HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A:** Research Reports
No. **31**; December 1994

ISSN 0783–5396
ISBN 951-22-2396-1

INVESTIGATIONS ON CAUTIOUS AUTOEPISTEMIC REASONING

Tomi Janhunen

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
FIN–02150 ESPOO, FINLAND

HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A**: Research Reports
No. **31**; December 1994

ISSN 0783–5396
ISBN 951-22-2396-1

Investigations on Cautious Autoepistemic Reasoning

TOMI JANHUNEN

Abstract: In this work, a formal model of knowledge representation systems, cautious autoepistemic logic, is investigated. This logic is a variant of autoepistemic logic which is a leading formal approach to knowledge representation systems. Cautious autoepistemic logic has some attractive properties that are not present in its ancestor. Firstly, every set of premises induces a unique set of conclusions. Moreover, this set of conclusions can be characterized iteratively. In this work, a proof theory is developed for cautious autoepistemic logic. The cumulativity of cautious autoepistemic reasoning is studied with the aid of the proof theory. It turns out that cautious autoepistemic reasoning is cumulative in restricted cases. Computational complexity of cautious autoepistemic logic is analyzed. Negative introspection which is the major subtask is shown to be a complete problem on the second level of the polynomial time hierarchy. An upper bound for the computational complexity of cautious autoepistemic reasoning is provided. This implies that such reasoning is in the worst case only slightly more complex than corresponding reasoning in the original autoepistemic logic. As an application, new model theoretic semantics for logic programs is provided in terms of cautious autoepistemic logic. The proposed cautious semantics is compared with the well-founded semantics. Results indicate that the cautious model is more accurate than the well-founded model. An alternative approach is also studied. It leads to a different model that also subsumes the well-founded model.

Keywords: nonmonotonic reasoning, autoepistemic logic, computational complexity, logic programming, well-founded semantics

Printing: TKK Monistamo; Otaniemi 1994

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

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

Contents

List of Figures

1 Introduction

Many applications in Artificial Intelligence (AI) involve representing *knowledge* in some form or another. Typically such a system consist of a *knowledge base* which is data structure capable of storing pieces of information and an *inference engine* which searches for answers to the queries made by the users of the system. However, it is often difficult to give a semantical justification for the inferences made by the engine. A reason for this is that reasoning in knowledge representation (KR) systems differs from classical logical reasoning in a profound way: reasoning in KR systems is nonmonotonic whereas classical logical reasoning is monotonic. Areas where nonmonotonic reasoning is a crucial feature include *logic programming* [19], *deductive databases* [1], *inheritance reasoning* [30], *truth maintenance systems* [6] and *diagnosis* [29].

Nonmonotonic logics [10] have been devised in order to semantically characterize such nonclassical reasoning. These logics can be considered as formal models of KR systems. They all use a logical language for knowledge representation in a declarative way. This leads to knowledge representation mechanisms that are independent of the implementation. In this work, we investigate *cautious autoepistemic logic* (CAEL) which is one of the formal models of knowledge representation. The following topics are of our main interest. Firstly, we study structural properties of cautious autoepistemic reasoning. Then we estimate time requirements of such reasoning by means of complexity theory. Finally, as an application we study logic programming with the aid of CAEL.

Let us start by an example of a KR system. The purpose of the example is twofold. First of all, it demonstrates important properties of KR systems. Moreover, by further development of the example we encounter problems that often arise in such systems. We have encoded our intuition on drinking habits of engineers to a knowledge base (1). The meaning of the provided rule is that engineers who are not known to be teetotallers are beer-drinkers. Furthermore, we have specified that John is an engineer by incorporating the fact `engineer(john)`.

```
IF engineer(x) AND not(teetotaller(x))
  THEN beer-drinker(x).                    (1)
FACT engineer(john).
```

Knowledge bases are typically managed through a collection of procedures including maintenance and capability to answer queries. We assume that we have implemented a forward chaining inference engine that derives new facts using the rules of our sample base. The engine maintains a set of facts against which the conditions of the rules are tested. This procedure starts its operation from the set of facts that is explicitly presented in the knowledge base.

The new facts derived by the procedure can be seen as implicit knowledge represented in the knowledge base.

Consider then answering the query `beer-drinker(john)`. To answer this query our system checks whether it can derive the fact using the rules and the facts of the knowledge base. It starts with the only explicit fact `engineer (john)`. Because the fact `teetotaller(john)` is not included in this set, the procedure applies the rule and adds the fact `beer-drinker(john)` to the set of facts it maintains. Consequently, the query is answered "yes". However, we might later on update our knowledge base and add the fact `teetotaller(john)` meaning that John *is* a teetotaller. Let us now reconsider the operation of the query answering procedure. The test that "`teetotaller(john)` is not present in the set of facts" fails this time preventing the application of the rule. In this way the previous conclusion that John is a beer-drinker is retracted and the corresponding query is given a negative answer. Despite the simplicity of our example, it has demonstrated two important aspects of KR systems in general:

- *Reasoning in such systems is often nonmonotonic.*

  Classical reasoning is monotonic in the sense that the set of conclusions grows monotonically when the set of premises is extended. However, we encountered a different behavior with our example. When we extended our knowledge base with the fact `teetotaller(john)` the previous conclusion `beer-drinker(john)` was lost. The reason for such a behavior is that the conclusion was based on failure to derive the fact that John is a teetotaller. This type of reasoning lacking the monotonicity property is known as nonmonotonic reasoning. Such reasoning forms an important field of research because the phenomenon is present in many AI applications.

- *Default knowledge supports reasoning in the lack of information.*

  Initially, our knowledge contained no information about the drinking habits of John. The *default* conclusion `beer-drinker(john)` was therefore based on *incomplete information*. Thus we can consider our rule as *default knowledge* about drinking habits of engineers. Default knowledge makes inferences possible when no explicit information is available as demonstrated by our example. Therefore retracting conclusions upon new pieces of information is justified. The use of default knowledge is also a matter of *efficiency* of knowledge representation. If engineers who are teetotallers are rare, it is reasonable to save storage space and encode these as exceptions in our knowledge base.

The facts and rules of our knowledge base have a somewhat different purposes in the system. The facts `engineer(John)` and `teetotaller(John)` describe the relations among the objects of the domain. These relations can be seen as a conventional *relational database* (extensional knowledge). The rules (intensional knowledge) describe how new facts can be inferred, i.e how to use the information contained in the relational database. This classification distinguishes *deductive databases* [1] from relational ones.

Let us now further develop our example by incorporating knowledge about the drinking habits of salesmen. To our knowledge, salesmen are teetotallers if they are not beer-drinkers. We have also heard that Jill is both an engineer and a salesman by profession. The resulting knowledge base has the structure given in (2).

```
IF engineer(x) AND not(teetotaller(x))
  THEN beer-drinker(x).
IF salesman(x) AND not(beer-drinker(x))
  THEN teetotaller(x).
FACT engineer(jill).
FACT salesman(jill).
```
(2)

It is quite reasonable that a person has several professions. However, this raises a *conflict* between the two rules and the order in which the rules are applied becomes significant. Consequently, there are two possible sets of conclusions associated with our knowledge base. The conclusion `beer-drinker (jill)` is contained in one and `teetotaller(jill)` in the other set of conclusions. Our query answering procedure might apply rules in the order they appear in the database and therefore drift to the first set of conclusions. Let us complete our list with two more characteristics:

- *The set of conclusions may not be unique.*

  This problem can be approached by two strategies. In the *cautious strategy*, the intersection of the sets of conclusions is considered as the intended set of conclusions. In our example, this means that we *do not know* whether Jill is a beer-drinker or a teetotaller. But undoubtedly, she is an engineer and a salesman. In the *brave strategy*, one of the sets of conclusions is preferred over the others. The choice may depend on some additional information that is acquired outside the system. In our example, we might ask Jill to clear the state of affairs.

- *Sometimes no set of conclusions can be identified.*

  In a complex knowledge base, the rules might interact such that only assuming that `not(property(a))` allows us to infer `property(a)`. If this is case, it is difficult to identify a set of conclusions: the fact `property(a)` should be included exactly when it is not included (as indicated by the rules) and vice versa.

It is possible to implement another query answering procedure that tries to apply the rules of the knowledge base (2) *backwards*. If the procedure is given the query `beer-drinker(jill)` other peculiarities arise. To find out whether the fact `beer-drinker(jill)` can be inferred the procedure checks whether the fact `teetotaller(jill)` can be inferred. But checking this requires recursively checking whether the fact `beer-drinker(jill)` can be inferred. In this manner, the procedure loops forever. A similar behavior is encountered if we translate our knowledge base into a *logic program* [19] and present the query `beer-drinker(jill)` to the interpreter. Thus depending on the query answering procedure used, different answers (if any) may be obtained. This is where the formal models of KR become important. The correct sets of conclusions associated with the knowledge base can be specified using such models independently of the query answering procedure. This speaks for the declarative ways of knowledge representation over the procedural ones. Classical logic is an insufficient model for KR systems because classical logic is monotonic, but reasoning in KR systems nonmonotonic. To capture nonmonotonic reasoning used in KR new kinds of logics have been developed. Leading such nonmonotonic logics include *circumscription* [21], *default logic* [28] and *autoepistemic logic* (AEL) [23]. We have chosen the autoepistemic framework because reasoning in many NMR systems can be presented as a form of autoepistemic reasoning. Furthermore, autoepistemic logic offers a uniform logical language for knowledge representation. Autoepistemic logic models the sets of beliefs of an introspective agent given a set of premises as its initial assumptions. These sets of beliefs are called *stable expansions* of the premises. The connection to KR systems is the following: the knowledge base forms the premises on which the agent (inference engine) performs introspection (answers queries). Stable expansions are the possible sets of conclusions associated with the knowledge base.

In this work, we analyze cautious autoepistemic logic [15, 16] which is a variant of the Moore's autoepistemic logic. CAEL differs from AEL in the principle used for negative introspection. This gives rise to *cautious expansions* that are the possible sets of conclusions given the premises such as the knowledge base above. CAEL has some attractive features that are not present in the original AEL.

1. Every set of premises has a unique cautious expansion associated with it. This solves the problems related to the existence of a set of conclusions. In our KR example (2) the unique cautious expansion contains the facts `engineer(jill)` and `salesman(jill)`, but the other two are left out. This justifies the name of the logic as the cautious reasoning strategy is evidently followed.

2. The unique cautious expansion can be seen as an *approximation* of the

intersection of stable expansions. Referring again to our example, we concluded that the knowledge base (2) has two alternative sets of conclusions associated with it. These correspond to the stable expansions of the knowledge base in the original AEL.

3. Furthermore, the unique cautious expansion can be constructed by an iterative procedure. This gives a *constructive* nature for cautious autoepistemic reasoning.

These features of CAEL are already established in [15].

1.1 The Research Problems

This work is concentrated on analyzing the properties of cautious autoepistemic reasoning. In particular, we are interested in computational complexity and structural properties. As an interesting application area we study the semantics of logic programs.

Computational Complexity of NMR

Recently, a lot of research effort has been devoted to the complexity analysis of NMR [7, 8, 13, 26]. This research is aimed at understanding the inherent computational complexity of NMR and developing efficient forms of NMR. Two orthogonal sources of computational complexity have been identified in the major formalisms: circumscription, default logic and autoepistemic logic. The one is the computational complexity of the underlying monotonic inference. The other is *conflict resolution* that was demonstrated by our earlier example (2). From this point of view, NMR is computationally more complex than similar monotonic reasoning. The complexity results indicate that if efficient forms of NMR are to achieved both sources of computational complexity have to be addressed.

Computational complexity of NMR can be reduced by imposing syntactic restrictions on the knowledge representation language. Niemelä and Rintanen have demonstrated this effect by identifying tractable subclasses of autoepistemic theories [27]. This leads to a natural trade-off situation: there is a connection between the expressibility of the knowledge representation language and the computational complexity of the required nonmonotonic inference. Another possibility is to use approximations of NMR. CAEL falls in this approximative category: the unique cautious expansion can be seen as an approximation of the intersection of *stable expansions* of original AEL. This justifies the name of the logic which therefore follows the cautious reasoning strategy. One of our interests is to examine how the weakened negative introspection scheme of CAEL affects the computational complexity of autoepistemic reasoning. It seems also possible that the computational complexity of autoepistemic reasoning can be reduced by introducing alternative principles for negative introspection.

Structural Properties of NMR

The structural properties of nonmonotonic consequence relations have also been of interest. A purely formal framework has been suggested by Kraus, Lehman and Magidor [18]. They introduce structural rules for nonmonotonic consequence relations. These rules can be used to classify NMR systems and to specify the desired properties of these systems in general. One of the most interesting properties is *cumulativity*. In a cumulative reasoning system, drawn conclusions can be used as "lemmas" and be freely inserted to the premises to support forthcoming inferences. The conditions on which CAEL satisfies the cumulativity property are studied in this work.

Applications in Logic Programming

Logic programming [19] is one of the successful NMR applications. In a simple form, logic programs consist of facts and unidirectional inference rules that allow derivation of new facts. This class of logic programs is known as *definite logic programs*. They have a clear model-theoretic declarative semantics. A sound and complete proof procedure called *SLD-resolution* gives a definite program its *procedural meaning*: the computation of the program is determined by its proof procedure. NMR comes into play when a form of negation is introduced. In *general logic programs* negation is understood as a *failure to prove*. Negation as failure is beneficial in many programming tasks, but unfortunately it leads to complications on semantical aspects. In this work, we provide a model-theoretic semantics for general logic programs in terms of the unique cautious expansion.

A lot of research work has been devoted to finding alternative semantics for general logic programs. Clark [4] proposed a semantics based on the idea of *predicate completion*. However, there are general logic programs whose predicate completion is inconsistent. In this way, the semantics proposed by Clark leaves the meaning of some general logic programs open. One possibility is to define the meaning of general logic programs in terms of autoepistemic logic. This involves translating general logic programs into autoepistemic theories. If the original AEL is used, this leads to a semantics equivalent to *stable model semantics* proposed by Gelfond and Lifschitz [12]. Unfortunately, such models are not unique and do not exist for every general logic program. Van Gelder et al [32] have approached the problem in terms of *well-founded models* logic programs. Every general logic program has a unique well-founded model. Here we have an analogy to CAEL where every set of premises has a unique cautious expansion associated with it. This supports the idea that a semantics of general logic programs can be provided in terms of CAEL.

1.2 The Contributions of the Work

In the following we present the contributions of this work to the previously described research problems. The structure of the work is also outlined.

- *Computational complexity of cautious autoepistemic reasoning*

  Various reasoning tasks of CAEL are mapped to the polynomial time hierarchy (PH) [2, 17]. To single out the two most important results, the decision problem corresponding to the negative introspection capability of the agent is shown to be a complete problem on the second level of the hierarchy. Determining whether a sentence belongs to the unique cautious expansion can be seen as the major reasoning task of CAEL. Limits for the computational complexity of the corresponding decision problem are given in terms of PH. These indicate that the problem is reducible to a problem on the second level of PH in polynomial time.

- *A proof theory for cautious autoepistemic reasoning*

  In autoepistemic logic and many other nonmonotonic logics, the sets of conclusions are defined using fixed point equations. This makes it difficult to develop a proof theory for such logics. However, cautious autoepistemic logic is exceptional in this sense. In this work, we provide cautious autoepistemic logic with a proof theory that resembles the one of classical logic. A proof rule for negative introspection has a central role in the proof system.

- *Cumulativity of cautious autoepistemic reasoning*

  The cumulativity of reasoning with respect to the unique cautious expansion is analyzed. It turns out that reasoning with respect to the unique expansion is not cumulative in general. The cumulativity property consists of satisfaction of rules *cut* and *cautious monotonicity*. The results presented in this work indicate that reasoning with respect to the unique expansion satisfies the cut rule. However, cautious monotonicity is satisfied only if additional restrictions are present.

- *New models for general logic programs*

Two formal models for general logic program are proposed. They are naturally defined to correspond to the unique cautious expansions of two slightly different translations of a logic program into autoepistemic logic. The resulting models are analogous to the well-founded model. The analysis presented in this work shows that the proposed models are at least as accurate as the well-founded model. The provided examples show that the proposed models may differ from the well-founded model. The two models are mutually incomparable in strength.

Structure of the Work

In **Section 2** we introduce cautious autoepistemic logic in detail. A finitary characterization is provided for the cautious expansions. Some supplementary results are provided to support forthcoming proofs. The proof-theoretical properties of CAEL are analyzed in **Section 3**. A notion of provability is developed. It is then used to analyze cumulativity of cautious autoepistemic reasoning. In **Section 4** decision procedures are devised for the major reasoning tasks of CAEL. The complexity of these problems is analyzed in terms of the hierarchy. A model-theoretic semantics for general logic programs is proposed in **Section 5**. The well-founded model is compared with the proposed cautious model. An alternative cautious model is devised and the proposed models are compared. Conclusions of the work are presented in **Section 6**.

2 Cautious Autoepistemic Logic

In this section, we present details of CAEL following [15, 16]. Some supplementary results needed in the rest of the work are also provided. Let us begin by introducing the original AEL proposed by Moore [22]. In AEL, the *autoepistemic language* $\mathcal{L}_{ae}$ is used for knowledge representation. This language is the propositional language extended with sentences of the form $L\phi$. The purpose of the *modal operator* L is to formalize a notion of belief. Autoepistemic logic models sets of beliefs of an agent given a set of premises as its initial assumptions. These sets of beliefs – called *stable expansions* – are defined by the fixed point equation given in Definition 2.1. The agent is capable of *positive and negative introspection*: it decides what it believes and disbelieves on the basis of the logical consequences of its premises Σ and its beliefs $L\Delta \cup \neg L\overline{\Delta}$.

Definition 2.1 (Moore [22]) *Let Σ be a set of premises. A set of sentences $\Delta \subseteq \mathcal{L}_{ae}$ is a stable expansion of Σ if and only if*

$$\Delta = \mathrm{Cn}(\Sigma \cup L\Delta \cup \neg L\overline{\Delta})$$

where $\overline{\Gamma} = \mathcal{L}_{ae} - \Gamma$, $\neg\Gamma = \{\neg\phi \,|\, \phi \in \Gamma\}$ and $L\Gamma = \{L\phi \,|\, \phi \in \Gamma\}$ for any set of sentences Γ.

If the agent believes the sentence ϕ, it includes $L\phi$ to its beliefs by positive introspection. If not, it includes $\neg L\phi$ to its beliefs by negative introspection. Stable expansions are sets of sentences that are reproduced by these introspection principles. Let us demonstrate the applicability of AEL in knowledge representation by Example 2.2. It provides a translation of our knowledge base (2) into autoepistemic sentences. The atomic propositions e, s, t, b denote that Jill is an engineer, a salesman, a teetotaller and a beer-drinker, respectively.

Example 2.2 Let Σ be the set of premises $\{Le\wedge\neg Lt \rightarrow b, Ls\wedge\neg Lb \rightarrow t, e\wedge s\}$. It has two stable expansions. In the first, the agent believes e, s and b and it does not believe t. Therefore it includes the sentences Le, Ls, Lb and $\neg Lt$ in its beliefs. Notice that the beliefs e, s and b are logical consequences of Σ and the beliefs of the agent, but t is not. Therefore the set of beliefs corresponds to a stable expansion of Σ. In the other expansion, the roles of t and b change.

This example demonstrates how autoepistemic reasoning follows the characteristics outlined in the introduction: there are two alternative sets of conclusions. If we apply the cautious reasoning strategy only the beliefs that are contained in both stable expansions (such as e and s above) are be accepted. Let us now proceed to the definition of CAEL and introduce some notation and terminology. The sentences of the form $L\phi$ are called *quasi-atoms*. *Quasi-literals* are either quasi-atoms or their negations. We define a function *qatom* for quasi-literals and sets of quasi-literals such that $\mathrm{qatom}(\chi) = \mathrm{qatom}(\neg\chi) = \chi$ for any quasi-atom χ. The sentence χ' is called the *base sentence* of the quasi-literals $L\chi'$ and $\neg L\chi'$. The *L-subformulae* of a sentence ϕ are denoted by $\mathrm{Sf}^{\mathrm{L}}(\phi)$. For a quasi-atom $L\phi$, $\mathrm{Sf}^{\mathrm{L}}(L\phi) = \{L\phi\} \cup \mathrm{Sf}^{\mathrm{L}}(\phi)$. For other forms of sentences we apply structural recursion: e.g. $\mathrm{Sf}^{\mathrm{L}}(\phi \wedge \psi) = \mathrm{Sf}^{\mathrm{L}}(\phi) \cup \mathrm{Sf}^{\mathrm{L}}(\psi)$. For example, $\mathrm{Sf}^{\mathrm{L}}(LLp \wedge Lr) = \{Lp, LLp, Lr\}$. This operation is naturally extended for sets of autoepistemic sentences.

The propositional consequence relation $\models$ is extended to its autoepistemic counterpart $\models_{\mathrm{ae}}$ by treating all quasi-atoms like atomic propositions. In autoepistemic logic, models are pairs $\langle I, \Delta \rangle$ where I is a standard propositional interpretation and Δ determines the truth values of quasi-atoms such that $\langle I, \Delta \rangle \models_{\mathrm{ae}} L\phi$ if and only if $\phi \in \Delta$. We let $\mathrm{Cn}(\Sigma)$ denote the set of logical consequences $\{\phi \mid \Sigma \models_{\mathrm{ae}} \phi\}$.

2.1 Cautious Expansions

In the original AEL, the agent believes the logical consequences of its premises and beliefs and disbelieves (negative introspection) everything else. In cautious autoepistemic logic, negative introspection is based on *enumeration-based introspection* [25]. The agent examines sentences in an order determined by an enumeration ε of the *enumeration base* Γ. In CAEL, the enumeration base contains sentences whose belief status has not been yet determined. Enumeration-based introspection is then used to construct sets of beliefs in the following manner. At each step an introspection rule is applied to a sentence in ε. The introspection of the agent is based directly on logical consequences of the premises and the results of previous introspection steps that specify a set of beliefs of the agent. If the sentence ϕ under introspection belongs to the set of beliefs, $L\phi$ is inferred (positive introspection) and otherwise $\neg L\phi$ is inferred (negative introspection). In this way a set of quasi-literals $\mathrm{B}^{\varepsilon}(\Sigma)$ is constructed. The corresponding set of beliefs is denoted by $\mathrm{EB}^{\varepsilon}(\Sigma)$. These concepts are clarified in the example 2.4 that uses the same translation as Example 2.2.

