executed. It returns true as $B' = \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ holds. Thus $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true and (**B**) holds.

First, let $\mathrm{SL}(\Sigma, B, F) = \emptyset$. Then by Proposition 8.1 (1) $\mathrm{SL}(\Sigma, B', F') = \emptyset$ where $\langle B', F'\rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$. Then by (**B**) $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true.

Assume that for B and F such that the cardinality of $\mathrm{SL}(\Sigma, B, F)$ is less than i, if $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true.

Let the cardinality of $\mathrm{SL}(\Sigma, B, F)$ be i and let $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. Let $\langle B', F'\rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$. If $\mathrm{SL}(\Sigma, B', F') = \emptyset$, then by (**B**) $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true.

Assume that $\mathrm{SL}(\Sigma, B', F') \neq \emptyset$. First we show that there is $L\chi \in \mathrm{SL}(\Sigma, B', F')$ such that $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B')$. Consider a sentence ψ_i which is the sentence with the smallest index (in ε_n) such that $L\psi_i \in \mathrm{SL}(\Sigma, B', F')$. We show that $\mathrm{Sf}^{\mathrm{L}}(\psi_i) \subseteq \text{q-atom}(B')$. Let $L\psi_j \in \mathrm{Sf}^{\mathrm{L}}(\psi_i)$. As ε_n is L-hierarchic, $j < i$. Therefore $L\psi_j \in \text{q-atom}(B') \cup F'$ by the minimality requirement for i. It sufficient to show that if $L\psi_j \in F'$, then $L\psi_j \in \text{q-atom}(B')$ because this implies that $L\psi_j \in \text{q-atom}(B')$ and thus $\mathrm{Sf}^{\mathrm{L}}(\psi_i) \subseteq \text{q-atom}(B')$. Let $L\psi_j \in F'$. By Proposition 8.1 (3) $L\psi_j \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\Sigma \cup \mathbf{B}_{j-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_j$ holds. We prove that

$$\mathbf{B}_{j-1}^{\varepsilon_n}(\Sigma) \subseteq B' \tag{8.6}$$

holds by showing that for all $l = 0, 1, \ldots, j-i$, $\mathbf{B}_l^{\varepsilon_n}(\Sigma) \subseteq B'$ holds. For $l = 0$ it holds trivially. Let $\mathbf{B}_l^{\varepsilon_n}(\Sigma) = \mathbf{B}_{l-1}^{\varepsilon_n}(\Sigma) \cup \{L\psi_l\}$. Then $\Sigma \cup \mathbf{B}_{l-1}^{\varepsilon_n}(\Sigma) \models_{\mathrm{ae}} \psi_l$ holds. By the inductive hypothesis on l $\Sigma \cup B' \models_{\mathrm{ae}} \psi_l$ holds. But as ε_n is L-hierarchic, $\mathrm{Sf}^{\mathrm{L}}(\psi_l) \subseteq \text{q-atom}(\mathbf{B}_{l-1}^{\varepsilon_n}(\Sigma)) \subseteq \text{q-atom}(B')$. So Proposition 8.1 (4) implies $L\psi_l \in \text{q-atom}(B')$. But $\neg L\psi_l \notin B'$ as $L\psi_l \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $B' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ by Proposition 8.1 (3). Thus $L\psi_l \in B'$.

Let $\mathbf{B}_l^{\varepsilon_n}(\Sigma) = \mathbf{B}_{l-1}^{\varepsilon_n}(\Sigma) \cup \{\neg L\psi_l\}$. Then $L\psi_l \notin B' \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. As $l < j < i$, by assumption that i is the smallest index (in ε_n) such that $L\psi_i \in \mathrm{SL}(\Sigma, B', F')$, $L\psi_l \in \text{q-atom}(B') \cup F'$. Then $L\psi_l \in \text{q-atom}(B')$ and $\neg L\psi_l \in B'$. Thus (8.6) holds.

As ε_n is L-hierarchic, by (8.6) $\mathrm{Sf}^{\mathrm{L}}(\psi_j) \subseteq \text{q-atom}(\mathbf{B}_{j-1}^{\varepsilon_n}(\Sigma)) \subseteq \text{q-atom}(B')$ and $\Sigma \cup B' \models_{\mathrm{ae}} \psi_j$ holds. By Proposition 8.1 (4) $L\psi_j \in \text{q-atom}(B')$. Hence $\mathrm{Sf}^{\mathrm{L}}(\psi_i) \subseteq \text{q-atom}(B')$. So there is $L\chi (= L\psi_i) \in \mathrm{SL}(\Sigma, B', F')$ such that $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B')$.

If $\neg L\chi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then $B' \cup \{\neg L\chi\} \cup F' \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. As the cardinality of $\mathrm{SL}(\Sigma, B' \cup \{\neg L\chi\}, F')$ is less than i, then by the inductive hypothesis $\mathbf{der}_{\mathrm{LE}}(\Sigma, B' \cup \{\neg L\chi\}, F', \phi)$ returns true which implies that $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true. If $\neg L\chi \notin \mathbf{B}_n^{\varepsilon_n}(\Sigma)$, then $L\chi \in \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $B' \cup F' \cup \{L\chi\} \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$. As the cardinality of $\mathrm{SL}(\Sigma, B', F' \cup \{L\chi\})$ is less than i, then by the inductive hypothesis $\mathbf{der}_{\mathrm{LE}}(\Sigma, B', F' \cup \{L\chi\}, \phi)$ returns true which implies that $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true. $\square$

As a direct consequence of Lemmata 8.2 and 8.3 we have the following theorem.

Theorem 8.4 *Let Σ be a finite set of sentences. Let B and F be sets of sentences such that* $\text{q-atom}(B) \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ *and let there be a fully L-hierarchic enumeration generating B from Σ. Then* $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ *returns true if and only if there exists an L-hierarchic Σ-acceptable enumeration ε_n of the set $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and* $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ *returns true.*

Proof. Let B and F be sets of sentences such that $\text{q-atom}(B) \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and let there be a fully L-hierarchic enumeration generating B from Σ. ($\Rightarrow$) If

$\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true, then by Lemma 8.2 there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. ($\Leftarrow$) Let ε_n be an L-hierarchic Σ-acceptable enumeration of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $B \cup F \subseteq \mathbf{B}_n^{\varepsilon_n}(\Sigma)$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true, then by Lemma 8.3 $\mathbf{der}_{\mathrm{LE}}(\Sigma, B, F, \phi)$ returns true. $\square$

The theorem implies that we can obtain decision procedures for the problems **SE**(LE), **DERb**(LE), and **DERb**(LE) if the underlying consequence relation $\models_{\mathrm{ae}}$ is decidable by giving an appropriate function **test** and starting with the empty set as the common part of the full sets to be considered.

Theorem 8.5 *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{ae}}$ be a finite set of sentences and $\phi \in \mathcal{L}_{\mathrm{ae}}$ a sentence.*

1. *If $\mathbf{test}(\Sigma, \Lambda, \phi)$ returns true, $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true if and only if Σ has an L-hierarchic expansion.*
2. *If $\mathbf{test}(\Sigma, \Lambda, \phi)$ returns true if and only if $\Sigma \cup \Lambda \models_{\mathrm{L}} \phi$, $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true if and only if $\Sigma \mid\!\sim_{\mathrm{b(LE)}} \phi$.*
3. *If $\mathbf{test}(\Sigma, \Lambda, \phi)$ returns true if and only if $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \phi$, $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true if and only if $\Sigma \not\mid\!\sim_{\mathrm{c(LE)}} \phi$.*

Proof. Clearly, $\text{q-atom}(\emptyset) \cup \emptyset \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and $\emptyset$ is generated from Σ by the empty enumeration which is fully L-hierarchic, of course. Then by Theorem 8.4 $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true if and only if there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true.

(1) Let $\mathbf{test}(\Sigma, \Lambda, \phi)$ return true for every input. ($\Rightarrow$) Let $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ return true. Then there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. By Theorem 5.10 there is an L-hierarchic expansion of Σ. ($\Leftarrow$) Let there exist an L-hierarchic expansion of Σ. Then by Theorem 5.11 there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. Hence $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true.

(2) Let $\mathbf{test}(\Sigma, \Lambda, \phi)$ return true if and only if $\Sigma \cup \Lambda \models_{\mathrm{L}} \phi$. ($\Rightarrow$) Let $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ return true. Then there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ and $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{L}} \phi$. By Theorem 5.10 there is an L-hierarchic expansion Δ of Σ such that

$$\Delta = \mathrm{SE}_{\Sigma}(\mathbf{B}_n^{\varepsilon_n}(\Sigma)) = \{\phi \in \mathcal{L}_{\mathrm{ae}} \mid \Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{L}} \phi\}. \tag{8.7}$$

Thus $\phi \in \Delta$ and $\Sigma \mid\!\sim_{\mathrm{b(LE)}} \phi$ holds. ($\Leftarrow$) Let $\Sigma \mid\!\sim_{\mathrm{b(LE)}} \phi$ hold. Then there exists an L-hierarchic expansion Δ of Σ containing ϕ. By Theorem 5.11 there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that Δ is given by (8.7). Thus $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \models_{\mathrm{L}} \phi$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. Hence $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true.

(3) Let $\mathbf{test}(\Sigma, \Lambda, \phi)$ return true if and only if $\Sigma \cup \Lambda \not\models_{\mathrm{L}} \phi$. ($\Rightarrow$) Let $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ return true. Then there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ and $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{L}} \phi$. By Theorem 5.10 there is an L-hierarchic expansion Δ of Σ given by (8.7). Thus $\phi \notin \Delta$ and $\Sigma \not\mid\!\sim_{\mathrm{c(LE)}} \phi$ holds. ($\Leftarrow$) Let $\Sigma \not\mid\!\sim_{\mathrm{c(LE)}} \phi$ hold. Then there exists an L-hierarchic expansion Δ of Σ not containing ϕ. By Theorem 5.11 there exists an L-hierarchic Σ-acceptable enumeration ε_n of $\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ such that Δ is given (8.7). Thus $\Sigma \cup \mathbf{B}_n^{\varepsilon_n}(\Sigma) \not\models_{\mathrm{L}} \phi$ and $\mathbf{test}(\Sigma, \mathbf{B}_n^{\varepsilon_n}(\Sigma), \phi)$ returns true. Hence $\mathbf{der}_{\mathrm{LE}}(\Sigma, \emptyset, \emptyset, \phi)$ returns true. □

The decision method reveals clearly the two sources of computational complexity in nonmonotonic reasoning:

- required monotonic reasoning and
- conflict resolution.

In the autoepistemic framework required monotonic reasoning involves deciding whether $\Sigma \models_{\mathrm{ae}} \phi$ holds for a finite set of sentences Σ and a sentence ϕ. The conflict resolution task emerges when deciding for which $L\chi$ subformula of Σ $\neg L\chi$ is included in the full set to be constructed. Such choices can be in conflict with each other: if one such $\neg L\chi$ is included in the full set to be constructed, that might make some χ' a consequence of the premises Σ and the full set and force $L\chi'$ to be added, instead of $\neg L\chi'$, to the full set. In the decision method the conflict resolution task is solved by a simple backtracking scheme. If $\neg L\chi'$ has been included in the full to be constructed but χ' becomes a consequence of the premises Σ and the full set, the method backtracks to the last choice point where some $\neg L\chi$ was added to the full set and retracts this choice. As for every $L\chi$ subformula of Σ, either $\neg L\chi$ or $L\chi$ belongs to the full set, the other possibility is that $L\chi$ belongs to the full set. The other alternative is explored by putting $L\chi$ as a "frozen" belief to be included in the full set later. The same backtracking mechanism can be exploited when searching for an expansion containing a given sentence (brave derivability) or an expansion not containing a given sentence (cautious derivability). When a full set has been found, the function **test** is invoked to determine whether the expansion induced by the full set satisfies the required condition, e.g. contains a given sentence. If this is not the case, i.e. **test** returns false, the method backtracks to the last choice point and tries the other alternative.

The decision procedures for the problems **SE**(E), **DERb**(E), and **DERb**(E) can be obtained easily by modifying the functions $\mathbf{is}_{\mathrm{LE}}$ and $\mathbf{der}_{\mathrm{LE}}$ slightly.

FUNCTION $\mathbf{is}_{\mathrm{E}}(\Sigma, B, F)$
INPUT: Finite sets of sentences Σ, B, and F.

> **REPEAT**
> $Bold := B$;
>   **FOR** every $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \text{q-atom}(Bold)$ **DO**
>   **IF** $\Sigma \cup B \models_{\mathrm{ae}} \chi$ **THEN**
> $B := B \cup \{L\chi\}$
>   **ELSE**
>   skip
> **UNTIL** $B = Bold$;
> return $\langle B, F \rangle$.

Figure 8.3: The function $\mathbf{is}_{\mathrm{E}}$ for positive introspection based on enumerations.

The only change is that there is no need to require that $\mathrm{Sf}^{\mathrm{L}}(\chi) \subseteq \text{q-atom}(B)$ for the new $(\neg)L\chi$ formulae to be added to B. The modified functions $\mathbf{is}_{\mathrm{E}}$ and $\mathbf{der}_{\mathrm{E}}$ are presented in Figures 8.3 and 8.4.

The decision procedures for the problems **SE**, **DERb**, and **DERb** can be devised by further modifying the function $\mathbf{is}_{\mathrm{E}}$. The change involves allowing the frozen sentences to be used in positive introspection because the groundedness requirements for Moore style expansions are not as strict as for enumeration-based expansions. The modified function $\mathbf{is}_{\mathrm{M}}$ is shown in Figure 8.5.

Next we consider the computational complexity of the decision method. First we note that the space complexity of the method is determined by the space complexity of the theorem prover implementing the consequence tests. If the theorem prover has a polynomial space bound, then the decision method has polynomial space complexity. This implies that the decision method yields a polynomial space decision procedure, e.g., for propositional autoepistemic reasoning because propositional consequence tests can be performed within a polynomial space bound.