Definition 2.3 (Niemelä [25]) *Let Σ be a set of premises and $\varepsilon = \psi_1, \psi_2, \ldots$ an enumeration of a subset Γ of $\mathcal{L}_{ae}$. Let $|\varepsilon|$ denote the length of ε. Define $\mathrm{B}^{\varepsilon}_0(\Sigma) = \emptyset$ and $\mathrm{B}^{\varepsilon}_{i+1}(\Sigma)$ for $i = 0, 1, \ldots, |\varepsilon|$ as follows:*

$$\mathrm{B}^{\varepsilon}_{i+1}(\Sigma) = \begin{cases} \mathrm{B}^{\varepsilon}_i(\Sigma) \cup \{L\psi_{i+1}\}, & \text{if } \Sigma \cup \mathrm{B}^{\varepsilon}_i(\Sigma) \models_{\mathrm{ae}} \psi_{i+1} \\ \mathrm{B}^{\varepsilon}_i(\Sigma) \cup \{\neg L\psi_{i+1}\}, & \text{if } \Sigma \cup \mathrm{B}^{\varepsilon}_i(\Sigma) \not\models_{\mathrm{ae}} \psi_{i+1} \end{cases}$$

Finally let

$$\mathrm{B}^{\varepsilon}(\Sigma) = \begin{cases} \mathrm{B}^{\varepsilon}_{|\varepsilon|}(\Sigma), & \textit{if } |\varepsilon| \textit{ is finite} \\ \bigcup_{i=0}^{\infty} \mathrm{B}^{\varepsilon}_{i}(\Sigma), & \textit{otherwise} \end{cases}$$

and the enumeration-based set of beliefs $\mathrm{EB}^{\varepsilon}(\Sigma) = \{\phi \mid \Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi\}$.

Example 2.4 Let Σ be the set of premises $\{Le \wedge \neg Lt \rightarrow b, Ls \wedge \neg Lb \rightarrow t, e \wedge s\}$ and ε the enumeration e, s, t, b. Because e is a logical consequence of Σ, $\mathrm{B}^{\varepsilon}_{1}(\Sigma)$ equals to Le. Quite similarly, $\mathrm{B}^{\varepsilon}_{2}(\Sigma)$ is constructed by adding Ls to $\mathrm{B}^{\varepsilon}_{1}(\Sigma)$ because s is a logical consequence of $\Sigma \cup \{Le\}$. The negative quasi-literal $\neg Lt$ is included in $\mathrm{B}^{\varepsilon}_{3}(\Sigma)$, since t is not a logical consequence of $\Sigma \cup \mathrm{B}^{\varepsilon}_{2}(\Sigma)$. Finally, $\mathrm{B}^{\varepsilon}_{4}(\Sigma)$ is formed by adding Lb to $\mathrm{B}^{\varepsilon}_{3}(\Sigma)$. Thus $\mathrm{B}^{\varepsilon}(\Sigma) = \{Le, Ls, \neg Lt, Lb\}$. Notice that we encountered exactly the same quasi-literals in Example 2.2. Because b is a logical consequence of $\Sigma \cup \{Le, Ls, \neg Lt, Lb\}$, b belongs to the set of beliefs $\mathrm{EB}^{\varepsilon}(\Sigma)$. Naturally, other enumerations give rise to different sets of beliefs $\mathrm{EB}^{\varepsilon}(\Sigma)$.

In CAEL, negative introspection is defined in terms of non-derivability with respect to enumeration-based introspection. Given a set of premises Σ and an enumeration base Γ, the agent disbelieves a sentence ϕ if and only if ϕ does not belong to any set of beliefs $\mathrm{EB}^{\varepsilon}(\Sigma)$ based on an enumeration of Γ. This central concept is formalized as a consequence relation in Definition 2.5. If it is the case that $\Sigma \not\dashv|^{\Gamma}_{\mathrm{ae}} \phi$, i.e. there is an enumeration ε of Γ such that $\phi \in \mathrm{EB}^{\varepsilon}(\Sigma)$, we say that ε *refutes* ϕ or ε is a *refuting enumeration* for ϕ. In Example 2.4, the sentence b is refutable from Σ using an enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. Thus we let $\mathrm{Ir}(\Sigma, \Gamma)$ denote the set $\{\phi \mid \Sigma \dashv|^{\Gamma}_{\mathrm{ae}} \phi\}$, i.e. the set of sentences that are *irrefutable* from Σ using enumerations of the *enumeration base* Γ. We let $\varepsilon \cdot \varepsilon'$ denote the *concatenation* of two enumerations ε and ε' (ε is necessarily finite). This operation is often needed to build refuting enumerations for sentences.

Definition 2.5 *Let* Σ *be a set of premises,* Γ *an enumeration base and* ϕ *any sentence. Then the relation*

$$\Sigma \dashv|^{\Gamma}_{\mathrm{ae}} \phi$$

holds if and only if there is no enumeration ε *of* Γ *such that* $\phi \in \mathrm{EB}^{\varepsilon}(\Sigma)$.

Cautious expansions represent the possible sets of beliefs of the agent. Formally they are pairs $\langle \Delta^{+}, \Delta^{-} \rangle$ where Δ^{+} contains the sentences that are *believed* and Δ^{-} those *disbelieved* by the agent. The positive introspection principle of original AEL is adopted in cautious autoepistemic reasoning: the agent believes sentences that logically follow from its premises Σ and beliefs $L\Delta^{+} \cup \neg L\Delta^{-}$. The agent disbelieves sentences that are irrefutable from $\Sigma \cup L\Delta^{+} \cup \neg L\Delta^{-}$ using enumerations of $\mathcal{L}_{ae} - (\Delta^{+} \cup \Delta^{-})$. The irrefutability of a sentence ϕ guarantees that adding $\neg L\phi$ to the premises is safe: the agent cannot derive ϕ by means of enumeration-based introspection. This is where the cautiousness of the agent becomes forth.

Definition 2.6 *Let Σ be a set of premises. A pair of subsets $\langle \Delta^+, \Delta^- \rangle$ of $\mathcal{L}_{ae}$ is a cautious (autoepistemic) expansion of Σ if and only if $\Delta^+ = \mathrm{Cn}(\Sigma \cup L\Delta^+ \cup \neg L\Delta^-)$ and $\Delta^- = \mathrm{Ir}(\Sigma \cup L\Delta^+ \cup \neg L\Delta^-, \mathcal{L}_{ae} - (\Delta^+ \cup \Delta^-))$.*

In the conventional AEL, the agent disbelieves every sentence that it does not believe. In this sense, we have *weakened* the negative introspection capability of the agent. There may exist *undetermined* sentences ϕ that are neither believed nor disbelieved by the agent, i.e. $\phi \not\in \Delta^+ \cup \Delta^-$. This phenomenon is demonstrated in Example 2.7. Such undetermined sentences favor the existence of expansions in CAEL. Notice that in Example 2.7 the set of premises Σ does not have stable expansions: if the agent does not believe the sentence p it should disbelieve p and vice versa. In fact, cautious expansions generalize the *stable expansions* of the original autoepistemic logic. To see this, notice that the requirements of Definition 2.6 reduce to the fixed point equation of Definition 2.1 when Δ^+ equals to $\overline{\Delta^-}$ and consequently the enumeration base $\mathcal{L}_{ae} - (\Delta^+ \cup \Delta^-)$ is empty.

Example 2.7 The set of premises $\Sigma = \{\neg Lp \rightarrow p\}$ has a cautious expansion where p is undetermined. Leaving p undetermined is justifiable because the agent cannot decide the status of p. If p does not belong to $\Delta^+ \cup \Delta^-$ it is included in the enumeration base $\mathcal{L}_{ae} - (\Delta^+ \cup \Delta^-)$. Notice that neither Lp nor $\neg Lp$ is included in $L\Delta^+ \cup \neg L\Delta^-$. The sentence p is refuted by any enumeration that begins with p ($\neg Lp$ is added immediately). Moreover, p is not a logical consequence of $\Sigma \cup L\Delta^+ \cup \neg L\Delta^-$.

2.2 Finitary Characterization

Cautious expansions are infinite sets of sentences which makes them somewhat hard to present and handle. Fortunately, cautious expansions can be provided a finitary characterization based on Niemelä style full sets [26, 16]. Full sets are useful in two ways. Firstly, they provide a convenient representation of cautious expansions. Moreover, properties of cautious expansions can be examined using full sets. The idea is to restrict reasoning to the sentences that appear in the L-subformulae $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ of the premises Σ. Furthermore, to perform the negative introspection it is sufficient to use enumerations of subsets of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. If $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is finite, the full sets and the needed enumerations are finite. The relations of Definition 2.8 provide the finitary mechanism for deciding whether a formula ϕ is believed or disbelieved by the agent, respectively. The two consequence relations are mutually exclusive, i.e. $\langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \phi$ implies $\langle\Sigma,\Lambda\rangle \not\models\!\!\!\dashv_{\mathrm{L}} \phi$. The sentences in the *enumeration base* $\mathrm{Eb}(\Sigma,\Lambda,\phi)$ are used in the irrefutability relation. Sentences ϕ that contain nested belief operators may have subbeliefs $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$.

Definition 2.8 *For a set of premises Σ, a set of quasi-literals Λ, and $\phi \in \mathcal{L}_{ae}$:*

1. $\langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \phi$ *iff* $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi) \models_{\mathrm{ae}} \phi$
2. $\langle\Sigma,\Lambda\rangle \dashv_{\mathrm{L}} \phi$ *iff* $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi) \dashv^{\mathrm{Eb}(\Sigma,\Lambda,\phi)}_{\mathrm{ae}} \phi$

where

$$\mathrm{SB}_{\Sigma,\Lambda}(\phi) = \{L\chi \in \mathrm{Sf}^{\mathrm{L}}(\phi) \mid \langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \chi\} \cup \{\neg L\chi \in \neg\mathrm{Sf}^{\mathrm{L}}(\phi) \mid \langle\Sigma,\Lambda\rangle \dashv_{\mathrm{L}} \chi\}$$

and the enumeration base $\mathrm{Eb}(\Sigma,\Lambda,\phi)$ *is*

$$\{\chi \mid L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi))\}.$$

Definition 2.9 *Let Σ be a set of premises. A set $\Lambda \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ is Σ-full if and only if the following conditions hold for every $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$:*

1. $L\chi \in \Lambda$ *iff* $\langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \chi$ 2. $\neg L\chi \in \Lambda$ *iff* $\langle\Sigma,\Lambda\rangle \dashv_{\mathrm{L}} \chi$

The Σ-*full sets* [1] are in one-to-one correspondence with cautious expansions of Σ [16, Theorem 6]. The mappings from between cautious expansions and full sets are provided in Theorem 2.10. Here $\Pi_{\Lambda}(\Gamma)$ denotes the set

$$\{L\chi \in \Gamma \mid L\chi \in \Lambda\} \cup \{\neg L\chi \in \Gamma \mid L\chi \in \Lambda\},$$

i.e. the *reduction* of a set of quasi-literals Γ with respect to a set of quasi-atoms Λ. For instance, $\Pi_{\{Lt,Lb\}}(\{Le, Ls, \neg Lt, Lb\}) = \{\neg Lt, Lb\}$.

Theorem 2.10 (Janhunen [15]) *Let Σ be the set of premises. The cautious expansions of Σ and Σ-full sets are in bijective relationship: (i) given a cautious expansion $\langle\Delta^+,\Delta^-\rangle$ of Σ, the corresponding Σ-full set is $\Lambda = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$ and (ii) given a Σ-full set Λ, the corresponding cautious expansion is $\langle\{\phi \mid \langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \phi\}, \{\phi \mid \langle\Sigma,\Lambda\rangle \dashv_{\mathrm{L}} \phi\}\rangle$.*

Full sets provide means for a systematic search for cautious expansions. One has to consider all consistent subsets of $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and check whether the conditions of Definition 2.9 hold. We end this section with Examples 2.11 and 2.12. The latter deals with our beer-drinker example in terms of cautious autoepistemic reasoning.

[1] In [16] a slightly different consequence relations were used in the definition of full sets. However, the consequence relations coincide for sentences χ such that $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$.

Example 2.11 Let Σ be the set of premises $\{Lp \rightarrow p\}$. Then there are three candidates for Σ-full sets: $\emptyset$, $\{\neg Lp\}$ and $\{Lp\}$. Consider the first one in further detail. As $\neg Lp$ does not belong to it $\langle \Sigma, \{\neg Lp\}\rangle \not\vdash_{\mathrm{L}} p$ should hold. This is the case if and only if there is an enumeration $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma) - \mathrm{qatom}(\emptyset)\}$ (this equals to $\{p\}$) such that $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} p$. Notice that $\mathrm{SB}_{\Sigma,\emptyset}(p) = \emptyset$ as $\mathrm{Sf}^{\mathrm{L}}(p) = \emptyset$. The only possible enumeration is $\varepsilon = p$. Enumeration-based introspection yields $\mathrm{B}^{\varepsilon}(\Sigma) = \{\neg Lp\}$ as p is not a logical consequence of Σ. Neither does it follow from $\Sigma \cup \{\neg Lp\}$. This implies that p is irrefutable from Σ using enumerations of $\{p\}$. Consequently $\emptyset$ is not Σ-full. A similar analysis on the last two candidates shows that they are both Σ-full.

Example 2.12 Let Σ be the set of premises $\{Le \wedge \neg Lt \rightarrow b, Ls \wedge \neg Lb \rightarrow t, e \wedge s\}$. Then there are three Σ-full sets out of 81 candidates: $\{Le, Ls, \neg Lt, Lb\}$, $\{Le, Ls, Lt, \neg Lb\}$ and $\{Le, Ls\}$. The third is of special interest. It is the intersection of the others and therefore follows the cautious reasoning strategy. The sentences b and t are refutable from $\Sigma \cup \{Le, Ls\}$ by the enumerations $\varepsilon_1 = t, e$ and $\varepsilon_2 = e, t$.

2.3 A Groundedness Condition for Cautious Expansions

Some of the cautious expansions are in a sense weakly grounded in the premises. Recall from Example 2.7 that $\{Lp \rightarrow p\}$ has cautious expansions corresponding to the full sets $\{Lp\}$ and $\{\neg Lp\}$. In the first cautious expansion, the belief in p is supported by the belief itself and thus weakly grounded in the premises. Marek and Truszczyński [20] have used a nonstandard modal logic $\mathbf{N}$ to rule out such weakly grounded *stable* expansions. This logic corresponds to propositional calculus augmented with the necessitation rule $\phi \vdash L\phi$. For instance, we can derive r from the premises $\{p, Lp \rightarrow r\}$ by applying the necessitation rule to p. We write $\Sigma \vdash_{\mathbf{N}} \phi$ to denote the $\mathbf{N}$-*derivability* of the sentence ϕ from the premises Σ. Similarly, $\mathrm{Cn}_{\mathbf{N}}(\Sigma)$ denotes the set of $\mathbf{N}$-derivable sentences from Σ. Following Marek and Truszczyński we impose a groundedness condition on cautious expansions in Definition 2.13. The idea behind the condition is that the beliefs Δ^+ should be $\mathbf{N}$-derivable from the premises and negative beliefs $\neg L\Delta^-$. This excludes cases where beliefs support themselves.

Definition 2.13 (Marek and Truszczyński [20]) *Given a set of premises Σ, a cautious expansion $\langle \Delta^+, \Delta^- \rangle$ of Σ $\mathbf{N}$-grounded if and only if it satisfies the equation $\Delta^+ = \mathrm{Cn}_{\mathbf{N}}(\Sigma \cup \neg L\Delta^-)$.*

The one-to-one correspondence of cautious expansions and full sets suggests a similar groundedness condition for full sets. In Definition 2.14, we give the condition suggested by Gottlob. It can be shown [15, Theorem 3.12] that a Σ-full set is N-grounded if and only if the corresponding cautious expansion of Σ is N-grounded. If we reconsider the $\{Lp \rightarrow p\}$-full set $\{Lp\}$ we notice that p is not N-derivable from Σ. Consequently, the corresponding cautious expansion does not satisfy the groundedness condition.

Definition 2.14 (Gottlob [14]) *Let Σ be a set of premises. A Σ-full set Λ is* N*-grounded if and only if it every $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ satisfies:*

$$L\chi \in \Lambda \text{ iff } \Sigma \cup [\Lambda]^{-} \vdash_{\mathbf{N}} \chi$$

where $[\Lambda]^{-} = \{\neg L\chi \mid \neg L\chi \in \Lambda\}$.

Restricting to N-grounded expansions has advantages: it leads to pleasant properties not present in Moore style autoepistemic logic. First of all, a partial order can be defined on N-grounded cautious expansions. Generally speaking, a cautious expansion $\langle \Delta_1^+, \Delta_1^- \rangle$ is smaller than or equal to $\langle \Delta_2^+, \Delta_2^- \rangle$ if and only if $\Delta_1^+ \subseteq \Delta_2^+$ and $\Delta_1^- \subseteq \Delta_2^-$. This partial order corresponds to *growth of information* in cautious expansions. The subset inclusion on N-grounded cautious expansions coincides with the subset inclusion on N-grounded Σ-full sets [16, Proposition 10]. In Example 2.12, the three cautious expansions are all N-grounded. The least one, namely the one that corresponds to the full set $\{Le, Ls\}$ is contained in the two other cautious expansions. Furthermore, the least N-grounded cautious expansion can be iteratively constructed by a monotonic operator [16, Theorem 14]. As a corollary of this every set of premises has a cautious expansion. This expansion is the intersection of all N-grounded cautious expansions. We shall also refer to the least N-grounded cautious expansion as the *least* expansion.

Let us now present the iterative construction for the least expansion. It is based on the operator given in Definition 2.15. The iteration is started from $\emptyset$. The scope of introspection is restricted in the construction by fixing beforehand a set of quasi-atoms Q that satisfies $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq Q$ and $\mathrm{Sf}^{\mathrm{L}}(Q) \subseteq Q$. The purpose of the construction is to compute a part of the least expansion as characterized in Proposition 2.16. This proposition is an instance of [15, Proposition 3.25]. The least N-grounded Σ-full set (or the *least* Σ-full set for short) is constructed by fixing the set of quasi-atoms Q to $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$. It should also be mentioned that the other fixed points of the operator $\mathrm{FCAE}_{\Sigma,\mathrm{Sf}^{\mathrm{L}}(\Sigma)}$ are Σ-full.

Definition 2.15 *Let Σ be a set of premises and Q a set of quasi-atoms such that $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq Q$ and $\mathrm{Sf}^{\mathrm{L}}(Q) \subseteq Q$. Then define an operator taking a consistent set of quasi-literals $\Lambda \subseteq Q \cup \neg Q$ as its argument:*

$$\mathrm{FCAE}_{\Sigma,Q}(\Lambda) = \{L\chi \in Q \mid \Sigma \cup \Lambda \models_{\mathrm{ae}} \chi\} \cup \{\neg L\chi \in \neg Q \mid \Sigma \cup \Lambda \mathrel{=\!\!\mid}_{\mathrm{ae}}^{\mathrm{Eb}(Q,\Lambda)} \chi\}$$

where the enumeration base $\mathrm{Eb}(Q, \Lambda)$ *equals to* $\{\psi \mid L\psi \in Q - \mathrm{qatom}(\Lambda)\}$.

Proposition 2.16 (Janhunen [15]) *Let* Σ *be the set of premises and the pair* $\langle \Delta^+, \Delta^- \rangle$ *the least expansion of* Σ. *Then* $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$, *i.e. the least fixed point of the operator* $\mathrm{FCAE}_{\Sigma,Q}$, *equals to* $\Pi_Q(L\Delta^+ \cup \neg L\Delta^-)$.

Example 2.17 Let Σ be the set of premises $\{Le \wedge \neg Lt \rightarrow b, Ls \wedge \neg Lb \rightarrow t, e \wedge s\}$ of Example 2.12. Let us fix $Q = \mathrm{Sf}^{\,\mathrm{L}}(\Sigma) = \{Le, Ls, Lt, Lb\}$. The least Σ-full set is computed in two steps. The iteration is started from $\Lambda_0 = \emptyset$. An application of the operator $\mathrm{FCAE}_{\Sigma,\mathrm{Sf}^{\mathrm{L}}(\Sigma)}$ yields $\Lambda_1 = \{Le, Ls\}$. This is already the least fixed point of the operator as both t and b are refutable from $\Sigma \cup \{Le, Ls\}$ using enumerations of $\{t, b\}$. If we enlarge Q with LLe the iteration yields the least fixed point $\{Le, LLe, Ls\}$. This indicates by proposition 2.16 that the belief Le is also contained in the least expansion. It is therefore believed that e is believed.

The consequence relations $\models_{\mathrm{L}}$ and $=\!\!|_{\mathrm{L}}$ of definition 2.8 provide us a tool to determine the membership of sentences in the cautious expansion characterized by a full set. The irrefutability relation of Definition 2.15 is somewhat different. It uses a fixed scope Q in the introspection. Therefore the enumeration base $\{\psi \mid L\psi \in Q - \mathrm{qatom}(\Lambda)\}$ is used. We will use a shorthand $\langle \Sigma, \Lambda \rangle =\!\!|^{\mathrm{Q}}_{\mathrm{ae}} \chi$ for the relation $\Sigma \cup \Lambda =\!\!|^{\mathrm{Eb}(Q,\Lambda)}_{\mathrm{ae}} \chi$ of Definition 2.15. The set Q will be clear from the context whenever it is used. The importance of the relation $=\!\!|^{\mathrm{Q}}_{\mathrm{ae}}$ lies in that it provides us the link to the results on the iterative construction presented in [15]. However, it is not very practical to carry around such a set Q all the time. Therefore we write $\langle \Sigma, \Lambda \rangle =\!\!|_{\mathrm{ae}} \phi$ in the case Q equals to $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})$. The resulting enumeration base is $\{\psi \mid L\psi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda)\}$ that contains also the subformulae of ϕ. This is practical because deciding whether ϕ is irrefutable by means of the relation $=\!\!|^{\mathrm{Q}}_{\mathrm{ae}}$ requires a set Q such that $\mathrm{Sf}^{\,\mathrm{L}}(\phi)$ is contained in it. The connection between the relations $=\!\!|_{\mathrm{ae}}$ and $=\!\!|_{\mathrm{L}}$ is the following: $\langle \Sigma, \Lambda \rangle =\!\!|_{\mathrm{L}} \phi$ holds if and only if $\langle \Sigma, \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi) \rangle =\!\!|_{\mathrm{ae}} \phi$ holds. The connection between the irrefutability relations $=\!\!|_{\mathrm{ae}}$ and $=\!\!|^{\mathrm{Q}}_{\mathrm{ae}}$ is cleared in Lemma 2.26 of the supplementary results that follow.

2.4 Supplementary Results

We end this section by some auxiliary results that are needed in the forthcoming proofs. Lemma 2.18 shows how a set quasi-literals Λ can be reduced without affecting the logical consequence ϕ of $\Sigma \cup \Lambda$. A similar results is formulated for refuting enumerations in Lemma 2.19 which is a slight generalization of [15, Proposition 2.15].

Lemma 2.18 *Let Σ be a set of premises, Λ a consistent set of quasi-literals and ϕ any sentence. Then $\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi$ if and only if $\Sigma \cup \Lambda' \models_{\mathrm{ae}} \phi$ where $\Lambda' = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda)$.*

Proof. ($\Rightarrow$) Niemelä [26] has proved this direction with $\Lambda'' = \mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\phi\})^2$. Because $\mathrm{Sf}^{\mathrm{qL}}(\Sigma \cup \{\phi\}) \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$, the claim for Λ' follows by the monotonicity of $\models_{\mathrm{ae}}$. ($\Leftarrow$) This direction follows by the monotonicity of $\models_{\mathrm{ae}}$. $\square$

Lemma 2.19 *Let Σ be a set of premises, Λ a consistent set of quasi-literals and ϕ any sentence. Assume there is an enumeration ε of a subset Γ of $\mathcal{L}_{ae} - \{\psi \mid L\psi \in \mathrm{qatom}(\Lambda)\}$ such that $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \phi$ holds. Then there is an enumeration ε' of*

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda')\}$$

such that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon'}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \phi$ where $\Lambda' = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda)$.