The number of calls to the theorem prover for the underlying consequence relation $\models_{\mathrm{ae}}$ provides a reasonable estimate of the worst case computational complexity of the decision method. We denote by $\mathrm{tpcalls}(\Sigma, B, F, \phi)$ the number of calls to the theorem prover for $\models_{\mathrm{ae}}$ in the decision method when the input to the method is Σ, B, F, ϕ. We show that if t is an upper bound for the number of calls to the underlying theorem prover in the function **test**, then

$$\mathrm{tpcalls}(\Sigma, B, F, \phi) \leq (2^{u+1} - 1)S^2 + 2^u t \tag{8.8}$$

where S is the cardinality of $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and u is the cardinality of $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ –

```
FUNCTION der_E(Σ, B, F, φ)
INPUT: Finite set of sentences Σ, B, and F and a sentence φ.

  ⟨B, F⟩ := is_E(Σ, B, F);
  IF for some ¬Lχ ∈ B, Σ ∪ B ∪ F ⊨_ae χ THEN
     return false
  ELSE IF Sf^L(Σ) ⊆ q-atom(B) ∪ F THEN
     IF F ⊆ B THEN
        return test(Σ, B, φ)
     ELSE
        return false
  ELSE
     BEGIN
        Take some Lχ ∈ Sf^L(Σ) − (q-atom(B) ∪ F);
        IF der_E(Σ, B ∪ {¬Lχ}, F, φ) returns false THEN
           return der_E(Σ, B, F ∪ {Lχ}, φ)
        ELSE
           return true
     END.
```

Figure 8.4: A skeleton for the decision procedures for autoepistemic reasoning based on enumerations.

```
FUNCTION is_M(Σ, B, F)
INPUT: Finite sets of sentences Σ, B, and F.

  REPEAT
     Bold := B;
     FOR every Lχ ∈ Sf^L(Σ) − q-atom(B) DO
        IF Σ ∪ B ∪ F ⊨_ae χ THEN
           B := B ∪ {Lχ}
        ELSE
           skip
  UNTIL B = Bold;
  return ⟨B, F⟩.
```

Figure 8.5: The function $\mathbf{is}_{\mathrm{M}}$ for positive introspection for Moore style stable expansions.

(q-atom(B) $\cup$ F). This can be shown by induction on u. If $u = 0$, then $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq$ q-atom$(B) \cup F$ and calls to the underlying theorem prover are made only during positive introspection, conflict testing and in the function **test**. Note that during positive introspection only $L\chi$ formulae are added to B. Thus the number of $\neg L\chi$ formulae for which the conflict test is performed is bounded from above by the cardinality of B which is denoted by b. Hence

$$\mathrm{tpcalls}(\Sigma, B, F, \phi) \leq (S-b)^2 + b + t \leq S^2 + t = (2^{0+1}-1)S^2 + 2^0 t.$$

If $u > 0$, then

$$\begin{aligned}&\mathrm{tpcalls}(\Sigma, B, F, \phi)\\ &\leq S^2 + \mathrm{tpcalls}(\Sigma, B' \cup \{\neg L\chi\}, F', \phi) + \mathrm{tpcalls}(\Sigma, B', F' \cup \{L\chi\}, \phi)\end{aligned}$$

where $\langle B', F'\rangle = \mathbf{is}_{\mathrm{LE}}(\Sigma, B, F)$ and $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B') \cup F')$. By Proposition 8.1 (1) $\mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B') \cup F') \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B) \cup F)$. Thus for the cardinality u' of $\mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B' \cup \{\neg L\chi\}) \cup F')$ which is the same as that of $\mathrm{Sf}^{\mathrm{L}}(\Sigma) - (\text{q-atom}(B') \cup F' \cup \{L\chi\})$, holds that $u' < u$. Then by the inductive hypothesis

$$\begin{aligned}\mathrm{tpcalls}(\Sigma, B, F, \phi) &\leq S^2 + 2((2^{u'+1}-1)S^2 + 2^{u'}t)\\ &= (2^{u'+2}-1)S^2 + 2^{u'+1}t\\ &\leq (2^{u+1}-1)S^2 + 2^u t.\end{aligned}$$

As a consequence of (8.8), an upper bound for the number of calls to the underlying theorem prover when deciding whether a set of premises Σ has an expansion is

$$(2^{S+1}-1)S^2$$

where S is the number of $L\chi$ subformulae of Σ. For brave and cautious derivability the function **test** decides $\models_{\mathrm{L}}$ and this test can be implemented using at most $p+1$ calls to the underlying theorem prover where p is the number of $L\chi$ subformulae of ϕ. Hence an upper bound for number of calls to the underlying theorem prover when deciding whether ϕ is derivable (in the brave or cautious sense) from Σ is

$$(2^{S+1}-1)S^2 + 2^S(p+1)$$

where S is the number of $L\chi$ subformulae of Σ and p is the number of $L\chi$ subformulae of ϕ.

In the new method the theorem prover is used extensively to propagate positive introspection. It is interesting to compare the novel decision method with a straightforward approach to finding full sets where the theorem prover is not used to prune the search space. For example, full sets corresponding to L-hierarchic expansions can be found by simply iterating over every subset Λ^+ of the $L\phi$ subformulae of the premises Σ and performing a groundedness test for each Λ^+, i.e. testing whether $\Lambda^+ \cup \{\neg L\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \Lambda^+\}$ corresponds to a Σ-full set generated by an L-hierarchic enumeration. In this approach an upper bound for the number of calls to a theorem prover for $\models_{\mathrm{ae}}$ is $2^S S^2$ where S is the number of $L\phi$ subformulae of Σ. This is because there are 2^S subsets Λ^+ and an upper bound for the number of theorem prover calls in the groundedness test is S^2. Hence in the worst case the novel decision method uses only two times as many theorem prover calls as the straightforward approach. It seems beneficial to attempt to propagate positive introspection because the extra calls to the theorem prover are very advantageous when successful: every time when it is discovered that a formula χ is a consequence of the premises and the set B and a new formula $L\chi$ is added to B or F in the function $\mathbf{is}_{\mathrm{LE}}$, the upper bound for the number of theorem prover calls in the remaining computation is cut by half. In the same way, the extra calls to implement a conflict test during every recursive call of the decision method can lead to significant gains if a conflict is detected: an exponential number of calls to the theorem prover can be saved. A similar straightforward approach can be used to find full sets corresponding to enumeration-based and Moore style expansions. For enumeration-based expansions an upper bound for the number of calls to the theorem prover in the straightforward approach is also $2^S S^2$ but for Moore style expansions it is $2^S S$ because the groundedness test is just a fullness test which can be implemented with at most S calls to the theorem prover. The straightforward approach can be optimized slightly in the case there are nested L operators in the premises. As full sets consist of $L\phi$ quasi-subformulae or their negations, it is sufficient to iterate over every subset of $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ instead of every subset of $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$. If there are nested L operators in the premises, $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ can be a proper subset of $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and the optimization decreases the number of iterations. This kind of an optimization is not included in the novel decision method as we have attempted to present a conceptually simple basic decision method.

Finally note that the mappings from other forms of nonmonotonic reasoning studied in Chapter 7 into autoepistemic logic are simple local linear mappings. Hence when using the novel method to implement decision procedures for these forms of nonmonotonic reasoning, the mappings do not introduce any significant source of complexity or increase in the problem size.

The decision method described above is completely general. It does not rely on any additional assumptions about the underlying logic which determines the consequence relation $\models_{\mathrm{ae}}$. For example, no normal form transformation of the premises is required. The method needs a theorem prover for $\models_{\mathrm{ae}}$ as a subroutine. However, no extra requirements are posed on the theorem prover.

Properties of the underlying logic as well as possible regularities of the sets of premises can be utilized to optimize the computationally costly subtasks of the decision method: required monotonic reasoning ($\models_{\rm ae}$) and conflict resolution. The theorem prover for $\models_{\rm ae}$ needed as a subroutine in the decision method can exploit any special properties of the underlying logic to achieve highest possible performance because the decision method does not require the theorem prover to satisfy any additional requirements. For example, the order of solving the subgoals during theorem proving is not restricted. If the set of premises is of some special form for which a special purpose theorem prover is available, this theorem prover can be applied immediately. It should be noted that the theorem prover for $\models_{\rm ae}$ is very near that for the underlying logic CL although the sentences given as input to the theorem prover can contain L operators. This is because the $L\chi$ sentences behave like atomic sentences in the propositional calculus as far as the consequence relation $\models_{\rm ae}$ goes. For example, if the underlying logic is a first-order predicate calculus, we can use as the theorem prover for $\models_{\rm ae}$ a theorem prover for the predicate calculus which treats $L\chi$ sentence as new nilary predicates. Furthermore, it may be possible to partition the set of premises in such a way that only part of the set of premises is essential when deciding $\models_{\rm ae}$. The simple logical character of the $L\chi$ sentences suggests such optimizations. For example, Proposition 7.3 gives the flavor of the possibilities here.

The theorem proving tasks needed in the decision method have a regular pattern which could also be exploited. The decision method employs a theorem prover for $\models_{\rm ae}$ in two places: during the positive introspection and when checking whether there are conflicting choices in the formulae included in B and F, i.e. whether B and F can still be extended to a full set after positive introspection. The theorem prover is used during positive introspection in the following way: for a $L\chi$ subformula of the premises Σ, it is tested whether $\Sigma \cup B \models_{\rm ae} \chi$ holds for a monotonically growing sequence of consistent sets of quasi-literals $B \subseteq {\rm Sf}^{\rm L}(\Sigma) \cup \neg{\rm Sf}^{\rm L}(\Sigma)$ until either $\Sigma \cup B \models_{\rm ae} \chi$ holds and $L\chi$ is added to B or $\neg L\chi$ is included in B. Also the conflict tests have a similar pattern. For all $\neg L\chi$ formulae included in B, it is tested whether $\Sigma \cup B \cup F \models_{\rm ae} \chi$ holds for a monotonically growing sequence of consistent sets of quasi-literals $B \cup F \subseteq {\rm Sf}^{\rm L}(\Sigma) \cup \neg{\rm Sf}^{\rm L}(\Sigma)$ until either $\Sigma \cup B \cup F \models_{\rm ae} \chi$ holds and a conflict has been found or the full set has been completed. These regularities suggest an area of further research: given a consistent set quasi-literals $\Lambda \subseteq {\rm Sf}^{\rm L}(\Sigma) \cup \neg{\rm Sf}^{\rm L}(\Sigma)$, how to take advantage of the computation which was needed to determine that $\Sigma \cup \Lambda \not\models_{\rm ae} \chi$ holds when deciding whether $\Sigma \cup \Lambda' \models_{\rm ae} \chi$ holds for some consistent extension $\Lambda' \subseteq {\rm Sf}^{\rm L}(\Sigma) \cup \neg{\rm Sf}^{\rm L}(\Sigma)$ of Λ.

Note that it is not necessary to perform the conflict test during every recursive call of the decision method. The test can be postponed until all $L\chi$

subformulae of Σ have been exhausted ($\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq$ q-atom$(B) \cup F$) without sacrificing the correctness of the method. In the decision method presented in this chapter we have shown how to cut the backtracking as early as possible when conflicts emerge instead of building a complete candidate for a full set first and performing the conflict test only after the candidate has been completed.

In the decision method conflict resolution takes place when there is a set of $L\chi$ subformulae of the premises Σ not yet in q-atom$(B) \cup F$ and it must be decided which $\neg L\chi$ formula is added next to the full set to be constructed. At this stage it is known that for such a formula $\neg L\chi$, χ does not follow from $\Sigma \cup B$. To avoid needless backtracking, a formula $\neg L\chi$ should be selected such that $\neg L\chi$ is not refutable from $B \cup F$, i.e. χ does not follow from Σ and any consistent extension of $B \cup F$. Then B can be extended safely by $\neg L\chi$ as a conflict involving $\neg L\chi$ is not possible. Note that also these refutability tests have a regular pattern in the decision method. For a $L\chi$ subformula of the premises, refutability of $\neg L\chi$ could be tested from a monotonically growing sequence of consistent sets of quasi-literals $B \cup F \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$. Here is another area of further study. Given a set of premises Σ and a consistent set of quasi-literal $\Lambda \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$, how to decide whether $\neg L\chi$ is refutable from Λ, i.e. how to decide whether there is a consistent extension $\Lambda' \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ of Λ such that $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \chi$ and how to make use of the computation which was required to determine that $\neg L\chi$ is refutable from Λ when deciding whether $\neg L\chi$ is refutable from a consistent extension $\Lambda' \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ of Λ. Using a method to decide refutability, $\neg L\chi$ formulae could be chosen for which no backtracking is needed nor tests for conflicts. For some classes of premises, it is indeed possible to devise efficient refutability tests. Niemelä and Rintanen [112] have investigated the class of stratified sets of premises and obtained a simple iterative decision method for stratified autoepistemic logic.

Chapter 9

Conclusions

The symbolic reasoning needed in knowledge-based systems diverges from classical logical reasoning: it does not satisfy the monotonicity property respected by classical logical reasoning. Many different forms of this kind of nonmonotonic reasoning have been proposed. The aim of this work is to provide a unified theoretical basis for nonmonotonic reasoning. We argue that autoepistemic logic, originally introduced by Moore, stands out as a promising candidate. Autoepistemic logic which models the beliefs of an ideal rational introspective agent fits well the framework of knowledge-based systems where the agent corresponds to the inference engine of the system and the initial assumptions of the agent correspond to the knowledge base.

Autoepistemic logic possesses many features that make it an attractive theoretical basis for nonmonotonic reasoning. It has a uniform syntax which implies a certain degree of conceptual simplicity. For example, there is no need for a separate syntactic class corresponding to default rules typical in nonmonotonic reasoning. The explicit belief operator in autoepistemic logic permits a unified treatment of default rules, queries, and integrity constraints which is not possible, e.g., in default logic. As opposed to other approaches to formalizing nonmonotonic reasoning based on modal logic, autoepistemic logic is rather uncomplicated. It is a direct extension of the underlying classical logic. The belief operator is governed by the straightforward rules of positive and negative introspection and no modal axioms for expressing the interaction between the belief operator and the connectives of the underlying logic are needed. Moreover, recent results show that autoepistemic logic covers a range of prominent forms of nonmonotonic reasoning.