Proof. Let $\varepsilon = \psi_1, \psi_2, ...$ be a (possibly infinite) enumeration of Γ and $\varepsilon' = \psi'_1, \psi'_2, ...$ its (possibly infinite) reduction with respect to

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})\}.$$

Let f be a function defined by the equalities $f(0) = 0$ and $\psi'_i = \psi_{f(i)}$ when $i > 0$. Intuitively, $f(i)$ gives the location of the i^{th} sentence of ε' in the enumeration ε.

Let us first show by induction on i that

$$\text{(i)} \qquad \mathrm{B}^{\varepsilon'}_i(\Sigma \cup \Lambda') = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda)).$$

The base case is clear, because both $\mathrm{B}^{\varepsilon'}_0(\Sigma \cup \Lambda')$ and $\mathrm{B}^{\varepsilon}_{f(0)}(\Sigma \cup \Lambda)$ are empty. Induction step follows. By the inductive hypothesis $\mathrm{B}^{\varepsilon'}_{i-1}(\Sigma \cup \Lambda')$ equals to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i-1)}(\Sigma \cup \Lambda))$. Because the sentences $\psi_{f(i-1)+1}, ..., \psi_{f(i)-1}$ do not belong to $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})\}$, the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i-1)}(\Sigma \cup \Lambda))$ equals to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda))$. Furthermore, Λ' is the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda)$. Thus $\Lambda' \cup \mathrm{B}^{\varepsilon'}_{i-1}(\Sigma \cup \Lambda')$ and $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda))$ are equal. Let us the prove inclusion in both directions. ($\subseteq$) Let $L\psi'_i \in \mathrm{B}^{\varepsilon'}_i(\Sigma \cup \Lambda')$. This implies $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon'}_{i-1}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \psi'_i$. Because $\Lambda' \cup \mathrm{B}^{\varepsilon'}_{i-1}(\Sigma \cup \Lambda')$ is a reduction of $\Lambda \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda)$ and $\psi'_i = \psi_{f(i)}$, it follows that $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \psi_{f(i)}$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus $L\psi_{f(i)} \in$

[2] Here $\mathrm{Sf}^{\mathrm{qL}}(\Sigma)$ denotes the quasi-L-subformulae of Σ. This operator differs from Sf^{L} in the way it handles sentences of the form $L\phi$. For the operator $\mathrm{Sf}^{\mathrm{qL}}$, $\mathrm{Sf}^{\mathrm{qL}}(L\phi) = \{L\phi\}$. Therefore $\mathrm{Sf}^{\mathrm{qL}}(\Sigma) \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma)$.

$\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda)$. Because $L\psi_{f(i)}$ is a member of $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})$ and $L\psi_{f(i)} = L\psi'_i$, $L\psi'_i$ belongs to $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda))$.

The other direction follows. ($\supseteq$) Let $L\psi'_i$ be a member of $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda))$. This is possible only if $L\psi_{f(i)}$ is a member of $\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda)$. Thus $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \psi_{f(i)}$ holds. The set of quasi-literals $\Lambda \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda)$ is consistent by construction. Furthermore, $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\psi_{f(i)}\})$ is a subset of $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})$ because $L\psi_{f(i)}$ belongs to $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})$. This implies that the reduction $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\psi_{f(i)}\})}(\mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda))$ is a subset of the reduction $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda))$. Moreover, $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\psi_{f(i)}\})}(\Lambda)$ is a subset of Λ'. Thus $\Sigma \cup \Lambda' \cup \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda)) \models_{\mathrm{ae}} \psi_{f(i)}$ follows by Lemma 2.18 and the monotonicity of $\models_{\mathrm{ae}}$. Recall that $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \Lambda))$ equals to $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i-1)}(\Sigma \cup \Lambda))$ and so $\Sigma \cup \Lambda' \cup \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i-1)}(\Sigma \cup \Lambda)) \models_{\mathrm{ae}} \psi_{f(i)}$ follows. Consequently, we obtain $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon'}_{i-1}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \psi'_i$ by the inductive hypothesis and the equality $\psi'_i = \psi_{f(i)}$. Thus $L\psi'_i$ belongs to $\mathrm{B}^{\varepsilon'}_i(\Sigma \cup \Lambda')$.

So we have shown that $L\psi'_i$ belongs to $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda))$ if and only if $L\psi'_i \in \mathrm{B}^{\varepsilon'}_i(\Sigma \cup \Lambda')$. Together with the inductive hypothesis this implies that $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda))$ equals to $\mathrm{B}^{\varepsilon'}_i(\Sigma \cup \Lambda')$, because in enumeration based introspection either $L\psi'_i$ or $\neg L\psi'_i$ is added to the set of quasi-literals being constructed. Consequently, we have proven (i) for all finite values of i.

Let us then show that

$$\text{(ii)} \qquad \mathrm{B}^{\varepsilon'}(\Sigma \cup \Lambda') = \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda)).$$

Consider first the case when ε' is finite. The difference $\mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda) - \mathrm{B}^{\varepsilon}_{f(|\varepsilon'|)}(\Sigma \cup \Lambda)$ is empty, since the enumeration ε does not contain sentences ψ such that $L\psi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})$ after the sentence $\psi_{f(|\varepsilon'|)}$. Thus $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda))$ equals to $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(|\varepsilon'|)}(\Sigma \cup \Lambda))$. By (i) this equals to $\mathrm{B}^{\varepsilon'}_{|\varepsilon'|}(\Sigma \cup \Lambda')$, i.e. the set $\mathrm{B}^{\varepsilon'}(\Sigma \cup \Lambda')$ by Definition 2.3. If ε' is infinite, $\mathrm{B}^{\varepsilon'}(\Sigma \cup \Lambda') = \bigcup_{i=0}^{\infty} \mathrm{B}^{\varepsilon'}_i(\Sigma \cup \Lambda')$ by Definition 2.3. By (i) this equals to $\bigcup_{i=0}^{\infty} \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda))$. This equals to $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\bigcup_{i=0}^{\infty} \mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \Lambda))$ by the *continuity* [19] of the reduction operator. We may replace $f(i)$ by i in this expression, because the reduction with respect $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})$ is present. Thus we get an equal expression $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\bigcup_{i=0}^{\infty} \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda))$. This equals to $\Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda))$ by Definition 2.3.

The proof of the lemma follows. Let ε be an enumeration of the enumeration base Γ such that $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \phi$ holds. The set $\Lambda \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda)$ is consistent as Γ is disjoint from $\{\psi \mid L\psi \in \mathrm{qatom}(\Lambda)\}$. Thus

$$\Sigma \cup \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\Lambda \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda)) \models_{\mathrm{ae}} \phi$$

follows by Lemma 2.18. By (ii) $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{B}^{\varepsilon}(\Sigma\cup\Lambda))$ equals to $\mathrm{B}^{\varepsilon'}(\Sigma\cup\Lambda')$ where ε' is an enumeration of a *subset* of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma\cup\{\phi\}) - \mathrm{qatom}(\Lambda')\}.$$

It follows that $\Sigma\cup\Lambda'\cup\mathrm{B}^{\varepsilon'}(\Sigma\cup\Lambda') \models_{\mathrm{ae}} \phi$. By the monotonicity of $\models_{\mathrm{ae}}$, this continues to hold even though we extend ε' to a full enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma\cup\{\phi\}) - \mathrm{qatom}(\Lambda')\}$. $\square$

Given a set of quasi-literals Λ, we say that an enumeration ε *generates* Λ from Σ, if it is the case that $\Lambda = \mathrm{B}^{\varepsilon}(\Sigma)$. Lemmas 2.20 [15, Lemma 3.5] and 2.21 [15, Lemma 3.6] provide generating enumerations for N-grounded full sets and subbeliefs based on them. Lemma 2.22 [15, Lemma 3.7] provides us the condition on which the difference of a set of quasi-literals having a generating enumeration and its arbitrary subset can be generated by means of an enumeration. Lemma 2.23 shows that the subbeliefs $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ are contained in the cautious expansion corresponding to Λ. This lemma is a corollary to [15, Lemma 2.21] and [15, Theorem 2.22]. We omit the proofs of Lemmas 2.20 - 2.22 because they are presented in [15]. In general, the relation $=\!\!|_{\mathrm{ae}}$ is nonmonotonic. However, there are situations described by Lemma 2.24 where the irrefutability relation $=\!\!|_{\mathrm{ae}}$ acts monotonically. This is the key property that guarantees the characterizability of the least expansion by the iterative construction.

Lemma 2.20 (Janhunen [15]) *Let Σ be a set of premises and Λ an* N*-grounded Σ-full set. Then there is an enumeration ε of $\{\psi \mid L\psi \in \mathrm{qatom}(\Lambda)\}$ such that $\mathrm{B}^{\varepsilon}(\Sigma) = \Lambda$.*

Lemma 2.21 (Janhunen [15]) *Let Σ be a set of premises and ϕ any sentence. If Λ is an* N*-grounded Σ-full set, then there is an enumeration ε of $\{\psi \mid L\psi \in \mathrm{qatom}(\mathrm{SB}_{\Sigma,\Lambda}(\phi) - \Lambda)\}$ such that $\mathrm{B}^{\varepsilon}(\Sigma\cup\Lambda) = \mathrm{SB}_{\Sigma,\Lambda}(\phi) - \Lambda$.*

Lemma 2.22 (Janhunen [15]) *Let Σ be a set of premises, Λ a set of quasi-literals such that $\neg L\chi \in \Lambda$ implies $\Sigma\cup\Lambda \not\models_{\mathrm{ae}} \chi$ and ε an enumeration of $\{\psi \mid L\psi \in \mathrm{qatom}(\Lambda)\}$ such that $\Lambda = \mathrm{B}^{\varepsilon}(\Sigma)$. Then for every subset Λ' of Λ there is an enumeration ε_d of $\{\psi \mid L\psi \in \mathrm{qatom}(\Lambda - \Lambda')\}$ such that $\Lambda - \Lambda' = \mathrm{B}^{\varepsilon_d}(\Sigma\cup\Lambda')$.*

Lemma 2.23 (Janhunen [15]) . *Let Σ be a set of premises, Λ a Σ-full set and $\langle\Delta^+,\Delta^-\rangle$ the corresponding cautious expansion. Then $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ equals to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\phi)}(L\Delta^+ \cup \neg L\Delta^-)$.*

Lemma 2.24 *Let Σ be a set of premises and Λ a consistent set of quasi-literals. If Λ' is a subset of Λ such that $\Lambda - \Lambda'$ has a generating enumeration ε_g from $\Sigma\cup\Lambda'$, then $\langle\Sigma,\Lambda'\rangle =\!\!|_{\mathrm{ae}} \chi$ implies $\langle\Sigma,\Lambda\rangle =\!\!|_{\mathrm{ae}} \chi$.*

Proof. Let $\langle \Sigma, \Lambda' \rangle \dashv_{\mathrm{ae}} \chi$ hold. Let us then assume that $\langle \Sigma, \Lambda \rangle \not\dashv_{\mathrm{ae}} \chi$. Then there is an enumeration ε of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi\}) - \mathrm{qatom}(\Lambda)\}$ such that $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \chi$. As ε_g generates the difference $\Lambda - \Lambda'$ there is an enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi\}) - \mathrm{qatom}(\Lambda')\}$, namely $\varepsilon_g \cdot \varepsilon$, such that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon_g \cdot \varepsilon}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \chi$. It follows that $\langle \Sigma, \Lambda' \rangle \not\dashv_{\mathrm{ae}} \chi$ which contradicts our previous assumption. Hence, $\langle \Sigma, \Lambda \rangle \dashv_{\mathrm{ae}} \chi$. □

We provide yet two auxiliary results. The first one given in Lemma 2.25 points out how we can reduce or expand the set of quasi-literals Λ in the relation $\langle \Sigma, \Lambda \rangle \dashv_{\mathrm{ae}} \phi$.

Lemma 2.25 *Let Σ be a set of premises, Λ a consistent set of quasi-literals, and ϕ any sentence. Then $\langle \Sigma, \Lambda \rangle \dashv_{\mathrm{ae}} \phi$ if and only if $\langle \Sigma, \Lambda' \rangle \dashv_{\mathrm{ae}} \phi$ where $\Lambda' = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda)$.*

Proof. Let ε be any enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda)\}$. We show by induction on i that

$$\mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda) = \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda').$$

For the base case, the claim holds when $i = 0$, as both sets are empty. Induction step follows. (i) Let $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda)$. Then $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \psi_i$ holds. The set of quasi-literals $\Lambda \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda)$ is consistent by construction. By Lemma 2.18 there is a set of quasi-literals

$$\begin{array}{rcll}
\Lambda'' & = & \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\psi_i\})}(\Lambda \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda)) & [\text{Lemma 2.18}] \\
& \subseteq & \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda)) & \left[{\scriptstyle L\psi_i \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}\right] \\
& = & \Lambda' \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda) & \left[{\scriptstyle \mathrm{qatom}(\mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda)) \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}\right] \\
& = & \Lambda' \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda') & [\text{The inductive hypothesis}]
\end{array}$$

such that $\Sigma \cup \Lambda'' \models_{\mathrm{ae}} \psi_i$ holds. Thus $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \psi_i$ follows by the monotonicity of $\models_{ae}$ and $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda')$. (ii) Let $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda')$. Then it follows that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \psi_i$. As $\Lambda' \subseteq \Lambda$ and $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda')$ equals to $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda)$ by the inductive hypothesis, the monotonicity of $\models_{ae}$ yields $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \psi_i$ and $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda)$. By the cases (i) and (ii) it holds that $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda)$ if and only if $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda')$. In the enumeration based introspection of ψ_i either $L\psi_i$ or $\neg L\psi_i$ is added in both sets. Together with the inductive hypothesis, this implies that $\mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda) = \mathrm{B}^{\varepsilon}_i(\Sigma \cup \Lambda')$.

($\Rightarrow$) Let us assume that $\langle \Sigma, \Lambda' \rangle \not\dashv_{\mathrm{ae}} \phi$. Then there is an enumeration ε of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda')\}$ such that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \phi$. Because $\Lambda' \subseteq \Lambda$ and $\mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda')$ equals to $\mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda)$ as proven above, $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \phi$ follows by the monotonicity of $\models_{ae}$. Since $\Lambda' = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\Lambda)$, the enumeration ε is also an enumeration of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda)\}$$

implying that $\langle\Sigma,\Lambda\rangle \not\dashv_{\rm ae} \phi$ holds.

($\Leftarrow$) Conversely, assume that $\langle\Sigma,\Lambda\rangle \not\dashv_{\rm ae} \phi$. Then there is an enumeration ε of $\{\psi \mid L\psi \in {\rm Sf}^{\rm L}(\Sigma\cup\{\phi\}) - {\rm qatom}(\Lambda)\}$ such that $\Sigma\cup\Lambda\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda) \models_{\rm ae} \phi$. Earlier we proved by induction that ${\rm B}^\varepsilon(\Sigma\cup\Lambda)$ equals to ${\rm B}^\varepsilon(\Sigma\cup\Lambda')$ so that $\Sigma\cup\Lambda\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda') \models_{\rm ae} \phi$ holds. The set of quasi-literals $\Lambda\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda')$ is consistent because ε contains no sentence χ such that $L\chi \in {\rm qatom}(\Lambda)$. Thus by Lemma 2.18 there is a set of quasi-literals $\Lambda'' = \Pi_{{\rm Sf}^{\rm L}(\Sigma\cup\{\phi\})}(\Lambda\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda')) \subseteq \Lambda'\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda')$ such that $\Sigma\cup\Lambda'' \models_{\rm ae} \phi$ holds. Then $\Sigma\cup\Lambda'\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda') \models_{\rm ae} \phi$ follows by the monotonicity of $\models_{ae}$. As $\Lambda' = \Pi_{{\rm Sf}^{\rm L}(\Sigma\cup\{\phi\})}(\Lambda)$ the enumeration ε is also an enumeration of $\{\psi \mid L\psi \in {\rm Sf}^{\rm L}(\Sigma\cup\{\phi\}) - {\rm qatom}(\Lambda')\}$ and $\langle\Sigma,\Lambda'\rangle \not\dashv_{\rm ae} \phi$ follows. □

Lemma 2.26 provides us means to switch between the irrefutability relations $\dashv_{\rm ae}$ and $\dashv^{\rm Q}_{\rm ae}$. This result is necessary because we will use some results given in [15] that are presented using the relation $\dashv^{\rm Q}_{\rm ae}$.

Lemma 2.26 *Let Σ be a set of premises and Q a set of quasi-atoms such that* ${\rm Sf}^{\rm L}(\Sigma) \subseteq Q$ *and* ${\rm Sf}^{\rm L}(Q) \subseteq Q$. *Let Λ be a consistent set of quasi-literals and ϕ any sentence. If* ${\rm Sf}^{\rm L}(\phi) \subseteq Q$, *then $\langle\Sigma,\Lambda\rangle \dashv^{\rm Q}_{\rm ae} \phi$ if and only if $\langle\Sigma,\Lambda\rangle \dashv_{\rm ae} \phi$.*

Proof. ($\Rightarrow$) Assume that $\langle\Sigma,\Lambda\rangle \not\dashv_{\rm ae} \phi$. Then there is an enumeration ε of $\{\psi \mid L\psi \in {\rm Sf}^{\rm L}(\Sigma\cup\{\phi\}) - {\rm qatom}(\Lambda)\}$ such that $\Sigma\cup\Lambda\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda) \models_{\rm ae} \phi$ holds. Because ${\rm Sf}^{\rm L}(\Sigma\cup\{\phi\}) \subseteq Q$, the enumeration ε can be extended to an enumeration ε' of $\{\psi \mid L\psi \in Q - {\rm qatom}(\Lambda)\}$ by concatenating the elements of

$$\{\psi \mid L\psi \in (Q - {\rm Sf}^{\rm L}(\Sigma\cup\{\phi\})) - {\rm qatom}(\Lambda)\}$$

to ε. Because ε prefixes ε', ${\rm B}^\varepsilon(\Sigma\cup\Lambda) \subseteq {\rm B}^{\varepsilon'}(\Sigma\cup\Lambda)$ and $\Sigma\cup\Lambda\cup{\rm B}^{\varepsilon'}(\Sigma\cup\Lambda) \models_{\rm ae} \phi$ follows by the monotonicity of $\models_{ae}$. Thus $\langle\Sigma,\Lambda\rangle \not\dashv^{\rm Q}_{\rm ae} \phi$. ($\Leftarrow$) Assume then that $\langle\Sigma,\Lambda\rangle \not\dashv^{\rm Q}_{\rm ae} \phi$, i.e. there is an enumeration ε of $\{\psi \mid L\psi \in Q - {\rm qatom}(\Lambda)\}$ such that $\Sigma\cup\Lambda\cup{\rm B}^\varepsilon(\Sigma\cup\Lambda) \models \phi$. This implies $\Sigma\cup\Lambda'\cup{\rm B}^{\varepsilon'}(\Sigma\cup\Lambda') \models \phi$ by Lemma 2.19. Here $\Lambda' = \Pi_{{\rm Sf}^{\rm L}(\Sigma\cup\{\phi\})}(\Lambda)$ and ε' is an enumeration of $\{\psi \mid L\psi \in {\rm Sf}^{\rm L}(\Sigma\cup\{\phi\}) - {\rm qatom}(\Lambda')\}$. Thus $\langle\Sigma,\Lambda'\rangle \not\dashv_{\rm ae} \phi$ holds. By Lemma 2.25 we obtain $\langle\Sigma,\Lambda\rangle \not\dashv_{\rm ae} \phi$. □

3 Provability in Cautious Autoepistemic Logic

We introduced earlier a notion of groundedness for cautious expansions in terms of the logic **N**. In [14] Gottlob has provided L-proofs of Definition 3.1 as an alternative characterization of provability in **N**. The idea is that each step of an L-proof corresponds to an application of the necessitation rule.

Definition 3.1 (Gottlob [14]) *Let Σ be a set of premises and ϕ a sentence. Then $\Sigma \vdash_{\mathbf{N}} \phi$ if and only if there exists a finite sequence $L\phi_1, ..., L\phi_n$ such that for $1 \leq i \leq n$*

$$\Sigma \cup \{L\phi_j \,|\, 1 \leq j < i\} \models_{\text{ae}} \phi_i$$

and $\Sigma \cup \{L\phi_1, ..., L\phi_n\} \models_{\text{ae}} \phi$. Such a sequence is called an L-proof of ϕ from Σ.

In this section we show how L-proofs are generalized to apply to cautious autoepistemic reasoning. It turns out that the least expansion can be given a characterization in terms of *generalized L-proofs*. Therefore generalized L-proofs provide a tool for analyzing the least expansion. Moreover, they are useful for the development of algorithms for cautious autoepistemic reasoning.

3.1 Generalized L-proofs

Relations $\vdash_{\mathbf{C}}$ and $\dashv_{\mathbf{C}}$ are introduced in Definition 3.2. They are together needed to define the notion of generalized L-proofs (or GL-proofs for short). The role of these relations is to define which sentences ϕ the agent believes ($\Sigma \vdash_{\mathbf{C}} \phi$) and disbelieves ($\Sigma \dashv_{\mathbf{C}} \phi$). Recall that in cautious autoepistemic reasoning it is quite possible that there are sentences neither believed or disbelieved by the cautious autoepistemic agent. This gives cautious autoepistemic logic a three-valued character. Notice that GL-proofs contain both negative and positive quasi-literals in contrast to L-proofs that consist of positive quasi-literals only. Furthermore, we can see GL-proofs as a step-by-step reasoning process. At each step a GL-proof is extended by either a positive or negative quasi-literal corresponding to one introspection decision.

Definition 3.2 *Let Σ be a set of premises and ϕ a sentence. Then $\Sigma \vdash_{\mathbf{C}} \phi$ ($\Sigma \dashv_{\mathbf{C}} \phi$) if and only if there exists a finite sequence of quasi-literals $\phi_1, ..., \phi_n$ such that for $1 \leq i \leq n$*

1. *if ϕ_i is a positive quasi-literal $L\chi$, then $\Sigma \cup \{\phi_j \,|\, 1 \leq j \leq i-1\} \models_{\text{ae}} \chi$*
2. *if ϕ_i is a negative quasi-literal $\neg L\chi$, then $\langle \Sigma, \{\phi_j \,|\, 1 \leq j \leq i-1\}\rangle \mathrel{=\!\!|}_{\text{ae}} \chi$ and*
3. *$\Sigma \cup \{\phi_1, ..., \phi_n\} \models_{\text{ae}} \phi$ ($\langle \Sigma, \{\phi_1, ..., \phi_n\}\rangle \mathrel{=\!\!|}_{\text{ae}} \phi$).*

Such a sequence is called a GL-proof (GL-irrefutation) of ϕ from Σ. The length of the GL-proof (GL-irrefutation) is n.

Recall that the irrefutability relation $\langle\Sigma, \Lambda\rangle \dashv_{\mathrm{ae}} \phi$ of Definition 3.2 uses the enumeration base $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\{\phi_1, ..., \phi_{i-1}\})\}$ for $1 \leq i \leq n$. This is how the refutation procedure is localized to every member of the GL-proof (GL-irrefutation) in turn. Notice also that the base sentences of the GL-proof (GL-irrefutation) are systematically removed from the enumeration base. An example of a GL-proof is given below.

Example 3.3 Let Σ be the set of premises $\{\neg Lp \rightarrow q, \neg Lq \rightarrow p, \neg Lr \rightarrow p, \neg Lq \rightarrow s\}$. Initially the enumeration base consists of the sentences $\{p, q, r\}$ if we consider refutability of p, q, r and s. None of these sentences are logical consequences of Σ. The sentences p and s are refutable by the enumeration $\varepsilon_1 = q, p, r$. The sentence q is refutable by the enumeration $\varepsilon_2 = p, q, r$. However, r is irrefutable from $\langle\Sigma, \emptyset\rangle$. Thus we derive $\neg Lr$ and consequently r is removed from the enumeration base. Because p is a logical consequence of $\Sigma \cup \{\neg Lr\}$, we get Lp by the necessitation rule. Now the enumeration base has shrunk to $\{q\}$ if we consider refutability of q. The enumeration $\varepsilon_3 = q$ does not refute q. Therefore we add $\neg Lq$ to our GL-proof. Now the sentence s is a logical consequence of $\Sigma \cup \{\neg Lr, Lp, \neg Lq\}$. Thus s has a GL-proof $\neg Lr, Lp, \neg Lq$ from Σ. Notice the critical role of $\neg Lr$ in the GL-proof. When it is included into the GL-proof q becomes irrefutable from $\langle\Sigma, \{\neg Lr\}\rangle$ though it is refutable from $\langle\Sigma, \emptyset\rangle$.

3.2 Conditions on Enumerations

We have defined irrefutability relations in terms of enumeration based introspection. Niemelä calls an enumeration $\varepsilon = \psi_1, ..., \psi_n$ Σ-*acceptable* if there is no i and ϕ such that $\neg L\phi \in \mathrm{B}_i^{\varepsilon}(\Sigma)$ but $\Sigma \cup \mathrm{B}_i^{\varepsilon}(\Sigma) \models \phi$. Niemelä has also proved that N-grounded stable expansions of a set of premises Σ have Σ-acceptable generating enumerations from the premises [25]. A similar acceptability condition can be formulated for a set of quasi-literals Λ: $\neg L\chi \in \Lambda$ implies $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} \chi$ (this condition appears already in Lemma 2.22). Thus the acceptability condition on an enumeration is actually an acceptability condition on the set of quasi-literals generated by the enumeration. The acceptability condition can be seen as a consistency condition on a set of quasi-literals Λ: if a negative belief $\neg L\chi$ is included in Λ the sentence χ should not become a logical consequence of $\Sigma \cup \Lambda$. Let us introduce one further requirement suggested by the enumerations associated with GL-proofs: a Σ-*irrefutable* enumeration $\varepsilon = \psi_1, ..., \psi_n$ satisfies the requirement

$$\neg L\psi_i \in \mathrm{B}_i^{\varepsilon}(\Sigma) \text{ implies } \langle\Sigma, \mathrm{B}_{i-1}^{\varepsilon}(\Sigma)\rangle \dashv_{\mathrm{ae}} \psi_i$$

for every $1 \leq i \leq n$. This condition is stronger than Σ-acceptability. In Lemma 3.4 it is proven that Σ-irrefutable enumerations are also Σ-acceptable. However, the converse of the result does not hold. For instance given the set of premises $\Sigma = \{\neg Lp \rightarrow q, \neg Lq \rightarrow p\}$, the enumeration $\varepsilon = p, q$ is Σ-acceptable but not Σ-irrefutable because the enumeration $\varepsilon' = q, p$ refutes p and $\langle \Sigma, \mathrm{B}_0^{\varepsilon}(\Sigma)\rangle \not\dashv_{\mathrm{ae}} p$ holds.

Lemma 3.4 *Let Σ be a set of premises and ε a finite Σ-irrefutable enumeration of sentences. Then ε is Σ-acceptable as well.*

Proof. Assume that $\varepsilon = \psi_1, ..., \psi_n$ is not Σ-acceptable. Then there is a value i and a formula ϕ such that $\neg L\phi \in \mathrm{B}_i^{\varepsilon}(\Sigma)$ but $\Sigma \cup \mathrm{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$. As $\neg L\phi$ belongs to $\mathrm{B}_i^{\varepsilon}(\Sigma)$ there is a value $j \leq i$ such that $\phi = \psi_j$ and $\Sigma \cup \mathrm{B}_{j-1}^{\varepsilon}(\Sigma) \not\models_{\mathrm{ae}} \psi_j$. The Σ-irrefutability of ε implies $\langle \Sigma, \mathrm{B}_{j-1}^{\varepsilon}(\Sigma)\rangle \dashv_{\mathrm{ae}} \psi_j$. On the other hand, $\Sigma \cup \mathrm{B}_i^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$ implies $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j$ as $\models_{\mathrm{ae}}$ is monotonic. Let ε_j be the enumeration $\psi_j, ..., \psi_n$ and Λ be the set of quasi-literals $\mathrm{B}_{j-1}^{\varepsilon}(\Sigma)$. Now $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \psi_j$ can be rewritten as $\Sigma \cup \Lambda \cup \mathrm{B}^{\varepsilon_j}(\Sigma \cup \Lambda) \models_{\mathrm{ae}} \psi_j$. Let $\Lambda' = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\psi_j\})}(\Lambda)$. By Lemma 2.19 we obtain $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon'}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \psi_j$ where ε' is an enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\psi_j\}) - \mathrm{qatom}(\Lambda')\}$. It follows that $\langle \Sigma, \Lambda'\rangle \not\dashv_{\mathrm{ae}} \psi_j$. By Lemma 2.25 we get $\langle \Sigma, \Lambda\rangle \not\dashv_{\mathrm{ae}} \psi_j$ or equivalently $\langle \Sigma, \mathrm{B}_{j-1}^{\varepsilon}(\Sigma)\rangle \not\dashv_{\mathrm{ae}} \psi_j$. But this is a contradiction. Therefore ε is Σ-acceptable. $\square$

The connection of irrefutable enumerations to GL-proofs is straightforward: the set of quasi-literals generated by a Σ-irrefutable enumeration can be used as a GL-proof or a GL-irrefutation from Σ. However, GL-proofs and -irrefutations are more general because the same sentence may appear several times in a proof or refutation. This is not the case in enumeration based introspection where each sentence is introspected only once. However, the quasi-literals generated by irrefutable enumerations and those included in the GL-proofs and -irrefutations are essentially the same as demonstrated by Lemma 3.5. Moreover, GL-proofs are consistent sets of quasi-literals. In Lemma 3.6, it is shown that the least fixed point $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ of the operator $\mathrm{FCAE}_{\Sigma,Q}$ can be generated by means of a Σ-irrefutable enumeration.

Lemma 3.5 *Let Σ be a set of premises and $\phi_1, ..., \phi_n$ a GL-proof or a GL-irrefutation from Σ. Then the set of quasi-literals $\{\phi_1, ..., \phi_n\}$ is consistent and it has a Σ-irrefutable generating enumeration from Σ.*

Proof. We prove by induction on i that $\{\phi_1, ..., \phi_i\}$ is consistent and it has a Σ-irrefutable generating enumeration ε_i from Σ. The base case $i = 0$ is clear as the respective set of quasi-literals is empty.

Induction step follows. By the inductive hypothesis $\{\phi_1, ..., \phi_{i-1}\}$ has a Σ-irrefutable generating enumeration ε_{i-1} from Σ. Then consider the i^{th} member ϕ_i of the GL-proof. If ϕ_i belongs to $\{\phi_1, ..., \phi_{i-1}\}$, ε_{i-1} can be chosen as ε_i. Otherwise we have two cases. (i) If $\phi_i = L\chi$, $\Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \models_{\text{ae}} \chi$ follows. This implies also that $\Sigma \cup \mathrm{B}^{\varepsilon_{i-1}}(\Sigma) \models_{\text{ae}} \chi$ because ε_{i-1} generates $\{\phi_1, ..., \phi_{i-1}\}$ from Σ. Thus ε_i can be constructed by adding χ at the end of ε_{i-1}. (ii) If $\phi_i = \neg L\chi$, it holds that $\langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \;=\!\!|_{\text{ae}}\; \chi$. This implies that $\langle \Sigma, \mathrm{B}^{\varepsilon_{i-1}}(\Sigma)\rangle \;=\!\!|_{\text{ae}}\; \chi$ as ε_{i-1} generates $\{\phi_1, ..., \phi_{i-1}\}$ from Σ. Furthermore, $\Sigma \cup \mathrm{B}^{\varepsilon_{i-1}}(\Sigma) \not\models_{\text{ae}} \chi$ follows by the mutual exclusion of $\models_{\text{ae}}$ and $=\!\!|_{\text{ae}}$. Thus ε_i can be constructed by adding χ at the end of ε_{i-1}.

Consider then the consistency of the GL-proof. By the inductive hypothesis $\{\phi_1, ..., \phi_{i-1}\}$ is consistent. Let us then consider the i^{th} member ϕ_i of the GL-proof. (i) If $\phi_i = L\chi$, it holds that $\Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \models_{\text{ae}} \chi$ implying that $\langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \;\not{=\!\!|}_{\text{ae}}\; \chi$ by the mutual exclusion of the consequence relations. Let us then assume that $\neg L\chi$ belongs to $\{\phi_1, ..., \phi_{i-1}\}$. Then for some j, $1 \leq j \leq i-1$, it holds that $\langle \Sigma, \{\phi_1, ..., \phi_{j-1}\}\rangle \;=\!\!|_{\text{ae}}\; \chi$. By Lemma 3.4 the generating enumeration of $\{\phi_1, ..., \phi_{i-1}\}$ is Σ-acceptable. Therefore the difference $\{\phi_1, ..., \phi_{i-1}\} - \{\phi_1, ..., \phi_{j-1}\}$ has a generating enumeration from $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\}$. This implies $\langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \;=\!\!|_{\text{ae}}\; \chi$ by Lemma 2.24. But this is a contradiction, and therefore $\neg L\chi \notin \{\phi_1, ..., \phi_{i-1}\}$. (ii) If $\phi_i = L\chi$, it holds that $\langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \;=\!\!|_{\text{ae}}\; \chi$ implying that $\Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \not\models_{\text{ae}} \chi$ by the mutual exclusion of the two consequence relations. Assuming that $L\chi \in \{\phi_1, ..., \phi_{i-1}\}$ implies $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \models_{\text{ae}} \chi$ for some j, $1 \leq j \leq i-1$. Thus we obtain a contradiction, namely $\Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \models_{\text{ae}} \chi$, by the monotonicity of $\models_{\text{ae}}$. Hence $L\chi \notin \{\phi_1, ..., \phi_{i-1}\}$. By the cases (i) and (ii) and the inductive hypothesis $\{\phi_1, ..., \phi_i\}$ is consistent. □

Lemma 3.6 *Let Σ be a set of premises and Q a set of quasi-atoms satisfying* $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq Q$ *and* $\mathrm{Sf}^{\mathrm{L}}(Q) \subseteq Q$. *The least fixed point of* $\mathrm{FCAE}_{\Sigma,Q}$ *has a Σ-irrefutable generating enumeration.*

Proof. This proof is closely related to a proof in [15, Proposition 3.19] and it is based on the operator $\mathrm{FCAE}_{\Sigma,Q}$ of Definition 2.15. By the major results presented in [15], the iterative sequence defined by $\Gamma_0 = \emptyset$ and $\Gamma_{i+1} = \mathrm{FCAE}_{\Sigma,Q}(\Gamma_i)$ $(i \geq 0)$ is non-decreasing with respect to subset inclusion and each Γ_{i+1} has a generating enumeration ε_{i+1} which is prefixed by ε_i, i.e. the enumeration generating Γ_i. We show the Σ-irrefutability of the generating enumerations ε_i $(i \geq 0)$ by induction on i. Notice that $\neg L\chi \in \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ implies $\langle \Sigma, \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})\rangle \;=\!\!|^{\mathrm{Q}}_{\text{ae}}\; \chi$ as $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ is a fixed point of $\mathrm{FCAE}_{\Sigma,Q}$. But then $\Sigma \cup \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q}) \not\models_{\text{ae}} \chi$ holds, since the consequence relations $=\!\!|^{\mathrm{Q}}_{\text{ae}}$ and $\models_{\text{ae}}$ of Definition 2.15 are mutually exclusive. This implies that the generating enumerations ε_i of Γ_i $(i \geq 0)$ are Σ-acceptable.

Induction base: i=0. The set $\Gamma_0 = \emptyset$ is generated by the empty enumeration $\varepsilon_0 = \lambda$ that vacuously satisfies the Σ-irrefutability condition.

Inductive hypothesis: The enumerations ε_i $(i < k)$ are Σ-irrefutable.

Induction step: $i = k$. By the inductive hypothesis, the enumeration ε_{k-1} is Σ-irrefutable. Let $\varepsilon_k = \psi_1, ..., \psi_n$ be the enumeration generating Γ_k from Σ. Assume that $\neg L\psi_i$ belongs to $\Gamma_k = \mathrm{B}^{\varepsilon_k}(\Sigma)$. If $i \leq |\varepsilon_{k-1}|$, the Σ-irrefutability condition is fulfilled, as ε_{k-1} is Σ-irrefutable. Otherwise, $\langle\Sigma, \Gamma_{k-1}\rangle \dashv^{\mathrm{Q}}_{\mathrm{ae}} \phi$ holds. Assume now that $\langle\Sigma, \mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma)\rangle \not\dashv^{\mathrm{Q}}_{\mathrm{ae}} \psi_i$. Then there is an enumeration ε_r of $\{\psi \,|\, L\psi \in Q - \mathrm{qatom}(\mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma))\}$ such that

$$\Sigma \cup \mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma) \cup \mathrm{B}^{\varepsilon_r}(\Sigma \cup \mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma)) \models_{\mathrm{ae}} \psi_i$$

holds. As $\Gamma_{k-1} \subseteq \mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma)$ and the enumeration ε_k is Σ-acceptable, there is an enumeration ε_d that generates the difference $\mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma) - \Gamma_{k-1}$ from $\Sigma \cup \Gamma_{k-1}$ by Lemma 2.22. Thus we obtain $\Sigma \cup \Gamma_{k-1} \cup \mathrm{B}^{\varepsilon_d \cdot \varepsilon_r}(\Sigma \cup \Gamma_{k-1}) \models_{\mathrm{ae}} \psi_i$. The concatenation of enumerations $\varepsilon_d \cdot \varepsilon_r$ is an enumeration of

$$\{\psi \,|\, L\psi \in Q - \mathrm{qatom}(\Gamma_{k-1})\}.$$

Thus $\varepsilon_d \cdot \varepsilon_r$ is a refuting enumeration for ψ_i from $\Sigma \cup \Gamma_{k-1}$, contradicting the fact that $\langle\Sigma, \Gamma_{k-1}\rangle \dashv^{\mathrm{Q}}_{\mathrm{ae}} \psi_i$. Hence, $\langle\Sigma, \mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma)\rangle \dashv^{\mathrm{Q}}_{\mathrm{ae}} \psi_i$. Recall that ε_k is an enumeration of a subset of $\{\psi \,|\, L\psi \in Q\}$. So $L\psi_i$ belongs to Q. Because $\mathrm{Sf}^{\,\mathrm{L}}(Q) \subseteq Q$, it follows that $\mathrm{Sf}^{\,\mathrm{L}}(\psi_i) \subseteq Q$. Thus by Lemma 2.26 it holds that $\langle\Sigma, \mathrm{B}^{\varepsilon_k}_{i-1}(\Sigma)\rangle \dashv_{\mathrm{ae}} \psi_i$ and consequently ε_k is Σ-irrefutable. □

3.3 Alternative Characterization of Cautious Autoepistemic Reasoning

Our next objective is to show that the conclusions drawn by GL-proofs and -irrefutations coincide with those of the least expansion $\langle\Delta^+, \Delta^-\rangle$. The soundness and completeness of the two relations with respect to the least expansion is considered in Propositions 3.7 and 3.8, respectively.

Proposition 3.7 *Let Σ be a set of premises and $\langle\Delta^+, \Delta^-\rangle$ the least expansion of Σ. Then (i) $\Sigma \vdash_{\mathbf{C}} \phi$ implies $\phi \in \Delta^+$ and (ii) $\Sigma \dashv_{\mathbf{C}} \phi$ implies $\phi \in \Delta^-$.*

Proof. The proposition is proved by induction on the length of (i) the GL-proof or (ii) the GL-irrefutation of ϕ from Σ. Let $\Lambda = \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$, i.e. the least Σ-full set.

Induction base: ϕ has a GL-proof (GL-irrefutation) of length 0. (i) Now $\Sigma \vdash_{\mathbf{C}} \phi$ implies $\Sigma \models_{\mathrm{ae}} \phi$. It follows that $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi) \models_{\mathrm{ae}} \phi$ by the monotonicity of $\models_{ae}$. Thus $\langle\Sigma, \Lambda\rangle \models_{\mathrm{L}} \phi$ and $\phi \in \Delta^+$ by Theorem 2.10. (ii) Let

$\Sigma \dashv_{\mathrm{C}} \phi$ hold. Then $\langle\Sigma, \emptyset\rangle \dashv_{\mathrm{ae}} \phi$ holds, because ϕ has a GL-irrefutation of length 0. Assume now that $\langle\Sigma, \Lambda\rangle \not\dashv_{\mathrm{L}} \phi$. Then there is an enumeration ε of $\{\psi \mid L\psi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi))\}$ such that $\phi \in \mathrm{EB}^{\varepsilon}(\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi))$. By Lemmas 2.20 and 2.21 the set of quasi-literals $\Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi)$ has a generating enumeration ε_g from Σ, because Λ is N-grounded. This implies there is an enumeration $\varepsilon_g \cdot \varepsilon$ of $\{\psi \mid L\psi \in \mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi\})\}$ such that $\Sigma \cup \mathrm{B}^{\varepsilon_g \cdot \varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$. Consequently, $\langle\Sigma, \emptyset\rangle \not\dashv_{\mathrm{ae}} \phi$ holds, which is a contradiction. Thus $\langle\Sigma, \Lambda\rangle \dashv_{\mathrm{L}} \phi$ follows and $\phi \in \Delta^-$.

Inductive hypothesis: The claims (i) and (ii) hold when ϕ has a GL-proof shorter than k.

Induction step: ϕ has a GL-proof (GL-irrefutation) $\phi_1, ..., \phi_k$. By the inductive hypothesis, $\{\phi_1, ..., \phi_k\} \subseteq L\Delta^+ \cup \neg L\Delta^-$, as the members of the GL-proof (GL-irrefutation) have GL-proofs and GL-irrefutations of length less than k.

(i) Now $\Sigma \vdash_{\mathrm{C}} \phi$ implies that $\Sigma \cup \{\phi_1, ..., \phi_k\} \models_{\mathrm{ae}} \phi$ where $\phi_1, ..., \phi_k$ is a GL-irrefutation of ϕ. Because $\{\phi_1, ..., \phi_k\}$ is a subset of $L\Delta^+ \cup \neg L\Delta^-$, we obtain $\Sigma \cup L\Delta^+ \cup \neg L\Delta^- \models_{\mathrm{ae}} \phi$ by the monotonicity of $\models_{ae}$. Thus $\phi \in \Delta^+$.

(ii) Let $\Sigma \dashv_{\mathrm{C}} \phi$ hold. This implies that $\langle\Sigma, \{\phi_1, ..., \phi_k\}\rangle \dashv_{\mathrm{ae}} \phi$ where $\phi_1, ..., \phi_k$ is a GL-irrefutation of ϕ from Σ. Then let $Q = \mathrm{Sf}^{\,\mathrm{L}}(\Sigma \cup \{\phi_1, ..., \phi_k, \phi\})$. By Proposition 2.16 the fixed point $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ equals to $\Pi_Q(L\Delta^+ \cup \neg L\Delta^-)$. As $\{\phi_1, ..., \phi_k\} \subseteq L\Delta^+ \cup \neg L\Delta^-$, it follows that $\{\phi_1, ..., \phi_k\} \subseteq \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$. By Lemmas 3.6 and 3.4 the least fixed point $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ has a Σ-acceptable generating enumeration. Therefore the difference $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q}) - \{\phi_1, ..., \phi_k\}$ has a generating enumeration ε_d by Lemma 2.22. Let us then assume that $\langle\Sigma, \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})\rangle \not\dashv^{\mathrm{Q}}_{\mathrm{ae}} \phi$ holds implying that there is an enumeration ε_r of $\{\psi \mid L\psi \in Q - \mathrm{qatom}(\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q}))\}$ such that $\phi \in \mathrm{EB}^{\varepsilon_r}(\Sigma \cup \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q}))$. Thus there is an enumeration $\varepsilon_d \cdot \varepsilon_r$ of $\{\psi \mid L\psi \in Q - \mathrm{qatom}(\{\phi_1, ..., \phi_k\})\}$ such that $\Sigma \cup \{\phi_1, ..., \phi_k\} \cup \mathrm{B}^{\varepsilon_d \cdot \varepsilon_r}(\Sigma \cup \{\phi_1, ..., \phi_k\}) \models_{\mathrm{ae}} \phi$ holds, implying that $\langle\Sigma, \{\phi_1, ..., \phi_k\}\rangle \not\dashv^{\mathrm{Q}}_{\mathrm{ae}} \phi$. As $\mathrm{Sf}^{\,\mathrm{L}}(\phi) \subseteq Q$, $\langle\Sigma, \{\phi_1, ..., \phi_k\}\rangle \not\dashv_{\mathrm{ae}} \phi$ follows by Lemma 2.26. But this is a contradiction and thus $\langle\Sigma, \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})\rangle \dashv^{\mathrm{Q}}_{\mathrm{ae}} \phi$ holds. By Lemma 2.26, $\langle\Sigma, \mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})\rangle \dashv_{\mathrm{ae}} \phi$. Then there is a set of quasi-literals $\Lambda' = \Pi_{\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})}(\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q}))$ such that $\langle\Sigma, \Lambda'\rangle \dashv_{\mathrm{ae}} \phi$ by Lemma 2.25. As the L-subformulae $\mathrm{Sf}^{\,\mathrm{L}}(\Sigma\cup\{\phi\})$ are contained in Q and $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ is the set $\Pi_Q(L\Delta^+ \cup \neg L\Delta^-)$ by Proposition 2.16, the set of quasi-literals Λ' equals to $\Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi)$ by Lemma 2.23. It follows that $\langle\Sigma, \Lambda\rangle \dashv_{\mathrm{L}} \phi$ and $\phi \in \Delta^-$ by Theorem 2.10. □

Proposition 3.8 *Let Σ be a set of premises and $\langle\Delta^+, \Delta^-\rangle$ the least* N*-grounded cautious expansion of Σ. Then (i) $\phi \in \Delta^+$ implies $\Sigma \vdash_{\mathrm{C}} \phi$ and (ii) $\phi \in \Delta^-$ implies $\Sigma \dashv_{\mathrm{C}} \phi$.*

Proof. (i) Let ϕ belong to Δ^+. It follows that $\langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \phi$ where $\Lambda = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$ is the least Σ-full set. Thus $\Sigma \cup \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi) \models_{\mathrm{ae}} \phi$ holds. By choosing $Q = \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$, the fixed point $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,Q})$ equals to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma\cup\{\phi\})}(L\Delta^+ \cup \neg L\Delta^-)$. By Proposition 2.23 this reduction equals to $\Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi)$. The fixed point $\mathrm{lfp}(\mathrm{FCAE}_{\Sigma,\mathrm{Sf}^{\mathrm{L}}(\Sigma\cup\{\phi\})})$ has a Σ-irrefutable generating enumeration ε by Lemma 3.6. As $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma) \models_{\mathrm{ae}} \phi$ holds, $\mathrm{B}^{\varepsilon}(\Sigma)$ forms a GL-proof (in the order determined by ε) for ϕ from Σ. Hence, $\Sigma \vdash_{\mathrm{C}} \phi$ holds. (ii) Similarly, let $\phi \in \Delta^-$ hold. Then $\langle\Sigma,\Lambda\rangle =\!|_{\mathrm{L}}\ \phi$ where $\Lambda = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$ is the least Σ-full set. Consequently, it follows that $\langle\Sigma, \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi)\rangle =\!|_{\mathrm{ae}}\ \phi$. The Σ-irrefutable enumeration ε is obtained in the same way as in the case (i). So $\langle\Sigma, \mathrm{B}^{\varepsilon}(\Sigma)\rangle =\!|_{\mathrm{ae}}\ \phi$ holds and $\mathrm{B}^{\varepsilon}(\Sigma)$ provides a GL-irrefutation for ϕ from Σ implying that $\Sigma \dashv_{\mathrm{C}} \phi$. □

By Theorem 3.9 the relations $\vdash_{\mathrm{C}}$ and $\dashv_{\mathrm{C}}$ are sound and complete with respect to the least expansion of Σ. These relations can replace the relations $\models_{\mathrm{L}}$ and $=\!|_{\mathrm{L}}$ when the membership of a sentence in the least expansion is determined. From this point of view, the new relations give an alternative characterization for the least expansion. Furthermore, GL-proofs and GL-irrefutations allow the development of step-by-step algorithms for cautious autoepistemic reasoning. This in contrast to the iterative construction (Definition 2.15). Let us yet emphasize that there is no fixed point condition present in Theorem 3.9. This is an exceptional property among autoepistemic logics.