Moore formalizes correct autoepistemic conclusions from a set of premises in terms of stable expansions of the premises. Stable expansions are defined non-constructively using a fixed point equation. This leads to expansions that are quite weakly grounded in the premises, i.e. contain conclusions that are justified by circular arguments. These conclusions are often unwanted and a tighter form of autoepistemic reasoning is desired. We propose an alternative definition of autoepistemic reasoning based on enumerations of sentences where we attempt to carry the notion of a proof to autoepistemic logic. In the fixed point approach, correct reasoning is defined by presenting a condition (the fixed point equation) directly on a whole set of correct conclusions. In the enumeration-based approach, correct autoepistemic reasoning is defined as a sequence of steps starting from the premises where each step is an application of an introspection rule of autoepistemic logic. The resulting enumeration-based expansions are a proper subclass of Moore style expansions which coincides with the class of N-expansions and iterative expansions. As a solution to the problem of developing a form of autoepistemic reasoning where correct conclusions are tightly grounded in the premises, we propose a subclass of enumeration-based expansions, the L-hierarchic expansions. Here certain sequences of introspection are eliminated to guarantee that when an introspection rule is applied to a sentence, all its subbeliefs have been subjected to introspection earlier in the sequence. This ensures that the result of introspection is firmly grounded in the premises. The L-hierarchic expansions are very tightly grounded in the premises unlike Moore style expansions and some of its variants. The L-hierarchic expansions also avoid peculiarities typical of other tightly grounded classes of expansions, e.g. dependence on the syntactic representation of the premises. Autoepistemic logic based on L-hierarchic expansions provides a promising unified theoretical basis for symbolic nonmonotonic reasoning needed in knowledge-based systems as leading forms of nonmonotonic reasoning are captured within this framework. We show that Reiter's default logic, important forms of circumscription, answer set semantics of logic programs with classical negation, nonmonotonic truth maintenance systems, and forms of abduction can be seen as special cases of autoepistemic logic based on L-hierarchic expansions using very simple and intuitive embeddings.

We develop the theory of autoepistemic reasoning in a general setting where an autoepistemic logic is built on top of a given classical logic. We consider three classes of expansions: Moore style, enumeration-based, and L-hierarchic expansions. We elaborate a finitary characterization for each type of expansions. The finitary characterizations form the basis of the decidability and complexity results as well as the computational treatment of autoepistemic logic. We show that autoepistemic reasoning based on each of the three types of expansions is decidable if the monotonic consequence relation given by the underlying classical logic is decidable. In the propositional case, decision problems related to the three classes of expansions are shown to be complete problems with respect to the second level of the polynomial time hierarchy. This implies that propositional autoepistemic reasoning is strictly harder than classical propositional reasoning unless the polynomial time hierarchy collapses. Based on the finitary characterizations a decision method is developed for autoepistemic reasoning in the general setting. The method is conceptually simple and requires as a subroutine only a theorem prover for the underlying monotonic consequence relation. The method also clearly isolates the sources of complexity in nonmonotonic reasoning and permits a principled approach to developing optimization techniques. Because of the central position of autoepistemic logic based on L-hierarchic expansions the decision method offers a unified reasoning technique for a large variety of forms of nonmonotonic reasoning.

The work suggests a number of interesting areas of further research. The L-hierarchic expansions have been defined in proof-theoretic terms. A semantical characterization is an open research question. Another interesting topic is the applicability of autoepistemic logic based on some other logic than a first-order calculus. For example, autoepistemic logic based on some types of temporal logics might be useful when dealing with frame axioms needed in reasoning about actions. The decision method developed for autoepistemic reasoning is a promising basis for a unified nonmonotonic reasoning engine. The method has a solid conceptual and theoretical foundation and it covers a large range of forms of nonmonotonic reasoning as the results in Chapter 7 imply. The method structures the steps needed in nonmonotonic reasoning and indicates where optimization techniques can be applied as discussed in Chapter 8.

The complexity results on autoepistemic reasoning show that in the worst case nonmonotonic reasoning in this general form may be computationally too expensive to be applied in practical systems. The theoretical framework provided by autoepistemic logic clarifies the sources of complexity in nonmonotonic reasoning and points to more efficiently implementable forms of nonmonotonic reasoning. There are two sources of complexity in nonmonotonic reasoning: required monotonic reasoning and conflict resolution. In the autoepistemic framework, the required monotonic reasoning is given by the underlying classical logic and conflict resolution means the determination of a proper order of applying introspection rules. The computational complexity of required monotonic reasoning can be reduced easily, e.g. by restricting the form of sentences in the set of premises. However, this does not imply fast nonmonotonic reasoning methods because the conflict resolution task might require too much computational resources. The autoepistemic basis of nonmonotonic reasoning developed in this work suggests two interesting approaches to subclasses of nonmonotonic reasoning for which the conflict resolution task can be efficiently solved.

The formalization of autoepistemic reasoning as a sequence of introspection steps suggests a notion of priorities where priorities restrict orders of introspection. The advantage of this notion of priorities is that it has a rather declarative interpretation. The priorities are not tied, e.g., to the method and strategy used by the theorem prover for the underlying monotonic reasoning. The computational advantage comes from the fact that the search base for conflict resolution is pruned. This notion of priorities seems to offer an interesting solution to the problem of choosing the "preferred" expansion when reasoning in the brave mode in the presence of multiple expansions. Priorities have already been studied in the stratified case [112]. It turns out that the priorities, i.e. preferred orders of introspection, are implicit in a stratified set of premises and a suitable order of introspection can be computed from the set of premises. The order of introspection provides a solution to the conflict resolution task. Hence, for stratified premises an iterative decision method can be developed the complexity of which is dominated by the complexity of required monotonic reasoning [112].

The other approach is directed to reasoning in the cautious mode where we are interested in the intersection of all expansions. In conflict resolution we have to determine when to apply the negative introspection rule and the difficulties arise since we might apply the rule too early so that we end up in a conflict because of some later introspection. If we weaken the negative introspection rule so that a conflict is not possible at a later stage, we are staying within the intersection of the expansions. The resulting set of conclusions is an approximation of the intersection. This kind of an approach seems to lead to an iterative reasoning method. The approach is related to the well-founded semantics of logic programs [149]. There appear to be many alternative ways of weakening the rule and some proposals have already been made [10, 54, 55, 117].

Bibliography

[1] K.R. Apt, H.A. Blair, and A. Walker. Towards a theory of declarative knowledge. In J. Minker, editor, *Foundations of Deductive Databases and Logic Programming*, pages 89–148. Morgan Kaufmann Publishers, Los Altos, 1988.

[2] A.B. Baker. Nonmonotonic reasoning in the framework of situation calculus. *Artificial Intelligence*, 49:5–23, 1991.

[3] J.L. Balcázar, I. Díaz, and J. Gabarró. *Structural Complexity I.* Springer-Verlag, Berlin, 1988.

[4] 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. The MIT Press, July 1991.

[5] R. Ben-Eliyahu and R. Dechter. Propositional semantics for default logic. In *Working Notes of the 4th International Workshop on on Nonmonotonic Reasoning*, pages 13–27, Plymouth, Vermont, USA, May 1992.

[6] P. Besnard. *An Introduction to Default Logic.* Springer-Verlag, Berlin, 1989.

[7] N. Bidoit. Negation in rule-based database languages: a survey. *Theoretical Computer Science*, 78:3–83, 1991.