Theorem 3.9 *Let Σ be a set of premises. Then $\langle\{\phi \,|\, \Sigma \vdash_{\mathrm{C}} \phi\}, \{\phi \,|\, \Sigma \dashv_{\mathrm{C}} \phi\}\rangle$ is the least expansion of Σ.*

Proof. Let $\langle\Delta^+,\Delta^-\rangle$ be the least expansion. By Propositions 3.7 and 3.8, (i) $\phi \in \Delta^+$ if and only if $\Sigma \vdash_{\mathrm{C}} \phi$ and (ii) $\phi \in \Delta^-$ if and only if $\Sigma \dashv_{\mathrm{C}} \phi$. This implies that $\Delta^+ = \{\phi \,|\, \Sigma \vdash_{\mathrm{C}} \phi\}$ and $\Delta^- = \{\phi \,|\, \Sigma \dashv_{\mathrm{C}} \phi\}$. □

3.4 Minimal GL-proofs

GL-proofs are also helpful when *decision procedures* for cautious autoepistemic reasoning are developed. This will be done in a later section. Here we provide two preliminary results on GL-proofs that are helpful when proving a decision procedure correct. We prove the existence of *minimal* GL-proofs whose base sentences contain only sentences from $\{\psi \,|\, L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})\}$ in Lemma 3.10. L-proofs have an analogous property [13]. Lemma 3.11 provides us with a sufficient condition to decide when a sentence ϕ is not believed (but not necessarily disbelieved) according to the least expansion of Σ. A dual result can be formulated for the derivability relation $\dashv_{\mathrm{C}}$ (notice that $\Sigma \dashv_{\mathrm{C}} \phi$ implies $\Sigma \vdash_{\mathrm{C}} \neg L\phi$).

Lemma 3.10 *Let Σ be a set of premises and ϕ a sentence such that $\Sigma \vdash_{\mathrm{C}} \phi$. Then ϕ has a GL-proof $\phi_1, ..., \phi_n$ from Σ whose base sentences are contained in $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})\}$.*

Proof. From $\Sigma \vdash_{\mathrm{C}} \phi$ it follows that ϕ has a GL-proof $\phi'_1, ..., \phi'_m$ from Σ. We prove by induction that the reduction of $\phi'_1, ..., \phi'_m$ with respect to $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$ is a GL-proof for ϕ from Σ.

Let $\phi_1, ..., \phi_n$ be the reduction of $\phi'_1, ..., \phi'_m$ with respect to $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$. Define a function $f : \{1, ..., n\} \to \{1, ..., m\}$ to give the correspondence between the two GL-proofs such that $\phi_i = \phi'_{f(i)}$. Every quasi-literal ϕ_i of the GL-proof $\phi_1, ..., \phi_n$ is either of the form $L\chi_i$ or $\neg L\chi_i$. This defines the *base sentences* $\chi_1, ..., \chi_n$ for the GL-proof $\phi_1, ..., \phi_n$. The base sentences $\chi'_1, ..., \chi'_m$ are similarly defined for the GL-proof $\phi'_1, ..., \phi'_m$. In the following, we prove that for each $0 < i \leq n$ it holds that

$$\begin{array}{lll} \text{(i)} & \Sigma \cup \{\phi'_1, ..., \phi'_{f(i)-1}\} \models_{\mathrm{ae}} \chi'_{f(i)} & \text{implies} \\ & \Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \models_{\mathrm{ae}} \chi_i & \text{and} \\ \text{(ii)} & \langle \Sigma, \{\phi'_1, ..., \phi'_{f(i)-1}\}\rangle \dashv_{\mathrm{ae}} \chi'_{f(i)} & \text{implies} \\ & \langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \dashv_{\mathrm{ae}} \chi_i. & \end{array}$$

Because of the relationship of the two GL-proofs, the set of quasi-literals $\{\phi_1, ..., \phi_{i-1}\}$ equals to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\{\phi'_1, ..., \phi'_{f(i)-1}\})$.

(i) Let $\Sigma \cup \{\phi'_1, ..., \phi'_{f(i)-1}\} \models_{\mathrm{ae}} \chi'_{f(i)}$ hold. As $\chi_i = \chi'_{f(i)}$ and the set of quasi-literals $\{\phi'_1, ..., \phi'_{f(i)-1}\}$ is consistent, it follows by Lemma 2.18 that $\Sigma \cup \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\phi'_1, ..., \phi'_{f(i)-1}\}) \models_{\mathrm{ae}} \chi_i$. Because $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})$ is contained in $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$ it follows that $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\phi'_1, ..., \phi'_{f(i)-1}\})$ is a subset of $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\{\phi'_1, ..., \phi'_{f(i)-1}\})$. By the monotonicity of $\models_{\mathrm{ae}}$, it follows that $\Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \models_{\mathrm{ae}} \chi_i$.

(ii) Let $\langle \Sigma, \{\phi'_1, ..., \phi'_{f(i)-1}\}\rangle \dashv_{\mathrm{ae}} \chi'_{f(i)}$ hold. Then $\langle \Sigma, \{\phi'_1, ..., \phi'_{f(i)-1}\}\rangle \dashv_{\mathrm{ae}} \chi_i$ holds, because $\chi_i = \chi'_{f(i)}$. We obtain $\langle \Sigma, \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\phi'_1, ..., \phi'_{f(i)-1}\})\rangle \dashv_{\mathrm{ae}} \chi_i$ by Lemma 2.25. Because $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$, the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\phi'_1, ..., \phi'_{f(i)-1}\})$ equals to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\phi_1, ..., \phi_{i-1}\})$. Therefore it holds that $\langle \Sigma, \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\phi_1, ..., \phi_{i-1}\})\rangle \dashv_{\mathrm{ae}} \chi_i$. By Lemma 2.25, this implies $\langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \dashv_{\mathrm{ae}} \chi_i$.

As $\phi'_1, ..., \phi'_m$ is a GL-proof of ϕ from Σ and the claims (i) and (ii) hold, we know that $\phi_1, ..., \phi_n$ satisfies the requirements of a GL-proof from Σ. What remains to be shown is that $\phi_1, ..., \phi_n$ is a GL-proof of ϕ from Σ. Now $\Sigma \cup \{\phi_1, ..., \phi_n\} \models_{\mathrm{ae}} \phi$ follows from $\Sigma \cup \{\phi'_1, ..., \phi'_m\} \models_{\mathrm{ae}} \phi$ by Lemma 2.18, because $\{\phi_1, ..., \phi_n\}$ is the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})}(\{\phi'_1, ..., \phi'_m\})$. □

Lemma 3.11 *Let Σ be a set of premises, ϕ any sentence and $\phi_1, ..., \phi_n$ a GL-proof from Σ such that (i) the base sentences of $\phi_1, ..., \phi_n$ are contained* *in* $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})\}$ *and (ii) every quasi-atom* $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\{\phi_1, ..., \phi_n\})$ *satisfies*

$$\Sigma \cup \{\phi_1, ..., \phi_n\} \not\models_{\mathrm{ae}} \chi \textit{ and } \langle \Sigma, \{\phi_1, ..., \phi_n\}\rangle \not\dashv_{\mathrm{ae}} \chi.$$

Then

$$\Sigma \cup \{\phi_1, ..., \phi_n\} \not\models_{\mathrm{ae}} \phi \textit{ implies } \Sigma \not\vdash_{\mathrm{C}} \phi.$$

Proof. Let $\Sigma \cup \{\phi_1, ..., \phi_n\} \not\models_{\mathrm{ae}} \phi$ hold. Assume then that $\Sigma \vdash_{\mathrm{C}} \phi$, i.e. ϕ is GL-provable from Σ. By Lemma 3.10 we may assume that a GL-proof $\phi'_1, ..., \phi'_m$ such that $\mathrm{qatom}(\{\phi'_1, ..., \phi'_m\}) \subseteq \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$ exists. Let us now prove by induction on i that $\{\phi'_1, ..., \phi'_i\}$ is contained in $\{\phi_1, ..., \phi_n\}$. The base case $i = 0$ is clear as the corresponding set of quasi-literals is empty. The induction step follows. By the inductive hypothesis, $\{\phi'_1, ..., \phi'_{i-1}\}$ is a subset of $\{\phi_1, ..., \phi_n\}$. Consider then the quasi-literal ϕ'_i. If ϕ'_i equals to $L\chi$, it holds that $\Sigma \cup \{\phi'_1, ..., \phi'_{i-1}\} \models_{\mathrm{ae}} \chi$ implying that $\Sigma \cup \{\phi_1, ..., \phi_n\} \models_{\mathrm{ae}} \chi$ by the monotonicity of $\models_{\mathrm{ae}}$. Now assuming that $L\chi \notin \{\phi_1, ..., \phi_n\}$ contradicts the property (ii) of $\phi_1, ..., \phi_n$. Thus $L\chi \in \{\phi_1, ..., \phi_n\}$. If ϕ'_i equals to $\neg L\chi$, it holds that $\langle \Sigma, \{\phi'_1, ..., \phi'_{i-1}\}\rangle \dashv_{\mathrm{ae}} \chi$. The GL-proof $\{\phi_1, ..., \phi_n\}$ is Σ-acceptable by Lemmas 3.5 and 3.4. Because $\{\phi'_1, ..., \phi'_{i-1}\}$ is a subset of $\{\phi_1, ..., \phi_n\}$, there is a generating enumeration for the difference of the two GL-proofs from $\Sigma \cup \{\phi'_1, ..., \phi'_{i-1}\}$ by Lemma 2.22. Consequently, $\langle \Sigma, \{\phi_1, ..., \phi_n\}\rangle \dashv_{\mathrm{ae}} \chi$ follows by Lemma 2.24. Again, assuming that $\neg L\chi \notin \{\phi_1, ..., \phi_n\}$ contradicts the property (ii) of $\phi_1, ..., \phi_n$. Thus $\neg L\chi \in \{\phi_1, ..., \phi_n\}$. Consequently, $\{\phi'_1, ..., \phi'_m\}$ is a subset of $\{\phi_1, ..., \phi_n\}$. Therefore $\Sigma \cup \{\phi'_1, ..., \phi'_m\} \models_{\mathrm{ae}} \phi$ implies $\Sigma \cup \{\phi_1, ..., \phi_n\} \models_{\mathrm{ae}} \phi$ by the monotonicity of $\models_{\mathrm{ae}}$. But this is a contradiction. Hence, $\Sigma \not\vdash_{\mathrm{C}} \phi$. □

3.5 Cumulativity of Cautious Autoepistemic Reasoning

Dix [5] has classified various semantics of logic programs on the basis of structural rules of nonmonotonic consequence relations which were proposed by Kraus, Lehman and Magidor [18]. Especially the rules *cautious monotonicity* and *cut* have been pointed out as rules that nonmonotonic inference should obey. They correspond to the inference rules

$$\frac{\alpha \mid\!\sim \beta, \alpha \mid\!\sim \gamma}{\alpha \wedge \beta \mid\!\sim \gamma} \text{ and } \frac{\alpha \mid\!\sim \beta, \alpha \wedge \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma},$$

respectively. Classical reasoning obeys both of these rules. By combining the rules we derive the following property: if $\alpha \mid\!\sim \beta$ then $\alpha \mid\!\sim \gamma$ if and only if $\alpha \wedge \beta \mid\!\sim \gamma$. This property is also referred to as *cumulativity* of inference. In a cumulative system drawn conclusions can be added directly to the premises without affecting the set of conclusions. We can understand sentences like β above as "lemmas" that support our forthcoming inferences. Dix provides an example implying that the original AEL does not satisfy the cumulativity property:

Example 3.12 (Dix [5]) Let Σ be the set of premises $\{\neg Lb \rightarrow a, \neg La \rightarrow b, \neg Lp \rightarrow p, La \rightarrow p\}$. It has only one stable expansion, namely the one corresponding to a Σ-full set[3] $\{La, \neg Lb, Lp\}$. The sentences a and p belongs to this expansion. However, if we add p to Σ there are two stable expansions corresponding to $\Sigma \cup \{p\}$-full sets $\{La, \neg Lb, Lp\}$ and $\{\neg La, Lb, Lp\}$. Consequently, the sentence a does not belong to the intersection of the stable expansions.

We shall investigate the cumulativity of cautious autoepistemic reasoning in terms of the derivability relation $\vdash_{\mathrm{C}}$. We are interested in the effects of inserting any sentence ϕ such that $\Sigma \vdash_{\mathrm{C}} \phi$ to the set of premises. Because the least expansion determined by $\vdash_{\mathrm{C}}$ and $\dashv_{\mathrm{C}}$ is unique for every set of premises, we need not consider the intersection of cautious expansions. As previously mentioned, the monotonic consequence relation $\models$ and its derivative $\models_{\mathrm{ae}}$ are cumulative. In other words, given a set of premises Σ and $\Sigma \models_{\mathrm{ae}} \phi$, then $\Sigma \models_{\mathrm{ae}} \psi$ if and only if $\Sigma \cup \{\phi\} \models_{\mathrm{ae}} \psi$. This follows easily from the assumption $\Sigma \models_{\mathrm{ae}} \phi$ and the following property of propositional logic: $\Sigma \cup \{\phi\} \models_{\mathrm{ae}} \psi$ if and only if $\Sigma \models_{\mathrm{ae}} \phi \rightarrow \psi$. In Lemma 3.13 we give a cumulativity condition for generating enumerations. GL-proofs have generating enumerations and satisfy the acceptability condition as it was shown in Lemma 3.5. Lemma 3.14 is a straightforward technical result but is included as it is needed in a forthcoming proof.

Lemma 3.13 *Let Σ be a set of premises and Λ a Σ-acceptable set of quasi-literals which has a generating enumeration ε from Σ. Let ϕ be any sentence such that $\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi$. Then ε generates Λ from $\Sigma \cup \{\phi\}$ and Λ is $\Sigma \cup \{\phi\}$-acceptable.*

Proof. It is shown by induction on i that $\mathrm{B}^{\varepsilon}_i(\Sigma) = \mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$. The base case is clear, as both sets are empty when $i = 0$. The induction step follows. Let $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma)$ so that $\Sigma \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$. By the inductive hypothesis $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma) = \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\})$ and $\Sigma \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \models_{\mathrm{ae}} \psi_i$ follows. Then we get $\Sigma \cup \{\phi\} \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \models_{\mathrm{ae}} \psi_i$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$. But what if $\neg L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma)$. As Λ is generated by ε from Σ, $\neg L\psi_i \in \Lambda$ holds. The Σ-acceptability of Λ yields $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} \psi_i$. As the relation $\models_{\mathrm{ae}}$ is cumulative, it follows that $\Sigma \cup \{\phi\} \cup \Lambda \not\models_{\mathrm{ae}} \psi_i$. Because ε generates Λ from Σ, the set $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma)$ is a subset of Λ. By the inductive hypothesis, $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma) = \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\})$ and $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \subseteq \Lambda$ follows. Therefore

[3] Stable expansions can also be characterized by full sets.

$\Sigma \cup \{\phi\} \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \not\models_{\mathrm{ae}} \psi_i$ holds by the monotonicity of $\models_{\mathrm{ae}}$ and $\neg L\psi_i$ belongs to $\mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$. Thus $\mathrm{B}^{\varepsilon}_i(\Sigma) = \mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$ for all $i \geq 0$. As a consequence, the enumeration ε generates Λ from $\Sigma \cup \{\phi\}$.

Let us now assume that Λ is not $\Sigma \cup \{\phi\}$-acceptable. Then there is a negative quasi-literal $\neg L\chi \in \Lambda$ such that $\Sigma \cup \{\phi\} \cup \Lambda \models_{\mathrm{ae}} \chi$. Since ϕ is a logical consequence of $\Sigma \cup \Lambda$ we get $\Sigma \cup \Lambda \models_{\mathrm{ae}} \chi$ by the cumulativity of $\models_{\mathrm{ae}}$. But this contradicts the Σ-acceptability of Λ. Consequently, Λ is $\Sigma \cup \{\phi\}$-acceptable. □

Lemma 3.14 *Let Σ be a set of premises and $\varepsilon = \psi_1, \ldots, \psi_n$ any enumeration of sentences. Let ϕ be any sentence such that $\Sigma \models_{\mathrm{ae}} \phi$. Then $\mathrm{B}^{\varepsilon}(\Sigma) = \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi\})$.*

Proof. This result follows easily from the cumulativity of $\models_{\mathrm{ae}}$. It is shown that $\mathrm{B}^{\varepsilon}_i(\Sigma) = \mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$ by induction on i. The base case is clear as both sets are empty when $i = 0$. Induction step follows. Consider the i^{th} sentence ψ_i in the enumeration. Now $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma)$ if and only if $\Sigma \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma) \models_{\mathrm{ae}} \psi_i$. This is the case if and only if $\Sigma \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \models_{\mathrm{ae}} \psi_i$ because $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma)$ equals to $\mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\})$ by the inductive hypothesis. As $\models_{\mathrm{ae}}$ is cumulative, $\Sigma \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \models_{\mathrm{ae}} \psi_i$ if and only if $\Sigma \cup \{\phi\} \cup \mathrm{B}^{\varepsilon}_{i-1}(\Sigma \cup \{\phi\}) \models_{\mathrm{ae}} \psi_i$, i.e. $L\psi_i \in \mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$. Consequently, $\mathrm{B}^{\varepsilon}_i(\Sigma)$ equals to $\mathrm{B}^{\varepsilon}_i(\Sigma \cup \{\phi\})$ for all $i \geq 0$. □

Now we have the necessary tools available to prove that the derivability relation $\vdash_{\mathbf{C}}$ satisfies the rule *cut*. By Proposition 3.15 we know that inserting a sentence ϕ such that $\Sigma \vdash_{\mathbf{C}} \phi$ to the premises Σ is sound. In other words, after the insertion no new sentence becomes GL-derivable from $\Sigma \cup \{\phi\}$. However, the completeness side does not hold in general. Some conclusions may be lost in the insertion as demonstrated in Example 3.16.

Proposition 3.15 *Let Σ be a set of premises, ϕ any sentence such that $\Sigma \vdash_{\mathbf{C}} \phi$ and ψ any sentence. Then $\Sigma \cup \{\phi\} \vdash_{\mathbf{C}} \psi$ implies $\Sigma \vdash_{\mathbf{C}} \psi$.*

Proof. Let $\phi_1, \ldots, \phi_n$ be a GL-proof for ϕ from Σ. Assume then that $\Sigma \cup \{\phi\} \vdash_{\mathbf{C}} \psi$ holds. Then there is a GL-proof $\psi_1, \ldots, \psi_m$ for ψ from $\Sigma \cup \{\phi\}$. Let $\chi_1, \ldots, \chi_m$ be the base sentences of the quasi-literals $\psi_1, \ldots, \psi_m$. We prove by induction for $i = 0, \ldots, m$ that $\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_i$ is a GL-proof from Σ. The base case $i = 0$ is clear, because $\phi_1, \ldots, \phi_n$ is a GL-proof from Σ.

Induction step follows. By the inductive hypothesis $\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_{i-1}$ is a GL-proof from Σ. (i) Let $\psi_i = L\chi_i$, i.e. $\Sigma \cup \{\phi\} \cup \{\psi_1, \ldots, \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ holds. It follows that $\Sigma \cup \{\phi\} \cup \{\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ by the monotonicity of $\models_{\mathrm{ae}}$. Now ϕ may be dropped as it is a logical consequence of $\Sigma \cup \{\phi_1, \ldots, \phi_n\}$ and $\models_{\mathrm{ae}}$ is cumulative. Thus we obtain $\Sigma \cup \{\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$.

(ii) Consider the case $\psi_i = \neg L\chi_i$. Now $\langle \Sigma \cup \{\phi\}, \{\psi_1, ..., \psi_{i-1}\}\rangle \mathrel{=\!\!|}_{\mathrm{ae}} \chi_i$ holds. Assuming that $\langle \Sigma, \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}\rangle \not\mathrel{=\!\!|}_{\mathrm{ae}} \chi_i$ holds, there is an enumeration ε of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) - \mathrm{qatom}(\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\})\}$$

such that $\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\} \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i$. Because ϕ is a logical consequence of $\Sigma \cup \{\phi_1, ..., \phi_n\}$ it can be added by the cumulativity of $\models_{\mathrm{ae}}$. Moreover $\mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\})$ equals to $\mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi\} \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\})$ by Lemma 3.14 because ϕ is a logical consequence of $\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$. Thus we obtain

$$\begin{array}{l}\Sigma \cup \{\phi\} \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\} \cup \\ \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi\} \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i.\end{array}$$

Recall that $\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}$ is a GL-proof from Σ. By Lemma 3.5 the set of quasi-literals $\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$ has a Σ-irrefutable generating enumeration ε_g. Consequently, $\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$ is Σ-acceptable by Lemma 3.4. Thus by Lemma 3.13 the enumeration ε_g generates $\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$ from $\Sigma \cup \{\phi\}$ and $\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$ is in addition $\Sigma \cup \{\phi\}$-acceptable. By Lemma 2.22 we know that $\{\phi_1, ..., \phi_n\} - \{\psi_1, ..., \psi_{i-1}\}$ has a generating enumeration ε_d from $\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_{i-1}\}$. Therefore we obtain

$$\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_{i-1}\} \cup \mathrm{B}^{\varepsilon_d \cdot \varepsilon}(\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i$$

where $\varepsilon_d \cdot \varepsilon$ is a enumeration of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) - \mathrm{qatom}(\{\psi_1, ..., \psi_{i-1}\})\}.$$

Let Λ' be the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi, \chi_i\})}(\{\psi_1, ..., \psi_{i-1}\})$. By Lemma 2.19 we know that there is an enumeration ε' of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi, \chi_i\}) - \mathrm{qatom}(\Lambda')\}$ such that $\Sigma \cup \{\phi\} \cup \Lambda' \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi\} \cup \Lambda') \models_{\mathrm{ae}} \chi_i$ holds. This implies that $\langle \Sigma \cup \{\phi\}, \Lambda'\rangle \not\mathrel{=\!\!|}_{\mathrm{ae}} \chi_i$ and we get $\langle \Sigma \cup \{\phi\}, \{\psi_1, ..., \psi_{i-1}\}\rangle \not\mathrel{=\!\!|}_{\mathrm{ae}} \chi_i$ by Lemma 2.25 as Λ' is the reduction of $\{\psi_1, ..., \psi_{i-1}\}$ with respect to $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi, \chi_j\})$. But this is a contradiction. Hence $\langle \Sigma, \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}\rangle \mathrel{=\!\!|}_{\mathrm{ae}} \chi_i$. By the cases (i) and (ii) and the inductive hypothesis, $\phi_1, ..., \phi_n, \psi_1, ..., \psi_i$ is a GL-proof from Σ.

So far we know that $\phi_1, ..., \phi_n, \psi_1, ..., \psi_m$ is a GL-proof from Σ. Now it holds that $\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_m\} \models_{\mathrm{ae}} \psi$. By the monotonicity of $\models_{\mathrm{ae}}$, we get $\Sigma \cup \{\phi\} \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_m\} \models_{\mathrm{ae}} \psi$. This implies by the cumulativity of the relation $\models_{\mathrm{ae}}$ that $\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_m\} \models_{\mathrm{ae}} \psi$. It follows that $\Sigma \vdash_{\mathrm{C}} \psi$. □

Example 3.16 Let Σ be the set of premises $\{\neg Lp \rightarrow q, \neg Lq \rightarrow p, \neg Lr \rightarrow p, \neg Lq \rightarrow s\}$. Then Lp is GL-provable from Σ. However, $\Sigma \vdash_{\mathrm{C}} s$ but $\Sigma \cup \{Lp\} \not\vdash_{\mathrm{C}} s$.

The reason for the unexpected behavior of the derivability relation $\vdash_{\mathrm{C}}$ is that p is not a logical consequence of $\Sigma \cup \{Lp\}$. When the enumerations of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{Lp\})\}$ that begin with p are used by the negative introspection $\neg Lp$ is added to the premises. Consequently, the theory becomes propositionally inconsistent. It follows that every sentence is refutable and the negative introspection becomes totally paralyzed. This might suggest that cautious autoepistemic reasoning is cumulative when the extended set of premises $\Sigma \cup \{\phi\}$ is propositionally consistent with the sets of quasi-literals generated by enumerations of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$. However, this is not the case. In the example below, the sentence $\neg Lp \rightarrow r$ is also GL-derivable from Σ. If it is added to Σ, the sets of quasi-literals generated by the enumerations of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\}$ remain propositionally consistent. But, it is still the case that $\Sigma \cup \{\neg Lp \rightarrow r\} \not\vdash_{\mathrm{C}} s$.

Our counterexample to cumulativity suggests that the GL-proof of ϕ contains a history of introspection decisions. Our aim is to show that ϕ can be inserted to the premises if the GL-proof is also included. Thus we resort to a more constrained notion of cumulativity formulated in Proposition 3.20. Some auxiliary results are needed in the proof. Proposition 3.17 shows that $\vdash_{\mathrm{C}}$ allows adding propositional consequences to the premises. This is not so simple a result that is first appears. For instance, we may add any propositional tautology that contains quasi-literals (such as $LLLp \vee \neg LLLp$). This affects the enumerations of the negative introspection, but no unexpected side-effects arise. Lemmas 3.18 and 3.19 give two important properties of GL-proofs that are needed in the proof of Proposition 3.20. First of all, GL-proofs can be concatenated to form new GL-proofs. Secondly, we can transfer GL-proofs to the premises without affecting derivability by $\vdash_{\mathrm{C}}$.

Proposition 3.17 *Let Σ be a set of premises, ϕ any sentence such that $\Sigma \models_{\mathrm{ae}} \phi$ and ψ any sentence. Then $\Sigma \vdash_{\mathrm{C}} \psi$ if and only if $\Sigma \cup \{\phi\} \vdash_{\mathrm{C}} \psi$.*

Proof. ($\Rightarrow$) Let $\psi_1, ..., \psi_n$ be a GL-proof of ψ from Σ. Let $\chi_1, ..., \chi_n$ be the base sentences of the GL-proof. We prove by induction on i that $\psi_1, ..., \psi_i$ is a GL-proof from $\Sigma \cup \{\phi\}$ as well. The only-if-direction of the proposition follows from this. The base case $i = 0$ is clear as the corresponding GL-proof is empty.

Induction step follows. By the inductive hypothesis $\psi_1, ..., \psi_{i-1}$ is a GL-proof from $\Sigma \cup \{\phi\}$. (i) Let $\Sigma \cup \{\psi_1, ..., \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ hold. Then it follows that $\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ by the cumulativity of $\models_{\mathrm{ae}}$. (ii) Consider then the case $\langle \Sigma, \{\psi_1, ..., \psi_{i-1}\} \rangle \dashv_{\mathrm{ae}} \chi_i$. Let us make an assumption that $\langle \Sigma \cup \{\phi\}, \{\psi_1, ..., \psi_{i-1}\} \rangle \not\dashv_{\mathrm{ae}} \chi_i$. Then there is an enumeration ε of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi, \chi_i\}) - \mathrm{qatom}(\{\psi_1, ..., \psi_{i-1}\})\}$$

such that $\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_{i-1}\} \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi\} \cup \{\psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i$ and we obtain $\Sigma \cup \{\psi_1, ..., \psi_{i-1}\} \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \{\psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i$ by Lemma 3.14 and cumulativity of $\models_{\mathrm{ae}}$. Let Λ' be the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\psi_1, ..., \psi_{i-1}\})$. By Lemma 2.19 there is an enumeration ε' of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) - \mathrm{qatom}(\Lambda')\}$$

such that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon'}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \chi_i$ holds. This implies that $\langle \Sigma, \Lambda' \rangle \not\models\!\!|_{\mathrm{ae}} \chi_i$. By Lemma 2.25 we get $\langle \Sigma, \{\psi_1, ..., \psi_{i-1}\} \rangle \not\models\!\!|_{\mathrm{ae}} \chi_i$ as Λ' is the reduction of $\{\psi_1, ..., \psi_{i-1}\}$ with respect to $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})$. But this is a contradiction. Hence, $\langle \Sigma \cup \{\phi\}, \{\psi_1, ..., \psi_{i-1}\} \rangle \dashv_{\mathrm{ae}} \chi_i$. By the cases (i) and (ii), the sequence $\psi_1, ..., \psi_i$ is a GL-proof from $\Sigma \cup \{\phi\}$. By this inductive proof $\psi_1, ..., \psi_n$ is a GL-proof for ψ from $\Sigma \cup \{\phi\}$.

($\Leftarrow$) This is a special case of Proposition 3.15 as $\Sigma \models_{\mathrm{ae}} \phi$ implies $\Sigma \vdash_{\mathsf{C}} \phi$. □

Lemma 3.18 *Let Σ be a set of premises, $\phi_1, ..., \phi_n$ and $\psi_1, ..., \psi_m$ two GL-proofs from Σ. Then their concatenation $\phi_1, ..., \phi_n, \psi_1, ..., \psi_m$ is also a GL-proof from Σ.*

Proof. Let $\chi_1, ..., \chi_m$ be the base sentences of the quasi-literals $\psi_1, ..., \psi_m$. We prove by induction on i that $\phi_1, ..., \phi_n, \psi_1, ..., \psi_i$ is a GL-proof from Σ. The base case $i = 0$ is clear, because $\phi_1, ..., \phi_n$ is a GL-proof from Σ.

Induction step follows. By the inductive hypothesis $\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}$ is a GL-proof from Σ. (i) Let $\psi_i = L\chi_i$ so that $\Sigma \cup \{\psi_1, ..., \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ holds. Thus we obtain $\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ by the monotonicity of $\models_{\mathrm{ae}}$. (ii) Let us then consider the case $\psi_i = \neg L\chi_i$ implying that $\langle \Sigma, \{\psi_1, ..., \psi_{i-1}\} \rangle \dashv_{\mathrm{ae}} \chi_i$ holds. Assume $\langle \Sigma, \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\} \rangle \not\models\!\!|_{\mathrm{ae}} \chi_i$ holds. Then there is an enumeration ε of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) - \mathrm{qatom}(\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\})\}$$

such that $\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\} \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i$. Recall that $\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}$ is a GL-proof from Σ. By Lemma 3.5 the set of quasi-literals $\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$ has a Σ-irrefutable generating enumeration ε_g. Consequently, the set $\{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\}$ is Σ-acceptable by Lemma 3.4. By Lemma 2.22 we know that $\{\phi_1, ..., \phi_n\} - \{\psi_1, ..., \psi_{i-1}\}$ has a generating enumeration ε_d from $\Sigma \cup \{\psi_1, ..., \psi_{i-1}\}$. It follows that $\Sigma \cup \{\psi_1, ..., \psi_{i-1}\} \cup \mathrm{B}^{\varepsilon_d \cdot \varepsilon}(\Sigma \cup \{\psi_1, ..., \psi_{i-1}\}) \models_{\mathrm{ae}} \chi_i$ where $\varepsilon_d \cdot \varepsilon$ is an enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) - \mathrm{qatom}(\{\psi_1, ..., \psi_{i-1}\})\}$. Let Λ' be the reduction $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})}(\{\psi_1, ..., \psi_{i-1}\})$. By Lemma 2.19 there is an enumeration ε' of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\}) - \mathrm{qatom}(\Lambda')\}$$

such that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \chi_i$. It follows that $\langle \Sigma, \Lambda' \rangle \not\dashv_{\mathrm{ae}} \chi_i$ This implies a contradiction $\langle \Sigma, \{\psi_1, ..., \psi_{i-1}\} \rangle \not\dashv_{\mathrm{ae}} \chi_i$ by Lemma 2.25 as Λ' is the reduction of $\{\psi_1, ..., \psi_{i-1}\}$ with respect to $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_i\})$. Therefore it holds that $\langle \Sigma, \{\phi_1, ..., \phi_n, \psi_1, ..., \psi_{i-1}\} \rangle \dashv_{\mathrm{ae}} \chi_i$. By the cases (i) and (ii) and the inductive hypothesis, $\phi_1, ..., \phi_n, \psi_1, ..., \psi_i$ is a GL-proof from Σ. The inductive proof implies that $\phi_1, ..., \phi_n, \psi_1, ..., \psi_m$ is a GL-proof from Σ. □

Lemma 3.19 *Let Σ be a set of premises and $\phi_1, ..., \phi_{m-1}, \phi_m, ..., \phi_n$ a GL-proof from Σ. Then $\phi_m, ..., \phi_n$ is a GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_{m-1}\}$.*

Proof. Let $\chi_1, ..., \chi_n$ be the base sentences of the GL-proof. We prove by induction on j ($j \geq m-1$) that $\phi_m, ..., \phi_j$ is a GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_{m-1}\}$. The base case $j = m-1$ holds, as the corresponding GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_{m-1}\}$ is empty.

The induction step follows. By the inductive hypothesis $\phi_m, ..., \phi_{j-1}$ forms a GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_{m-1}\}$. (i) In the case $\phi_j = L\chi_j$ it holds that $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \models_{\mathrm{ae}} \chi_j$. This can be rewritten as

$$\Sigma \cup \{\phi_1, ..., \phi_{m-1}\} \cup \{\phi_m, ..., \phi_{j-1}\} \models_{\mathrm{ae}} \chi_j.$$

(ii) The case $\phi_j = \neg L\chi_j$ follows. Now $\langle \Sigma, \{\phi_1, ..., \phi_{j-1}\} \rangle \dashv_{\mathrm{ae}} \chi_j$ holds. Assuming $\langle \Sigma \cup \{\phi_1, ..., \phi_{m-1}\}, \{\phi_m, ..., \phi_{j-1}\} \rangle \not\dashv_{\mathrm{ae}} \chi_j$ there is an enumeration ε of

$$B = \{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi_1, ..., \phi_{m-1}, \chi_j\}) - \mathrm{qatom}(\{\phi_m, ..., \phi_{j-1}\})\}$$

such that $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\}) \models_{\mathrm{ae}} \chi_j$ holds. Now extra care is required. It is possible that ε contains some of the base sentences $\chi_1, ..., \chi_{m-1}$. In the last part of the proof (iii) we show that ε can be reduced to an enumeration ε' of

$$B' = \{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi_j\}) - \mathrm{qatom}(\{\phi_1, ..., \phi_{j-1}\})\}$$

such that $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \cup \mathrm{B}^{\varepsilon'}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\}) \models_{\mathrm{ae}} \chi_j$ still holds. It follows that $\langle \Sigma, \{\phi_1, ..., \phi_{j-1}\} \rangle \not\dashv_{\mathrm{ae}} \chi_j$ which is a contradiction. Thus we obtain

$$\langle \Sigma \cup \{\phi_1, ..., \phi_{m-1}\}, \{\phi_m, ..., \phi_{j-1}\} \rangle \dashv_{\mathrm{ae}} \chi_j.$$

By the cases (i) and (ii) and the inductive hypothesis, $\phi_m, ..., \phi_j$ is a GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_{m-1}\}$.

(iii) The difference $B - B'$ of the two enumeration bases is exactly the set

$$\{\psi \mid L\psi \in \mathrm{qatom}(\{\phi_1, ..., \phi_{m-1}\} - \{\phi_m, ..., \phi_{j-1}\})\}.$$

Let us denote the sentences of the difference by $\{\chi'_1, ..., \chi'_k\}$. Their order is determined by the enumeration ε. The sentences $\chi'_1, ..., \chi'_k$ are the base sentences of $\phi_1, ..., \phi_{m-1}$ that appear in ε. Let $f : \{0, ..., k\} \to \{0, ..., |\varepsilon|\}$ be a function that gives the position of the i^{th} sentence χ'_i in the enumeration ε and $f(0) = 0$. We will prove by induction for $i = 0, ..., k$ that there is an enumeration ε'_i of a *subset* of B' such that

$$\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\}) = \{\phi_{g(1)}, ..., \phi_{g(i)}\} \cup \mathrm{B}^{\varepsilon'_i}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$$

where the function $g : \{1, ..., k\} \to \{1, ..., m-1\}$ gives the corresponding member of the GL-proof $\phi_1, ..., \phi_{m-1}$ for the sentences $\chi'_1, ..., \chi'_k$. The enumeration ε'_0 is the empty enumeration λ. The base case $i = 0$ holds as all the three sets in the equation are empty.

Induction step follows. By the inductive hypothesis there is an enumeration ε'_{i-1} such that $\mathrm{B}^{\varepsilon}_{f(i-1)}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$ equals to $\{\phi_{g(1)}, ..., \phi_{g(i-1)}\} \cup \mathrm{B}^{\varepsilon'_{i-1}}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. Now ε'_i can be formed by concatenating the elements of ε that are between χ'_{i-1} and χ'_i at the end of ε' (if $i = 0$ the elements before χ'_1 are concatenated). The introspection of these sentences yield exactly the same literals, because their introspection is based on equal theories by the inductive hypothesis. Therefore $\mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$ equals to $\{\phi_{g(1)}, ..., \phi_{g(i-1)}\} \cup \mathrm{B}^{\varepsilon'_i}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. It remains to be shown that the introspection of the sentence χ'_i yields $\phi_{g(i)}$.

If $\phi_{g(i)} = L\chi'_i$, then χ'_i is a logical consequence of $\Sigma \cup \{\phi_1, ..., \phi_{g(i)-1}\}$. The monotonicity of $\models_{\mathrm{ae}}$ implies that the sentence χ'_i is a logical consequence of $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. Therefore $L\chi'_i$ is included in $\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. If $\phi_{g(i)} = \neg L\chi'_i$, then $\langle \Sigma, \{\phi_1, ..., \phi_{g(i)-1}\}\rangle \mathrel{=\!\!|}_{\mathrm{ae}} \chi'_i$ holds. Let us then assume that $L\chi'_i$ is included (incorrectly) in $\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. This requires that

$$\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \cup \mathrm{B}^{\varepsilon}_{f(i)-1}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\}) \models_{\mathrm{ae}} \chi'_i$$

holds. This implies that $\Sigma \cup \{\phi_1, ..., \phi_{j-1}\} \cup \mathrm{B}^{\varepsilon'_i}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\}) \models_{\mathrm{ae}} \chi'_i$ by the construction of ε'_i. Because $\phi_1, ..., \phi_{j-1}$ is a GL-proof from Σ it has a Σ-irrefutable generating enumeration from Σ by Lemma 3.5. Furthermore, this enumeration is Σ-acceptable by Lemma 3.4. Thus $\{\phi_1, ..., \phi_{j-1}\} - \{\phi_1, ..., \phi_{g(i)-1}\}$ has a generating enumeration ε_g from $\Sigma \cup \{\phi_1, ..., \phi_{g(i)-1}\}$ by Lemma 2.22. It follows that

$$\Sigma \cup \{\phi_1, ..., \phi_{g(i)-1}\} \cup \mathrm{B}^{\varepsilon_g \cdot \varepsilon'_i}(\Sigma \cup \{\phi_1, ..., \phi_{g(i)-1}\}) \models_{\mathrm{ae}} \chi'_i.$$

Let Λ' be the reduction $\Pi_{\Sigma \cup \{\chi'_i\}}(\{\phi_1, ..., \phi_{g(i)-1}\})$. By Lemma 2.25 there is an enumeration ε'' of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi'_i\}) - \mathrm{qatom}(\Lambda')\}$$

such that $\Sigma \cup \Lambda' \cup \mathrm{B}^{\varepsilon''}(\Sigma \cup \Lambda') \models_{\mathrm{ae}} \chi'_i$ implying that $\langle \Sigma, \Lambda' \rangle \not\dashv_{\mathrm{ae}} \chi'_i$. By Lemma 2.19 we obtain $\langle \Sigma, \{\phi_1, ..., \phi_{g(i)-1}\} \rangle \not\dashv_{\mathrm{ae}} \chi'_i$ as Λ' is the reduction of $\{\phi_1, ..., \phi_{g(i)-1}\}$ with respect to $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\chi'_i\})$. Therefore we derive a contradiction from our assumption and $\neg L\chi'_i$ is included in $\mathrm{B}^{\varepsilon}_{f(i)}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. The two analyzed cases imply that our inductive claim holds for i. The enumeration ε' is constructed by adding the remaining elements of ε (that appear after χ'_k) at the end of ε'_k. Introspection of these sentences generate same quasi-literals in both enumerations as $\mathrm{B}^{\varepsilon}_{f(k)}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$ equals to $\{\phi_{g(1)}, ..., \phi_{g(k)}\} \cup \mathrm{B}^{\varepsilon'_k}(\Sigma \cup \{\phi_1, ..., \phi_{j-1}\})$. □

Proposition 3.20 *Let Σ be a set of premises, ϕ a sentence that has a GL-proof $\phi_1, ..., \phi_n$ from Σ and ψ any sentence. Then $\Sigma \cup \{\phi_1, ..., \phi_n, \phi\} \vdash_{\mathbf{C}} \psi$ if and only if $\Sigma \vdash_{\mathbf{C}} \psi$.*

Proof. ($\Rightarrow$) Notice that $\phi_1, ..., \phi_n$ is also a GL-proof of $\phi_1 \wedge ... \wedge \phi_n \wedge \phi$ from Σ. Because $\{\phi_1, ..., \phi_n, \phi\}$ and $\{\phi_1 \wedge ... \wedge \phi_n \wedge \phi\}$ are logically equivalent and their L-subformulae are exactly the same $\Sigma \cup \{\phi_1, ..., \phi_n, \phi\} \vdash_{\mathbf{C}} \psi$ implies $\Sigma \cup \{\phi_1 \wedge ... \wedge \phi_n \wedge \phi\} \vdash_{\mathbf{C}} \psi$. Then $\Sigma \vdash_{\mathbf{C}} \psi$ follows by Proposition 3.15.

($\Leftarrow$) Now $\Sigma \vdash_{\mathbf{C}} \psi$ implies that ψ has a GL-proof $\psi_1, ..., \psi_m$ from Σ. By Lemma 3.18 the concatenation of the two GL-proofs $\phi_1, ..., \phi_n, \psi_1, ..., \psi_m$ is a GL-proof from Σ. By Lemma 3.19, the sequence $\psi_1, ..., \psi_m$ is also a GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_n\}$. Thus ψ has a GL-proof from $\Sigma \cup \{\phi_1, ..., \phi_n\}$ and $\Sigma \cup \{\phi_1, ..., \phi_n\} \vdash_{\mathbf{C}} \psi$ follows. As ϕ is a logical consequence of $\Sigma \cup \{\phi_1, ..., \phi_n\}$ we obtain $\Sigma \cup \{\phi_1, ..., \phi_n, \phi\} \vdash_{\mathbf{C}} \psi$ by Proposition 3.17. □

It is evident that the derivability relation $\dashv_{\mathbf{C}}$ is cumulative in the same sense as $\vdash_{\mathbf{C}}$ (the results on GL-proofs generalize to GL-irrefutations as well). It remains open whether the derivability relations are cumulative in more restricted cases, e.g. if propositional atoms are added to the premises. The notion of cumulativity that requires inclusion of GL-proofs is close to the one used by Brewka [3]. He devised a cumulative version of default logic [28]. In this approach, a history of the applied default rules is encoded as *supports* for assertions. It is then shown that such assertions with the support included can be added to the default theory without affecting its extensions (sets of conclusions in the terminology of default logic). However, this notion of cumulativity is somewhat different from the one used by Dix.