[8] N. Bidoit and C. Froidevaux. General logical databases and programs: Default logic semantics and stratification. *Information and Computation*, 91:15–54, 1991.

[9] N. Bidoit and C. Froidevaux. Negation by default and unstratified logic programs. *Theoretical Computer Science*, 78:85–112, 1991.

[10] P.A. Bonatti. Autoepistemic logics as a unifying framework for the semantics of logic programs. In *Proceedings of the 1992 Joint International Conference and Symposium on Logic Programming*, Washington, D.C., USA, November 1992.

[11] G. Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense.* Cambridge University Press, Cambridge, 1991.

[12] M. Cadoli and M. Lenzerini. The complexity of closed world reasoning and circumscription. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 550–555, Boston, MA, USA, July 1990. The MIT Press.

[13] A. Chandra and D. Harel. Horn clause queries and generalizations. *Journal of Logic Programming*, 2:1–15, 1985.

[14] B.F. Chellas. *Modal Logic: an Introduction.* Cambridge University Press, Cambridge, 1980.

[15] K.L. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 293–322. Plenum Press, New York, 1978.

[16] M. Davis. The mathematics of non-monotonic reasoning. *Artificial Intelligence*, 13:73–80, 1980.

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

[18] W.F. Dowling and J.H. Gallier. Linear-time algorithms for testing the satisfiability of propositional Horn formulae. *Journal of Logic Programming*, 3:267–284, 1984.

[19] J. Doyle. A truth maintenance system. *Artificial Intelligence*, 12:231–272, 1979.

[20] T. Eiter and G. Gottlob. Propositional circumscription and extended closed world reasoning are Π_2^p-complete. Technical Report CD-TR 91/20, Christian Doppler Labor für Expertensysteme, Institut für Informationssysteme, Technische Universität Wien, Vienna, Austria, May 1991.

[21] T. Eiter and G. Gottlob. The complexity of logic-based abduction. Technical Report CD-TR 92/35, Christian Doppler Labor für Expertensysteme, Institut für Informationssysteme, Technische Universität Wien, Vienna, Austria, July 1992.

[22] T. Eiter and G. Gottlob. Reasoning with parsimonious and moderately grounded expansions. *Fundamenta Informaticae*, 17:31–53, 1992.

[23] C. Elkan. A rational reconstruction of nonmonotonic truth maintenance systems. *Artificial Intelligence*, 43:219–234, 1990.

[24] D.W. Etherington. Formalizing nonmonotonic reasoning systems. *Artificial Intelligence*, 31:41–85, 1987.

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

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

[27] Y. Fujiwara and Honiden S. Relating TMS to autoepistemic logic. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1199–1205, Detroit, USA, August 1989. Morgan Kaufmann Publishers.

[28] M.R. Garey and D.S. Johnson. *Computers and Intractability*. W.H. Freeman and Company, San Francisco, 1979.

[29] M. Gelfond. On stratified autoepistemic theories. In *Proceedings of the 6th National Conference on Artificial Intelligence*, pages 207–211, Seattle, USA, July 1987. American Association for Artificial Intelligence.

[30] M. Gelfond. Autoepistemic logic and formalization of commonsense reasoning, preliminary report. In *Proceedings of the 2nd Workshop on Non-Monotonic Reasoning*, pages 176–186, Grassau, FRG, June 1988. Springer-Verlag.

[31] M. Gelfond. Strong introspection. In *Proceedings of the 9th National Conference on Artificial Intelligence*, pages 386–391. The MIT Press, July 1991.

[32] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proceedings of the 5th International Conference on Logic Programming*, pages 1070–1080, Seattle, USA, August 1988. The MIT Press.

[33] M. Gelfond and V. Lifschitz. Logic programs with classical negation. In *Proceedings of the 7th International Conference on Logic Programming*, pages 579–597, Jerusalem, Israel, June 1990. The MIT Press.

[34] M. Gelfond, V. Lifschitz, H. Przymusinska, and M. Truszczyński. Disjunctive defaults. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 230–237, Cambridge, MA, USA, April 1991. Morgan Kaufmann Publishers.

[35] M. Gelfond and H. Przymusinska. Formalization of inheritance reasoning in autoepistemic logic. *Fundamenta Informaticae*, 13:430–443, 1990.

[36] M. Gelfond, H. Przymusinska, and T. Przymusinski. On the relationship between circumscription and negation as failure. *Artificial Intelligence*, 38:75–94, 1989.

[37] M.L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann Publishers, Los Altos, 1987.

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

[39] G. Gottlob. Complexity results for nonmonotonic logics. Technical Report CD-TR 91/24, Christian Doppler Laboratory for Expert Systems, Institut für Informationssysteme, Technische Universität Wien, Vienna, Austria, August 1991.

[40] G. Gottlob. Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425, June 1992.

[41] G. Gottlob. The power of beliefs or translating default logic into standard autoepistemic logic. Presented at the European Workshop on Theoretical Foundations of Knowledge Representation and Reasoning held in conjunction with ECAI'92, August 3–4, Vienna, Austria, 1992.

[42] J.Y. Halpern and Y. Moses. Towards a theory of knowledge and ignorance: Preliminary report. In *Proceedings of Non-Monotonic Reasoning Workshop*, pages 125–143, New Paltz, New York, October 1984. American Association for Artificial Intelligence.

[43] S. Hanks and D. McDermott. Nonmonotonic logic and temporal projection. *Artificial Intelligence*, 33:379–412, 1987.

[44] N. Helft, K. Inoue, and D. Poole. Query answering in circumscription. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 426–431, Sydney, Australia, August 1991. Morgan Kaufmann Publishers.

[45] C. Hewitt. Planner: a language for manipulating models and proving theorems in a robot. A.I. Memo 168, MIT, Artificial Intelligence Project MAC, Cambridge, MA, 1968.

[46] G.E. Hughes and M.J. Cresswell. *A Companion to Modal Logic*. Methuen and Co., London, 1984.

[47] T. Imielinski. Results on translating defaults to circumscription. *Artificial Intelligence*, 32:131–146, 1987.

[48] D.S. Johnson. A catalog of complexity classes. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A. Algorithms and Complexity, pages 67–161. Elsevier, 1990.

[49] U. Junker. A correct non-monotonic ATMS. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1049–1054, Detroit, Michigan, USA, August 1989. Morgan Kaufmann Publishers.

[50] 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*, pages 278–283, Boston, MA, USA, July 1990. The MIT Press.

[51] U. Junker and K. Konolige. Computing the extensions of autoepistemic and default logics with a truth maintenance system. TASSO-Report 16, Gesellschaft für Mathematik und Datenverarbeitung, St. Augustin, Germany, December 1990.

[52] A.C. Kakas and P. Mancarella. Generalized stable models: a semantics for abduction. In *Proceedings of the 9th European Conference on Artificial Intelligence*, pages 385–391, Stockholm, Sweden, August 1990. Pitman Publishing.

[53] A.C. Kakas and P. Mancarella. Negation as stable hypothesis. In *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 275–288, Washington, D.C., USA, July 1991. The MIT Press.