4 On the Complexity of Cautious Autoepistemic Reasoning

Let us now turn our attention to the computational complexity of cautious autoepistemic reasoning. Each reasoning task will be formulated as a *decision problem*. Such a problem involves deciding the membership of an element in some set A of interest. For example, in cautious autoepistemic reasoning, we might want to decide whether a sentence belongs to the least expansion of a set of premises. The answer to a decision problem is either "yes" or "no" depending whether the membership test succeeds or fails. We shall identify a decision problem with the set of its "yes" instances in the rest of this section. *Turing-machines* will be used to solve decision problems and to *accept* specific sets.

4.1 Polynomial Time Hierarchy

Our aim is to classify the various reasoning tasks of CAEL with respect to the *polynomial time hierarchy* (PH) (see [2] for detailed treatment). Each level $k \geq 0$ of polynomial time hierarchy consists of problem classes Σ^{P}_k, Π^{P}_k and Δ^{P}_k. The intuition is that the higher a problem lies in PH the greater time complexity it has. The lowest level of PH (level 0) consists of the problems that can be solved by a *deterministic* Turing-machine in time polynomial with respect to size of its input, i.e. $\Sigma^{\mathrm{P}}_0 = \Pi^{\mathrm{P}}_0 = \Delta^{\mathrm{P}}_0 = \mathbf{P}$. The problems in the class **NP** (nondeterministic polynomial) are solvable by a *nondeterministic* Turing-machine in polynomial time, i.e. the "yes" instances of the problem can be accepted by a nondeterministic Turing-machine in polynomial time.

In computational complexity analysis, the *reducibility* of a problem into another is a very central concept. A frequently used reduction is *polynomial time (many-to-one) reducibility*. A set A_1 is said to be polynomial time (many-to-one) reducible to a set A_2 in the case there exists a function $f : A_1 \rightarrow A_2$ computable by a deterministic Turing-machine in polynomial time such that $a \in A_1$ if and only if $f(a) \in A_2$. One can understand such a function f as a polynomial-time transformation between the two sets. Reducibilities can be used to define the hardest problems in a class of problems in the following sense. We say that a set A is $\mathbf{C}$-*hard* (with respect to polynomial time many-to-one reducibility) if every set in $\mathbf{C}$ can be reduced in polynomial time to A. In the sequel, only many-to-one reducibility will be used to define hardness with respect to a class of sets. A set A is $\mathbf{C}$-*complete* if A belongs to $\mathbf{C}$ and A is $\mathbf{C}$-hard. For example, an NP-complete set A belongs to the class **NP** and all problems of the class **NP** are polynomial time many-to-one reducible to A.

In order to define the remaining classes of PH, *oracle Turing-machines* are introduced. An oracle machine is either deterministic or nondeterministic Turing-machine that is capable of carrying out unit time membership tests in an *oracle set*. We will use $\mathbf{P}(\mathbf{C})$ to denote the class of sets accepted by deterministic oracle machines using oracles from the class $\mathbf{C}$ in polynomial time. Similarly, the notation $\mathbf{NP}(\mathbf{C})$ is used in the nondeterministic case. Oracle sets typically correspond to some decision problem of interest. In this way, oracle machines provide us a tool to study the remaining time complexity of a problem assuming some subproblem can be solved in a unit time. This implies the possibility of identifying several sources of time complexity in one problem. Oracle machines give rise to another transformation between sets as follows. A set A_1 is polynomial time deterministic (nondeterministic) Turing-reducible to a set A_2 if A_1 accepted in polynomial time by a deterministic (nondeterministic) Turing-machine with oracle set A_2.

We may now define the rest of the hierarchy. The NP-complete sets form the class $\Sigma_1^{\mathrm{p}} = \mathbf{NP}(\mathbf{P})$. For a class of sets $\mathbf{C}$ the class **co-C** consists of the complements of the sets in $\mathbf{C}$. The class Π_1^{p} is the class **co-NP**. The class Δ_1^{p} contains problems that are polynomial time deterministic Turing-reducible to classes Σ_0^{p} (or Π_0^{p}). It follows that $\Delta_1^{\mathrm{p}} = \mathbf{P}$. The other levels are defined similarly. Given any $k > 0$, define $\Sigma_k^{\mathrm{p}} = \mathbf{NP}(\Sigma_{k-1}^{\mathrm{p}})$, $\Pi_k^{\mathrm{p}} = \mathbf{co}\text{-}\Sigma_{\mathbf{k-1}}^{\mathrm{p}}$ and $\Delta_k^{\mathrm{p}} = \mathbf{P}(\Sigma_k^{\mathrm{p}})$. The classes of PH form a hierarchy of subclasses as follows $\Sigma_{k_1}^{\mathrm{p}} \cup \Pi_{k-1}^{\mathrm{p}} \subseteq \Delta_k^{\mathrm{p}} \subseteq \Sigma_k^{\mathrm{p}} \cap \Pi_k^{\mathrm{p}}$. Every class of PH is closed with respect to polynomial time many-to-one reducibility. Each of the classes Δ_k ($k \geq 0$) is closed with respect to polynomial time deterministic Turing-reducibility.

4.2 Decision Problems in Cautious Autoepistemic Logic

In Definition 4.1 we give the relevant problems that are needed in the forthcoming analysis. The problem **CN** [26] corresponds to deciding whether ϕ is a logical consequence of Σ by means of the extended consequence relation $\models_{ae}$. Niemelä [24] has shown that **CN** is polynomial time deterministic Turing-reducible to **SAT** in the propositional case. This result holds also in cautious autoepistemic logic as propositional logic is used for monotonic inference. The problem **IR** amounts to finding out the irrefutability of a sentence in the sense of relation $\Dashv_{\mathrm{ae}}$. The problems **CNL** and **IRL** correspond to the consequence relations $\models_{\mathrm{L}}$ and $\Dashv_{\mathrm{L}}$ that determine the cautious expansion characterized by a full set. The problem **CFULL** is the problem of determining full sets for finite subsets of $\mathcal{L}_{ae}$. Because cautious expansions are a generalization of stable expansions, this problem generalizes the problem **FULL** [26] presented in the context of stable expansions. The **N**-derivable sentences are coded as pairs in **CNN**. **GND** corresponds to checking up the **N**-groundedness of a set of quasi-literals.

Definition 4.1 *Let Σ be a finite set of sentences of $\mathcal{L}_{ae}$, $\phi \in \mathcal{L}_{ae}$ any sentence and Λ a finite set of quasi-literals.*

1. **CN** *is the set of pairs $\langle \Sigma, \phi \rangle$ such that $\Sigma \models_{\mathrm{ae}} \phi$.*
2. **IR** *is the set of triples $\langle \Sigma, \Lambda, \phi \rangle$ such that $\langle \Sigma, \Lambda \rangle \dashv_{\mathrm{ae}} \phi$.*
3. **CNL** *is the set of triples $\langle \Sigma, \Lambda, \phi \rangle$ such that $\langle \Sigma, \Lambda \rangle \models_{\mathrm{L}} \phi$.*
4. **IRL** *is the set of triples $\langle \Sigma, \Lambda, \phi \rangle$ such that $\langle \Sigma, \Lambda \rangle \dashv_{\mathrm{L}} \phi$.*
5. **CFULL** *is the set of pairs $\langle \Sigma, \Lambda \rangle$ such that Λ is Σ-full.*
6. **CNN** *is the set of pairs $\langle \Sigma, \phi \rangle$ such that $\Sigma \vdash_{\mathbf{N}} \phi$.*
7. **GND** *is the set of pairs $\langle \Sigma, \Lambda \rangle$ such that Λ is $\mathbf{N}$-grounded in Σ.*

The problems **BAE**, **GBAE**, **CAE**, and **GCAE** given in Definition 4.2 are different membership tests with respect to cautious expansions of a finite set of premises Σ. The problems **BAE** and **GBAE** represent brave reasoning strategy (membership in one of the expansions) while the problems **CAE** and **GCAE** correspond to the cautious strategy (membership in every expansion). The problems **GBAE** and **GCAE** consider **N**-grounded cautious expansions only. Notice that the problem **GCAE** is of major interest in cautious autoepistemic reasoning as it is equivalent to determining whether ϕ belongs to the least expansion of Σ.

Definition 4.2 *Let Σ be a finite set of sentences of $\mathcal{L}_{ae}$, $\phi \in \mathcal{L}_{ae}$ any sentence and Λ a finite set of quasi-literals.*

8. **BAE** *is the set of pairs $\langle \Sigma, \phi \rangle$ such that ϕ belongs to some cautious expansion of Σ.*
9. **GBAE** *is the set of pairs $\langle \Sigma, \phi \rangle$ such that ϕ belongs to some $\mathbf{N}$-grounded cautious expansion of Σ.*
10. **CAE** *is the set of pairs $\langle \Sigma, \phi \rangle$ such that ϕ belongs to every cautious expansion of Σ.*
11. **GCAE** *is the set of pairs $\langle \Sigma, \phi \rangle$ such that ϕ belongs to every $\mathbf{N}$-grounded cautious expansion of Σ, i.e. ϕ belongs to the least expansion of Σ.*

4.3 Classification of Algorithms

The algorithms of this section are presented in a pascal-like programming notation. However, slight extensions are required to allow concise description of both deterministic and nondeterministic Turing-machines. The idea is, of course, that the Turing-machines described by the algorithms can be constructed, but here we rely on our intuition on how to program Turing-machines rather than present the actual Turing-machines. The nondeterministic choice is presented by the **Choose** e **From** S -statement. When this statement is executed, the variable e is assigned to an arbitrary element of the set S. The computation of the machine halts in an accepting or rejecting state when the statements **Accept** and **Reject**, respectively, are executed. Notice also that **For**-loops are assumed to execute a statement for all elements in a set (the order is not relevant) and for all elements in an enumeration (the enumeration determines the order).

The enumeration-based introspection is a very central notion in cautious autoepistemic logic. It is the basis for the irrefutability relation that defines the negative introspection capability of the agent. In Figure 1 we give an algorithm [26] for constructing the set $\mathrm{B}^{\varepsilon}(\Sigma)$ given a finite set of sentences Σ and a finite enumeration ε of sentences. Niemelä has shown that $\mathsf{Introspect}(\Sigma, \varepsilon)$ returns the set $\mathrm{B}^{\varepsilon}(\Sigma)$ by making a total of $|\varepsilon|$ queries to the oracle for the problem **CN**. A nondeterministic Turing-machine accepting $\overline{\mathbf{IR}}$ is presented in Figure 2. In Proposition 4.3 we use the machine in question to show that $\overline{\mathbf{IR}}$ is polynomial time nondeterministic Turing-reducible to **CN**. The hardness of $\overline{\mathbf{IR}}$ with respect to Σ_2^{p} is addressed in Proposition 4.4.

Proposition 4.3 *The set* $\overline{\mathbf{IR}}$ *is polynomial time nondeterministic Turing-reducible to the set* **CN**.

Proof. We show that the nondeterministic Turing-machine Refute described in Figure 2 accepts $\overline{\mathbf{IR}}$ in a time polynomial to the size of its input. For a set of quasi-atoms Λ, we use $\mathrm{Enums}(\Lambda)$ to denote the set of enumerations of $\{\psi \mid L\psi \in \Lambda\}$. The set of sentences $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Gamma)\}$ is finite in this case because Σ and Γ are and for each sentence χ the set $\mathrm{Sf}^{\mathrm{L}}(\chi)$ is finite. This computation requires a time polynomial to the size of the input (in fact linear). It follows that the machine Refute can nondeterministically choose a finite enumeration ε of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Gamma)\}$ in polynomial time (notice that the set $\mathrm{Enums}(\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Gamma))$ need not be constructed explicitly). The set $\mathrm{B}^{\varepsilon}(\Sigma \cup \Gamma)$ can be computed in polynomial time as function $\mathsf{Introspect}$ makes a total of $|\varepsilon|$ queries to the oracle for **CN**. Finally, the query to the oracle for **CN** happens in a unit time. So we can conclude that the machine Refute reaches a halting state in time polynomial to the size of the input.

Function Introspect(Σ, ε): a finite set of quasi-literals;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  EB: a finite set of quasi-literals;
  ε: a finite enumeration of sentences;
  χ: a sentence of $\mathcal{L}_{ae}$.
Begin
  EB:=$\emptyset$;
  For $\chi \in \varepsilon$ **Do**
    If $\langle \Sigma \cup \mathrm{EB}, \chi \rangle \in$ **CN Then**
      EB:=EB $\cup \{L\chi\}$
    Else
      EB:=EB $\cup \{\neg L\chi\}$;
  Introspect:=EB
End

Figure 1: Enumeration-based introspection

Machine Refute(Σ, Γ, ϕ);
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Γ,BE: finite sets of quasi-literals;
  ε: a finite enumeration of sentences;
  ϕ: a sentence of $\mathcal{L}_{ae}$;
Begin
  Choose ε **From** Enums$(\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Gamma))$;
  BE:=Introspect$(\Sigma \cup \Gamma, \varepsilon)$;
  If $\langle \Sigma \cup \Gamma \cup \mathrm{BE}, \phi \rangle \in$ **CN Then**
    Accept
  Else
    Reject
End

Figure 2: An oracle machine accepting $\overline{\mathbf{IR}}$

(i) If the machine Refute halts in an accepting state there is an enumeration ε of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Gamma)\}$$

such that $\phi \in \mathrm{EB}^{\varepsilon}(\Sigma \cup \Gamma)$, i.e. $\langle \Sigma, \Gamma \rangle \not\models_{\mathrm{ae}} \phi$. It follows that $\langle \Sigma, \Gamma, \phi \rangle$ is in $\overline{\mathbf{IR}}$. (ii) Conversely, if $\langle \Sigma, \Gamma, \phi \rangle \in \overline{\mathbf{IR}}$, $\langle \Sigma, \Gamma \rangle \not\models_{\mathrm{ae}} \phi$ holds and there is an enumeration ε of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Gamma)\}$$

such that $\Sigma \cup \Gamma \cup \mathrm{B}^{\varepsilon}(\Sigma \cup \Gamma) \models_{\mathrm{ae}} \phi$. Clearly, there is an accepting computation of Refute in which the particular ε is chosen. By the cases (i) and (ii), the machine Refute accepts the set $\overline{\mathbf{IR}}$ in polynomial time. As Refute is a nondeterministic Turing-machine, $\overline{\mathbf{IR}}$ is polynomial time nondeterministic Turing-reducible to **CN**. □

Proposition 4.4 *The set $\mathbf{QBF}_{2,\exists}$ is polynomial time many-to-one reducible to the set $\overline{\mathbf{IR}}$.*

Proof. The set $\mathbf{QBF}_{2,\exists}$ consists of *valid* quantified boolean formulae having the form $\exists x_1 ... \exists x_n \forall y_1 ... \forall y_m E(x_1, ..., x_n, y_1, ..., y_m, \top, \bot)$ where $m, n \geq 0$ and E is a *boolean expression* in *propositional variables* $x_1, ..., x_n$ and $y_1, ..., y_m$ and *propositional constants* true $\top$ and false $\bot$. The polynomial time transformation that we employ in the proof is close to the one used by Gottlob [13]. The transformation is as follows: for a quantified boolean formula $\phi = \exists x_1 ... \exists x_n \forall y_1 ... \forall y_m E(x_1, ..., x_n, y_1, ..., y_m, \top, \bot)$, we define $\mathrm{AE}_{\mathrm{QBF}}(\phi)$ as a triple

$$\begin{aligned}&\langle \{\neg L \neg p_1 \rightarrow p_1, \neg L p_1 \rightarrow \neg p_1, ..., \neg L \neg p_n \rightarrow p_n, \neg L p_n \rightarrow \neg p_n\}, \emptyset,\\ &\quad E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)\rangle\end{aligned}$$

where $p_1, ..., p_n$, $q_1, ..., q_m$ and p are distinct atomic propositions. It should be clear that this function can be computed in a time polynomial in $|\phi|$. In the following, we prove that for any quantified boolean formula ϕ, $\phi \in \mathbf{QBF}_{2,\exists}$ if and only if $\mathrm{AE}_{\mathrm{QBF}}(\phi) \in \overline{\mathbf{IR}}$.

($\Rightarrow$) Let $\phi = \exists x_1 ... \exists x_n \forall y_1 ... \forall y_m E(x_1, ..., x_n, y_1, ..., y_m, \top, \bot)$ belong to $\mathbf{QBF}_{2,\exists}$. Then there is a truth assignment v to the propositional variables $x_1, ..., x_n$ such that the boolean expression $E(v(x_1), ..., v(x_n), y_1, ..., y_m, \top, \bot)$ is true for any boolean values of $y_1, ..., y_m$. Let V_i ($0 < i \leq n$) be a conjunction of literals formed from $p_1, ..., p_i$ that determines the truth values of $p_1, ..., p_i$ in correspondence to the truth values $x_1, ..., x_i$ determined by v. It follows that $V_n \models E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ and the implication $V_n \rightarrow E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ is a tautology. Let Σ be the first element in the triple $\mathrm{AE}_{\mathrm{QBF}}(\phi)$, i.e. the set of premises in our translation.

Next we construct an enumeration of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\} = \{p_1, ..., p_n, \neg p_1, ..., \neg p_n\}$$

that refutes $E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ from Σ. Let ε_0 be the empty enumeration λ. The enumeration ε_i is constructed either by catenating p_i and $\neg p_i$ to ε_{i-1} if $v(x_i) = \bot$ or by catenating $\neg p_i$ and p_i to ε_{i-1} if $v(x_i) = \top$. Let $\varepsilon = \varepsilon_n$. Notice that neither p_i nor $\neg p_i$ is a logical consequence of $\Sigma \cup \mathrm{B}^{\varepsilon}_{2(i-1)}(\Sigma)$. Consequently, if p_i is before $\neg p_i$ in ε the enumeration based introspection yields $\neg L p_i$ and $L \neg p_i$, i.e. $\neg p_i$ becomes a logical consequence of $\Sigma \cup \mathrm{B}^{\varepsilon}_{2i}(\Sigma)$. If the two sentences appear in the opposite order, the enumeration based introspection yields $\neg L \neg p_i$ and $L p_i$, i.e. p_i is a logical consequence of $\Sigma \cup \mathrm{B}^{\varepsilon}_{2i}(\Sigma)$. Thus, $\Sigma \cup \mathrm{B}^{\varepsilon}_{2i}(\Sigma) \models_{\mathrm{ae}} V_i$. It follows that V_n is a logical consequence of $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma)$ and so is $E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$. Thus ε is a refuting enumeration for $E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ and $\langle \Sigma, \emptyset \rangle \not\models_{\mathrm{ae}} E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ follows. Hence $\mathrm{AE}_{\mathrm{QBF}}(\phi) \in \overline{\mathbf{IR}}$.

($\Leftarrow$) Let $\mathrm{AE}_{\mathrm{QBF}}(\phi) \in \overline{\mathbf{IR}}$ where

$$\phi = \exists x_1 ... \exists x_n \forall y_1 ... \forall y_m E(x_1, ..., x_n, y_1, ..., y_m, \top, \bot).$$

Then it holds that $\langle \Sigma, \emptyset \rangle \not\models_{\mathrm{ae}} E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ where Σ is the first element in the triple $\mathrm{AE}_{\mathrm{QBF}}(\phi)$ and $E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ an autoepistemic sentence formed of propositions $p_1, ..., p_n$, $q_1, ..., q_n$ and p. Let ε be the refuting enumeration in question. Clearly, ε is an enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)\} = \{p_1, ..., p_n, \neg p_1, ..., \neg p_n\}$. Depending on the relative order of the sentences p_i and $\neg p_i$ in the enumeration ε, either p_i or $\neg p_i$ is a logical consequence of $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma)$, i.e. the set of sentences $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma)$ uniquely determines the truth values of $p_1, ..., p_n$. On the contrary, the set of sentences $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma)$ does not determine the truth values of $q_1, ..., q_m$. Let V be the conjunction of the literals formed from $p_1, ..., p_n$ that determines the truth values of $p_1, ..., p_n$ in accordance with $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma)$. As $E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ logically follows from $\Sigma \cup \mathrm{B}^{\varepsilon}(\Sigma)$, $V \rightarrow E(p_1, ..., p_n, q_1, ..., q_m, p \vee \neg p, p \wedge \neg p)$ is a tautology. Thus $E(x_1, ..., x_n, y_1 ... y_m, \top, \bot)$ evaluates to true for any values of $y_1, ..., y_m$ if the variables $x_1, ..., x_n$ are assigned to the truth values corresponding to those of $p_1, ..., p_n$. Hence ϕ is a valid quantified boolean formula and it belongs to $\mathbf{QBF}_{2,\exists}$. □

Oracles that are capable of solving several problems are needed in the forthcoming analysis. To cope with queries to oracles for different problems we introduce the following notation. The set $\{\langle i, e \rangle \mid e \in S_i \text{ and } 1 \leq i \leq n\}$ is denoted by $S_1 \& ... \& S_n$. Intuitively, an oracle for $S_1 \& ... \& S_n$ can solve all the problems $S_1, ..., S_n$ in unit time. A query $e \in S_i$ made to an oracle for S_i is now represented as a query $\langle i, e \rangle \in S_1 \& ... \& S_n$ to an oracle for $S_1 \& ... \& S_n$.

In Proposition 4.6 we show that the sets **CNL** and **IRL** are polynomial time Turing-reducible to the set **CN&IR**. In Lemma 4.5 it is shown that the subbeliefs of a sentence ϕ can be constructed level by level.