[54] J. Kalinski. Autoepistemic expansions with incomplete belief introspection. In *Proceedings of the 1st International Workshop on Nonmonotonic and Inductive Logic*, pages 223–232, Karlsruhe, Germany, December 1990. Springer-Verlag. Lecture Notes in Artificial Intelligence 543.

[55] J. Kalinski. Weak autoepistemic reasoning. Presented at the European Workshop on Theoretical Foundations of Knowledge Representation and Reasoning held in conjunction with ECAI'92, August 3–4, Vienna, Austria, 1992.

[56] M. Kaminski. Embedding a default system into nonmonotonic logic. *Fundamenta Informaticae*, 14:345–354, 1991.

[57] H. Kautz. The logic of persistence. In *Proceedings of the 5th National Conference on Artificial Intelligence*, pages 401–405, Philadelphia, USA, August 1986. Morgan Kaufmann Publishers.

[58] H. Kautz and B. Selman. Hard problems for simple default logics. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 189–197, Toronto, Canada, May 1989. Morgan Kaufmann Publishers.

[59] H. Kautz and B. Selman. Hard problems for simple default logics. *Artificial Intelligence*, 49:243–279, 1991.

[60] K. Konolige. Hierarchic autoepistemic theories for nonmonotonic reasoning: Preliminary report. In *Proceedings of the 2nd Workshop on Non-Monotonic Reasoning*, pages 42–59, Grassau, FRG, June 1988. Springer-Verlag.

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

[62] K. Konolige. Errata: On the relation between default and autoepistemic logic [Artificial Intelligence 35 (1988) 343-382]. *Artificial Intelligence*, 41:115, 1989.

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

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

[65] K. Konolige. Ideal introspective belief. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 635–641, San Jose, CA, USA, July 1992. The MIT Press.

[66] H.J. Levesque. The interaction with incomplete knowledge bases: a formal treatment. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 240–245, Vancouver, Canada, August 1981. Morgan Kaufmann Publishers.

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

[68] H.J. Levesque. All I know: Abridged report. In *Proceedings of the 6th National Conference on Artificial Intelligence*, pages 426–431, Seattle, USA, July 1987. Morgan Kaufmann Publishers.

[69] H.J. Levesque. A knowledge-level account of abduction. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1061–1067, Detroit, USA, August 1989. Morgan Kaufmann Publishers.

[70] H.J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42:263–309, 1990.

[71] V. Lifschitz. Computing circumscription. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 121–127, Los Angeles, California, USA, August 1985. Morgan Kaufmann Publishers.

[72] V. Lifschitz. Formal theories of action. In M.L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 410–432. Morgan Kaufmann Publishers, Los Altos, 1987.

[73] V. Lifschitz. Pointwise circumscription. In M.L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 179–193. Morgan Kaufmann Publishers, Los Altos, 1987.

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

[75] V. Lifschitz. Nonmonotonic databases and epistemic queries. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 381–386, Sydney, Australia, August 1991. Morgan Kaufmann Publishers.

[76] F. Lin and Y. Shoham. Argument systems: a uniform basis for nonmonotonic reasoning. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 245–255, Toronto, Canada, May 1989. Morgan Kaufmann Publishers.

[77] F. Lin and Y. Shoham. Epistemic semantics for fixed-point nonmonotonic logics. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 111–120, Pacific Grove, USA, March 1990. Morgan Kaufmann Publishers.

[78] F. Lin and Y. Shoham. A logic of knowledge and justified assumptions. *Artificial Intelligence*, 57:271–289, 1992.

[79] W. Marek. Stable theories in autoepistemic logic. *Fundamenta Informaticae*, 12:243–254, 1989.

[80] W. Marek, G.F. Shvarts, and M. Truszczyński. Modal nonmonotonic logics: Ranges, characterization, computation. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 395–404, Cambridge, MA, USA, April 1991. Morgan Kaufmann Publishers.

[81] W. Marek and V.S. Subrahmanian. The relationship between logic program semantics and non-monotonic reasoning. In *Proceedings of the 6th International Conference on Logic Programming*, pages 600–617, Lisbon, Portugal, June 1989. The MIT Press.

[82] W. Marek and V.S. Subrahmanian. The relationship between stable, supported, default and autoepistemic semantics for general logic programs. *Theoretical Computer Science*, 103:365–386, 1992.

[83] W. Marek and M. Truszczyński. Autoepistemic logic. Technical Report 115-88, Department of Computer Science, University of Kentucky, Lexington, Kentucky, USA, 1988.

[84] W. Marek and M. Truszczyński. Relating autoepistemic and default logics. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 276–288, Toronto, Canada, May 1989.

[85] W. Marek and M. Truszczyński. Modal logic for default reasoning. In *Preprints of the 3rd International Workshop on Nonmonotonic Reasoning*, pages 247–258, South Lake Tahoe, California, May 1990.

[86] W. Marek and M. Truszczyński. Modal logic for default reasoning. *Annals of Mathematics and Artificial Intelligence*, 1:275–302, 1990.

[87] W. Marek and M. Truszczyński. Autoepistemic logic. *Journal of the ACM*, 38:588–619, 1991.

[88] 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, Washington, D.C., USA, July 1991. The MIT Press.

[89] W. Marek and M. Truszczyński. More on modal aspects of default logic. *Fundamenta Informaticae*, 17:99–116, 1992.

[90] J. McCarthy. Programs with common sense. In *Proceedings of the Symposium on Mechanisation of Thought Processes*, pages 77–84, London, November 1958. Has appeared also in *Semantic Information Processing*, M. Minsky, Ed., The MIT Press, Cambridge, 1968, pages 403–418 and in *Readings in Knowledge Representation*, R. Brachman and H. Levesque, Eds., Morgan Kaufmann Publishers, Los Altos, 1985, pages 300–307.

[91] J. McCarthy. Epistemological problems of artificial intelligence. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pages 1038–1044, Cambridge, MA, USA, August 1977.

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

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

[94] J. McCarthy and P.J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, 1969.

[95] D. McDermott. Non-monotonic logic II. *Journal of the ACM*, 29:33–57, 1982.

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

[97] M. Minsky. A framework for representing knowledge. In J. Haugeland, editor, *Mind Design*, pages 95–128. The MIT Press, Cambridge, 1981.

[98] R.C. Moore. Semantical considerations on nonmonotonic logic. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 272–279, Karlsruhe, FRG, August 1983. Morgan Kaufmann Publishers.

[99] R.C. Moore. Possible-world semantics for autoepistemic logic. In *Proceedings of the Workshop on Non-Monotonic Reasoning*, pages 344–354, New Paltz, USA, October 1984. American Association for Artificial Intelligence.

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

[101] R.C. Moore. Autoepistemic logic. In P. Smets, E.H. Mamdani, and D. Dubois, editors, *Non-Standard Logics for Automated Reasoning*, pages 105–136. Academic Press, London, 1988.

[102] P.H. Morris. Autoepistemic stable closure and contradiction resolution. In *Proceedings of the 2nd Workshop on Non-Monotonic Reasoning*, pages 60–73, Grassau, FRG, June 1988. Springer-Verlag.

[103] I. Niemelä. Decision procedure for autoepistemic logic. In *Proceedings of the 9th International Conference on Automated Deduction*, pages 675–684, Argonne, USA, May 1988. Springer-Verlag.

[104] I. Niemelä. On the complexity of the decision problem in propositional nonmonotonic logic. In *Proceedings of the 2nd Workshop on Computer Science Logic*, pages 226–239, Duisburg, FRG, October 1988. Springer-Verlag.

[105] I. Niemelä. Decision problem in autoepistemic logic. In *Proceedings of the Workshop on Alternatives of Logic Programming*, pages 174–188, Kuopio, Finland, May 1989. Kuopion yliopiston painatuskeskus.

[106] I. Niemelä. A new characterization of autoepistemic reasoning. In *Proceedings of the Finnish Artificial Intelligence Symposium*, pages 64–73, Oulu, Finland, June 1990. Finnish Artificial Intelligence Society.

[107] 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.

[108] I. Niemelä. Constructive tightly grounded autoepistemic reasoning. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 399–404, Sydney, Australia, August 1991. Morgan Kaufmann Publishers.

[109] I. Niemelä. Autoepistemic logic as a basis for automating nonmonotonic reasoning. Lecture Notes of the Summer School on Partial Semantics and Non-Monotonic Reasoning (Linköping, Sweden, May 25–29, 1992), 1992.

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

[111] I. Niemelä. A unifying framework for nonmonotonic reasoning. In *Proceedings of the 10th European Conference on Artificial Intelligence*, pages 334–338, Vienna, Austria, August 1992. John Wiley.

[112] I. Niemelä and J. Rintanen. On the impact of stratification on the complexity of nonmonotonic reasoning. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, pages 627–638, Cambridge, MA, USA, October 1992. Morgan Kaufmann Publishers.

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

[114] H. Przymusinska and T.C. Przymusinski. Semantic issues in deductive databases and logic programs. In R. Banerji, editor, *Formal Techniques in Artificial Intelligence*, pages 321–367. North-Holland, Amsterdam, 1990.

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

[116] T.C. Przymusinski. Three-valued formalizations of non-monotonic reasoning and logic programming. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 341–348, Toronto, Canada, May 1989. Morgan Kaufmann Publishers.

[117] T.C. Przymusinski. Autoepistemic logics of closed beliefs and logic programming. In *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 3–20, Washington, D.C., USA, July 1991. The MIT Press.

[118] T.C. Przymusinski. Three-valued nonmonotonic formalisms and semantics of logic programming. *Artificial Intelligence*, 49:309–343, 1991.

[119] P.K. Rathmann and M. Winslett. Circumscribing equality. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 468–473, Detroit, USA, August 1989. Morgan Kaufmann Publishers.

[120] M. Reinfrank, O. Dressler, and G. Brewka. On the relation between truth maintenance and autoepistemic logic. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1206–1212, Detroit, Michigan, August 1989. Morgan Kaufmann Publishers.

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

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

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

[124] R. Reiter. On asking what a database knows. In *Proceedings of the Symposium on Computational Logic*, pages 95–113, Brussels, Belgium, November 1990. Springer-Verlag.

[125] E. Sandewall. An approach to the frame problem and its implementation. In *Machine Intelligence 7*, pages 195–204. Wiley, New York, 1972.

[126] 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.

[127] G. Schwarz. Autoepistemic logic of knowledge. In *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 260–274, Washington, D.C., USA, July 1991. The MIT Press.

[128] G. Schwarz. Bounding introspection in nonmonotonic logic. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, pages 581–590, Cambridge, MA, USA, October 1992. Morgan Kaufmann Publishers.

[129] G. Schwarz. Minimal model semantics for nonmonotonic modal logics. In *Proceedings of the Seventh Annual IEEE Symposium on Logic in Computer Science*, pages 34–43, Santa Cruz, California, USA, June 1992. IEEE Computer Society Press.

[130] G. Schwarz. Reflexive autoepistemic logic. *Fundamenta Informaticae*, 17:157–173, 1992.

[131] G. Schwarz and M. Truszczyński. Modal logic S4F and the minimal knowledge paradigm. In *Proceedings of the 4th Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 184–198, Pacific Grove, California, USA, March 1992. Morgan Kaufmann Publishers.

[132] C. Schwind. A tableaux-based theorem prover for a decidable subset of default logic. In *Proceedings of the 10th International Conference on Automated Deduction*, pages 528–542, Kaiserslautern, FRG, July 1990. Springer-Verlag.

[133] B. Selman and H.J. Levesque. Abductive and default reasoning: A computational core. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 343–348, Boston, MA, USA, July 1990. The MIT Press.

[134] Y. Shoham. Chronological ignorance: Experiments in nonmonotonic temporal reasoning. *Artificial Intelligence*, 36:279–331, 1988.

[135] G. Shvarts. Autoepistemic modal logics. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 97–109, Pacific Grove, USA, March 1990. Morgan Kaufmann Publishers.

[136] R. Stalnaker. A note on non-monotonic modal logic. Unpublished manuscript, 1980.

[137] R.F. Stärk. On the existence of fixpoints in Moore's autoepistemic logic and the non-monotonic logic of McDermott and Doyle. In *Proceedings of the 4th Workshop on Computer Science Logic*, pages 354–365, Heidelberg, Germany, October 1990. Springer-Verlag.

[138] J. Stillman. It's not my default: the complexity of membership problems in restricted propositional default logics. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 571–578, Boston, Massachusetts, USA, July 1990. The MIT Press.

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

[140] M. Tiomkin and M. Kaminski. Nonmonotonic default modal logics. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 73–83, Pacific Grove, California, March 1990. Morgan Kaufmann Publishers.

[141] M. Tiomkin and M. Kaminski. Nonmonotonic default modal logics. *Journal of the ACM*, 38(4):963–984, October 1991.

[142] D.S. Touretzky. *The Mathematics of Inheritance Systems.* Morgan Kaufmann Publishers, Los Altos, 1986.

[143] D.S. Touretzky, J.F. Horty, and R.H. Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 476–482, Milan, Italy, August 1987. Morgan Kaufmann Publishers.

[144] M. Truszczyński. Embedding default logic into modal nonmonotonic logics. In *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 151–165, Washington, D.C., USA, July 1991. The MIT Press.

[145] M. Truszczyński. Modal interpretations of default logic. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 393–398, Sydney, Australia, August 1991. Morgan Kaufmann Publishers.

[146] M. Truszczyński. Modal nonmonotonic logic with restricted application of the negation as failure to prove rule. *Fundamenta Informaticae*, 14:355–366, 1991.

[147] A. Van Gelder. Negation as failure using tight derivations for general logic programs. In J. Minker, editor, *Foundations of Deductive Databases and Logic Programming*, pages 149–176. Morgan Kaufmann Publishers, Los Altos, 1988.

[148] A. Van Gelder, K.A. Ross, and J.S. Schlipf. Unfounded sets and the well-founded semantics for general logic programs, extended abstract.

In *Proceedings of the 7th Symposium on Principles of Database Systems*, pages 221–230, Austin, Texas, March 1988. ACM Press.

[149] A. Van Gelder, K.A. Ross, and J.S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, July 1991.

[150] A. Zhang and W. Marek. On the classification and existence of structures in default logic. *Fundamenta Informaticae*, 13:485–499, 1990.