Function $\mathsf{SubB}(\Sigma, \Lambda, \phi)$: set of quasi-literals;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ,SB,SB2: finite sets of quasi-literals;
  ϕ, χ: sentences of $\mathcal{L}_{ae}$;
  Test: Boolean;
Begin
  SB:=$\emptyset$;
  For $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\phi)$ **Do**
    Begin
      $\langle$Test, SB2$\rangle$:=$\mathsf{CnL}(\Sigma, \Lambda, \chi)$;
      SB:=SB $\cup$ SB2;
      If Test **Then**
        SB:=SB $\cup$ $\{L\chi\}$
      Elif $\langle 2, \langle \Sigma, \Lambda \cup \mathrm{SB2}, \chi\rangle\rangle \in$ **CN&IR Then**
        SB:=SB $\cup$ $\{\neg L\chi\}$
    End;
  SubB:=SB
End

Figure 3: Computation of subbeliefs

Function $\mathsf{CnL}(\Sigma, \Lambda, \phi)$: Boolean $\times$ set of quasi-literals;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ,SB: finite sets of quasi-literals;
  ϕ: a sentence of $\mathcal{L}_{ae}$;
Begin
  SB:=$\mathsf{SubB}(\Sigma, \Lambda, \phi)$;
  If $(\langle 1, \langle \Sigma \cup \Lambda \cup \mathrm{SB}, \phi\rangle\rangle \in$ **CN&IR**) **Then**
    CnL:= $\langle$True, SB$\rangle$
  Else
    CnL:= $\langle$False, SB$\rangle$
End

Figure 4: A decision procedure for **CNL**

Function $\mathsf{IrL}(\Sigma, \Lambda, \phi)$: Boolean;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ,SB: finite sets of quasi-literals;
  ϕ: a sentence of $\mathcal{L}_{ae}$;
Begin
  SB:=$\mathsf{SubB}(\Sigma, \Lambda, \phi)$;
  If ($\langle 2, \langle \Sigma, \Lambda \cup \mathrm{SB}, \phi\rangle\rangle \in$ **CN&IR**) **Then**
    IrL:= True
  Else
    IrL:= False
End

Figure 5: A decision procedure for **IRL**

Lemma 4.5 *Let Σ be a set of premises, Λ a consistent set of quasi-literals and ϕ any sentence. Then* $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ *equals to*

$$\{L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi) \mid \langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \chi\} \cup \{\neg L\chi \in \neg\mathrm{Sf}^{\,\mathrm{qL}}(\phi) \mid \langle\Sigma,\Lambda\rangle \mathrel{=\!\!|}_{\mathrm{L}} \chi\} \cup \bigcup_{L\psi\in\mathrm{Sf}^{\,\mathrm{qL}}(\phi)} \mathrm{SB}_{\Sigma,\Lambda}(\psi).$$

Proof. By Definition 2.8 the subbeliefs $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ consist of the set of quasi-literals

$$\{L\chi \in \mathrm{Sf}^{\,\mathrm{L}}(\phi) \mid \langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \chi\} \cup \{\neg L\chi \in \neg\mathrm{Sf}^{\,\mathrm{L}}(\phi) \mid \langle\Sigma,\Lambda\rangle \mathrel{=\!\!|}_{\mathrm{L}} \chi\}.$$

Therefore the major concern are the quasi-literals ϕ' such that $\mathrm{qatom}(\phi') \in \mathrm{Sf}^{\,\mathrm{L}}(\phi) - \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$ (recall that $\mathrm{Sf}^{\,\mathrm{qL}}(\phi) \subseteq \mathrm{Sf}^{\,\mathrm{L}}(\phi)$). Let us check the inclusion of such quasi-literals in both directions.

Let $L\chi \in \mathrm{Sf}^{\,\mathrm{L}}(\phi) - \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$ hold. (i) If $L\chi \in \mathrm{SB}_{\Sigma,\Lambda}(\phi)$, it follows that $\langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \chi$. Then there is some $L\chi' \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$ such that $L\chi \in \mathrm{Sf}^{\,\mathrm{L}}(\chi')$. This implies that $L\chi \in \mathrm{SB}_{\Sigma,\Lambda}(\chi')$ and $L\chi \in \bigcup_{L\psi\in\mathrm{Sf}^{\,\mathrm{qL}}(\phi)} \mathrm{SB}_{\Sigma,\Lambda}(\psi)$. (ii) Similarly, $\neg L\chi \in \mathrm{SB}_{\Sigma,\Lambda}(\phi)$ implies $L\chi \in \mathrm{SB}_{\Sigma,\Lambda}(\chi')$ for some $L\chi' \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$. Consequently, $\neg L\chi \in \bigcup_{L\psi\in\mathrm{Sf}^{\,\mathrm{qL}}(\phi)} \mathrm{SB}_{\Sigma,\Lambda}(\psi)$. By the cases (i) and (ii) $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ is contained in the set of quasi-literals claimed by the lemma.

(iii) Let $L\chi \in \bigcup_{L\psi\in\mathrm{Sf}^{\,\mathrm{qL}}(\phi)} \mathrm{SB}_{\Sigma,\Lambda}(\psi)$. Then for some $L\psi \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$, $L\chi \in \mathrm{Sf}^{\,\mathrm{L}}(\psi)$ and $\langle\Sigma,\Lambda\rangle \models_{\mathrm{L}} \chi$ holds. But then $L\chi \in \mathrm{Sf}^{\,\mathrm{L}}(\phi)$ and $L\chi \in \mathrm{SB}_{\Sigma,\Lambda}(\phi)$. (iv) If it is the case that $\neg L\chi \in \bigcup_{L\psi\in\mathrm{Sf}^{\,\mathrm{qL}}(\phi)} \mathrm{SB}_{\Sigma,\Lambda}(\psi)$ we get $\neg L\chi \in \mathrm{SB}_{\Sigma,\Lambda}(\phi)$ by similar arguments. Now (iii) and (iv) establish the inclusion in the other direction. Therefore the expressions for $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ are equal. $\Box$

Proposition 4.6 *The sets* **CNL** *and* **IRL** *are polynomial time Turing-reducible to the set* **CN&IR**.

Proof. The proof is based on the functions SubB, CnL and IrL presented in Figures 3, 4 and 5. The purpose of SubB is to compute the subbeliefs of a given sentence. The functions CnL and IrL serve as decision procedures for the recursive consequence relations $\models_{\mathrm{L}}$ and $=\!\mid_{\mathrm{L}}$. Notice that CnL returns a pair $\langle \mathrm{Test}, \mathrm{SB}_{\Sigma,\Lambda}(\phi)\rangle$ where Test is the answer (a boolean value) to the decision problem. This is to avoid unnecessary recomputation of subbeliefs. We let $\mathrm{BD}(\phi)$ denote the maximum nesting level of the modal operator L in a sentence ϕ. For example, $\mathrm{BD}(L(Lp \vee \neg LLr) \wedge LLp) = 3$. At first we consider the correctness of the functions SubB and CnL. We show by induction on $\mathrm{BD}(\phi)$ that (i) $\mathsf{SubB}(\Sigma, \Lambda, \phi)$ returns $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$ and (ii) $\mathsf{CnL}(\Sigma, \Lambda, \phi)$ returns $\langle \mathrm{True}, \mathrm{SB}_{\Sigma,\Lambda}(\phi)\rangle$ if and only if $\langle \Sigma, \Lambda\rangle \models_{\mathrm{L}} \phi$ holds.

Induction base: $\mathrm{BD}(\phi) = 0$. (i) In the function SubB the variable SB is assigned to $\emptyset$ and the **For**-loop is not executed because $\mathrm{Sf}^{\,\mathrm{qL}}(\phi)$ is empty. Now $\mathrm{SB}_{\Sigma,\Lambda}(\phi) = \emptyset$ which is also returned by SubB. (ii) In the function CnL, the variable SB is correctly assigned to $\emptyset$, i.e. the subbeliefs of ϕ. Then CnL makes a query to an oracle for **CN** and returns $\langle \mathrm{True}, \emptyset\rangle$ if and only if $\Sigma \cup \Lambda \models_{\mathrm{ae}} \phi$ or equivalently, $\langle \Sigma, \Lambda\rangle \models_{\mathrm{L}} \phi$ as $\mathrm{SB}_{\Sigma,\Lambda}(\phi) = \emptyset$. Therefore CnL operates correctly when $\mathrm{BD}(\phi) = 0$.

Inductive hypothesis: The claims (i) and (ii) hold when $\mathrm{BD}(\phi) < k$.

Induction step: $\mathrm{BD}(\phi) = k$. (i) As $\mathrm{BD}(\chi) < k$ holds for each $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\Sigma)$ the inductive hypothesis can be applied. Thus $L\chi$ is inserted to SB if and only if $\langle \Sigma, \Lambda\rangle \models_{\mathrm{L}} \chi$ holds. The subbeliefs returned by the call $\mathsf{CnL}(\Sigma, \Lambda, \chi)$ are stored in SB2. Instead making another recursive call $\mathsf{IrL}(\Sigma, \Lambda, \chi)$ (which would result in recomputing $\mathrm{SB}_{\Sigma,\Lambda}(\chi)$) a query to an oracle for **CN&IR** is made. Consequently, $\neg L\chi$ is inserted to SB if and only if $\langle \Sigma, \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\chi)\rangle =\!\mid_{\mathrm{ae}} \chi$, i.e. $\langle \Sigma, \Lambda\rangle =\!\mid_{\mathrm{L}} \chi$ holds. Moreover, for each $L\chi \in \mathrm{Sf}^{\,\mathrm{qL}}(\phi)$, the subbeliefs $\mathrm{SB}_{\Sigma,\Lambda}(\chi)$ are included in SB (by adding the elements of SB2 to SB in the **For**-loop). By Lemma 4.5 this implies that SB is finally assigned to $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$. Thus SuB returns correctly the subbeliefs $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$. (ii) In the function CnL, the variable SB is correctly assigned to $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$. Then a query to the oracle for **CN&IR** is made. As a result, $\mathsf{CnL}(\Sigma, \Lambda, \phi)$ returns $\langle \mathrm{True}, \mathrm{SB}_{\Sigma,\Lambda}(\phi)\rangle$ if and only if $\Sigma \cup \Lambda \models_{\mathrm{L}} \phi$ holds.

The correctness of IrL is verified next. By the preceding induction we know that the variable SB is correctly assigned to $\mathrm{SB}_{\Sigma,\Lambda}(\phi)$. Then a query to an oracle for **CN&IR** is made. The function IrL returns True if and only if $\langle \Sigma, \Lambda \cup \mathrm{SB}_{\Sigma,\Lambda}(\phi)\rangle =\!\mid_{\mathrm{ae}} \phi$ holds, or equivalently $\langle \Sigma, \Lambda\rangle =\!\mid_{\mathrm{L}} \phi$ holds.

Let us then evaluate the time required by the algorithms. First of all, the set

$\mathrm{Sf}^{\mathrm{L}}(\phi)$ is finite. Let n_ϕ be a function that gives the number of occurrences of $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\phi)$ in the sentence ϕ. Then define $n(\phi)$ as the sum

$$\sum_{L\chi \in \mathrm{Sf}^{\mathrm{L}}(\phi)} n_\phi(L\chi),$$

i.e. the total number of occurrences of the elements of $\mathrm{Sf}^{\mathrm{L}}(\phi)$ in ϕ. This sum equals to the number of L operators in the sentence ϕ. It is therefore clear that n grows linearly in the size of ϕ. Let us then prove by induction on $\mathrm{BD}(\phi)$ that at most $n(\phi)$ recursive calls to CnL (and subsequently SubB) are made. The base case is clear, because $\mathrm{BD}(\phi) = 0$ implies $n(\phi) = 0$ and $\mathrm{Sf}^{\mathrm{qL}}(\phi) = \emptyset$. So the **For**-loop of the function SubB is not executed and no recursive calls to CnL are made.

The induction step follows. The **For**-loop of the function SubB makes explicitly $|\mathrm{Sf}^{\mathrm{qL}}(\phi)|$ recursive calls to CnL. Let $L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\phi)$. Now $\mathrm{BD}(\chi)$ is less than $\mathrm{BD}(\phi)$. By the inductive hypothesis a recursive call $\mathsf{CnL}(\Sigma, \Lambda, \chi)$ involves at most $n(\chi)$ recursive calls to CnL. Therefore the total number of recursive calls is at most

$$|\mathrm{Sf}^{\mathrm{qL}}(\phi)| + \sum_{L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\phi)} (n(\chi)).$$

This is lower than $n(\phi)$ which equals to

$$\sum_{L\chi \in \mathrm{Sf}^{\mathrm{qL}}(\phi)} (n_\phi(L\chi) + n_\phi(L\chi) \times n(\chi)).$$

Therefore at most $n(\phi)$ recursive calls to CnL are made. It follows that the number of recursive calls grows linearly in the size of ϕ as $n(\phi)$ does.

Notice that the query $\langle 2, \langle \Sigma, \Lambda \cup \mathrm{SB2}, \chi \rangle \rangle \in$ **CN&IR** in the function SubB is made at most as many times as CnL is recursively called. By the same argument, the total number of executions of the **For**-loop in the function SuB is bounded by $n(\phi)$. Furthermore, each recursion instance of CnL makes an additional oracle query for **CN&IR**. This implies that when the function CnL is called, the total number of queries to an oracle for **CN&IR** is bounded by $2 \times n(\phi)$. Consequently, computing the functions SubB and CnL requires polynomial time for a deterministic oracle machine. This implies that computing the function IrL requires polynomial time for a deterministic oracle machine. Thus the reduction result is obtained. □

Recall that full sets were defined in terms of relations $\models_{\mathrm{L}}$ and $\dashv_{\mathrm{L}}$. Therefore its is natural to reduce the problem **CFULL** to the problems **CNL** and **IRL**. In proposition 4.7 it is shown that a polynomial time Turing-reduction is possible. By Theorem 2.10 we can determine the membership of any sentence

Machine CFull(Σ, Λ);
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ: a set of quasi-literals belonging to Cand(Σ);
  χ: a sentence of $\mathcal{L}_{ae}$;
Begin
  For $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ **Do**
    If ($L\chi \in \Lambda$ **and** $\langle 1, \langle \Sigma \cup \Lambda, \chi\rangle\rangle \in$ **CN&IR**) **or**
      ($\neg L\chi \in \Lambda$ **and** $\langle 2, \langle \Sigma, \Lambda, \chi\rangle\rangle \in$ **CN&IR**) **Then**
      Continue
    Elif ($\langle 1, \langle \Sigma \cup \Lambda, \chi\rangle\rangle \in$ **CN&IR or** $\langle 2, \langle \Sigma, \Lambda, \chi\rangle\rangle \in$ **CN&IR**) **Then**
      Reject;
  Accept
End

Figure 6: An oracle machine accepting **CFULL**

in a cautious expansion given the corresponding full set and the recursive consequence relation $\models_{\mathrm{L}}$. In Proposition 4.8 we show that $\overline{\textbf{CAE}}$ is polynomial time nondeterministic Turing-reducible to **CNL&CFULL**.

Proposition 4.7 *The set* **CFULL** *is polynomial time Turing-reducible to the set* **CN&IR**.

Proof. The machine CFull of Figure 6 executes its **For**-loop $|\mathrm{Sf}^{\mathrm{L}}(\Sigma)|$ number of times. In each round, at most two queries are made to an oracle for **CN&IR**. This implies that the machine CFull halts in polynomial time with respect to $|\mathrm{Sf}^{\mathrm{L}}(\Sigma)|$. Let Γ be a subset of $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ that gives the quasi-atoms of $\mathrm{Sf}^{\mathrm{L}}(\Sigma)$ that have been verified to satisfy the fullness conditions (initially $\Gamma = \emptyset$). The loop invariant is as follows: for all $L\chi \in \Gamma$ it holds that $L\chi \in \Lambda$ if and only if $\langle \Sigma, \Lambda\rangle \models_{\mathrm{L}} \chi$ and $\neg L\chi \in \Lambda$ if and only if $\langle \Sigma, \Lambda\rangle \dashv_{\mathrm{L}} \chi$. If the machine CFull halts in a rejecting state, there is a quasi-literal $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ (the one being currently processed) violating the Σ-fullness condition. If the machine accepts its input, $\Gamma = \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ and Λ is Σ-full by the loop invariant. □

Proposition 4.8 *The set* $\overline{\textbf{CAE}}$ *is polynomial time nondeterministic Turing-reducible to the set* **CNL&CFULL**.

Proof. In the algorithm Cand(Σ) denotes the candidate full sets that are consistent subsets of $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\Sigma)$. (i) If the machine Co-Cae of Figure 7 halts in an accepting state, the set of quasi-literals Λ is Σ-full and $\langle \Sigma, \Lambda\rangle \not\models_{\mathrm{L}} \phi$ holds. By Theorem 2.10 this implies that Σ has a cautious expansion to which

Machine Co-Cae(Σ, ϕ);
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ: a set of quasi-literals belonging to Cand(Σ);
  ϕ: a sentence of $\mathcal{L}_{ae}$;
Begin
  Choose Λ **From** Cand(Σ);
  If $\langle 2, \langle \Sigma, \Lambda \rangle\rangle \notin$ **CNL&CFULL Then**
    Reject;
  If $\langle 1, \langle \Sigma, \Lambda, \phi \rangle\rangle \notin$ **CNL&CFULL Then**
    Accept
  Else
    Reject;
End

Figure 7: An oracle machine accepting $\overline{\mathbf{CAE}}$

ϕ does not belong. Hence, $\langle \Sigma, \phi \rangle \in \overline{\mathbf{CAE}}$. (ii) Conversely, if $\langle \Sigma, \phi \rangle \in \overline{\mathbf{CAE}}$ there is a cautious expansion $\langle \Delta^+, \Delta^- \rangle$ of Σ such that $\phi \notin \Delta^+$. By Theorem 2.10 there is a Σ-full set $\Lambda = \Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma)}(L\Delta \cup \neg L\Delta^-)$ such that $\langle \Sigma, \Lambda \rangle \not\models_{\mathrm{L}} \phi$ holds. As Λ belongs Cand(Σ), the machine Co-Cae has an accepting computation on $\langle \Sigma, \Lambda \rangle$. By (i) and (ii) the machine Co-Cae accepts exactly the set $\overline{\mathbf{CAE}}$. As the computation requires two queries to the oracle for **CNL&CFULL** and the machine is nondeterministic, the reducibility result follows. □

It is time to evaluate the effect of the N-groundedness to the computational complexity. In Proposition 4.9 it is shown that $\overline{\mathbf{GND}}$ is polynomial time Turing-reducible to **CNN**. Consequently, the set $\overline{\mathbf{GCAE}}$ is polynomial time nondeterministic Turing-reducible to **CNL&CFULL&GND**.

Proposition 4.9 *The set* **GND** *is polynomial time Turing-reducible to the set* CNN.

Proof. The polynomial time reduction is given in Figure 8. Because Λ is finite, the set of negative quasi-literals $[\Lambda]^- = \{\neg L\chi \mid \neg L\chi \in \Lambda\}$ can be determined in a time linear to $|\Lambda|$. The **For**-loop of the algorithm is executed at most $|\mathrm{Sf}^{\mathrm{L}}(\Sigma)|$ times. Each loop execution involves (at most) two queries to an oracle for **CNN**. Hence, the reduction takes place in polynomial time. If the machine Gnd halts in a rejecting state, there is at least one $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ violating the groundedness condition. If the machine accepts Λ, it is clearly N-grounded. Conversely, if Λ is N-grounded, it has an accepting computation on Gnd. Therefore the reducibility results follows. □

Machine $\mathsf{Gnd}(\Sigma, \Lambda)$;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ: a set of quasi-literals belonging to $\mathrm{Cand}(\Sigma)$;
  χ: a sentence of $\mathcal{L}_{ae}$;
Begin
  For $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma)$ **Do**
    If $(L\chi \in \Lambda$ **and**$\langle \Sigma \cup [\Lambda]^{-}, \chi\rangle \notin$ **CNN**$)$ **or**
      $(L\chi \notin \Lambda$ **and**$\langle \Sigma \cup [\Lambda]^{-}, \chi\rangle \in$ **CNN**$)$ **Then**
      Reject;
  Accept
End

Figure 8: An oracle machine accepting **GND**

Proposition 4.10 *The set* $\overline{\mathbf{GCAE}}$ *is polynomial time nondeterministic Turing-reducible to the set* **CNL&CFULL&GND**.

Proof. A nondeterministic Turing-machine accepting $\overline{\mathbf{GCAE}}$ can be obtained by modifying the machine $\mathsf{Co\text{-}Cae}$ in Figure 7. First of all, an oracle for **CNL&CFULL&GND** is needed. An additional N-groundedness test is added: it is required that $\langle 3, \langle \Sigma, \Lambda\rangle\rangle \in$ **CNL&CFULL&GND**. Consequently, the modified machine $\mathsf{Co\text{-}GCae}$ accepts its input if and only if there is an N-grounded cautious expansion of Σ to which ϕ does not belong, i.e. $\langle \Sigma, \phi\rangle \in \overline{\mathbf{GCAE}}$. On the other hand, if $\langle \Sigma, \phi\rangle$ belongs to $\overline{\mathbf{GCAE}}$, there is an N-grounded cautious expansion $\langle \Delta^{+}, \Delta^{-}\rangle$ such that $\phi \notin \Delta^{+}$. Then there is an accepting computation of $\mathsf{Co\text{-}GCae}$ where Λ is assigned to $\Pi_{\mathrm{Sf}^{\mathrm{L}}(\Sigma)}(L\Delta^{+} \cup \neg L\Delta^{-})$. The machine $\mathsf{Co\text{-}GCae}$ accepts its input in a time polynomial to the size of its input by the proof of Proposition 4.8. $\square$

The reduction result provided by Proposition 4.10 can be improved on the basis of the iterative construction. Proposition 4.12 is sufficient to show the membership of **GCAE** in Δ_3^{p}. In Lemma 4.11 we prove correct the function Gl that computes GL-proofs for sentences. The results on minimal GL-proofs are needed in the proof. The operation of these central algorithms is demonstrated in Example 4.13.

Lemma 4.11 *Let* Σ *be a finite set of premises and* Q *a finite set of quasi-atoms satisfying* $\mathrm{Sf}^{\mathrm{L}}(\Sigma) \subseteq Q$ *and* $\mathrm{Sf}^{\mathrm{L}}(Q) \subseteq Q$. *Let* Λ *be the set of quasi-literals returned by the call* $\mathsf{Gl}(\Sigma, Q)$ *of the algorithm in Figure 9. Then (i)* Λ *is finite, (ii) the quasi-literals in* Λ *form a GL-proof from* Σ*, (iii)* $L\chi \in Q - \mathrm{qatom}(\Lambda)$ *implies* $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} \chi$ *and* $\langle \Sigma, \Lambda\rangle \not\dashv_{\mathrm{ae}} \chi$. *Furthermore, (iv) the*

Function $\mathsf{Gl}(\Sigma, Q)$: set of quasi-literals;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ: a set of quasi-literals;
  Q, Q': a finite set of quasi-atoms;
  χ: a sentence of $\mathcal{L}_{ae}$;
Begin
  $\Lambda{:=}\emptyset$;
  Repeat
    $Q'{:=}Q$;
    For $L\chi \in Q$ **Do**
      If $\langle 1, \langle \Sigma \cup \Lambda, \chi \rangle\rangle \in$ **CN&IR Then**
        $\Lambda := \Lambda \cup \{L\chi\}$, $Q := Q - \{L\chi\}$
      Elif $\langle 2, \langle \Sigma, \Lambda, \chi \rangle\rangle \in$ **CN&IR Then**
        $\Lambda := \Lambda \cup \{\neg L\chi\}$, $Q := Q - \{L\chi\}$
  Until $Q' = Q$;
  $\mathsf{Gl}{:=}\Lambda$
End

Figure 9: Computation of GL-proofs

Machine $\mathsf{Gcae}(\Sigma, \phi)$: Boolean;
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ: a finite set of quasi-literals;
  ϕ: a sentence of $\mathcal{L}_{ae}$;
Begin
  $\Lambda{:=}\mathsf{Gl}(\Sigma, \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}))$;
  If $\langle 1, \langle \Sigma \cup \Lambda, \phi \rangle\rangle \in$ **CN&IR Then**
    Accept
  Else
    Reject
End

Figure 10: An oracle machine accepting **GCAE**

the number of calls to an oracle for **CN&IR** *is quadratic in* $|Q|$.

Proof. The sets Λ and Q are always updated simultaneously in the algorithm of Figure 9. We shall refer to their values by Λ_i and Q_i where i gives the number of updates made. Initially $\Lambda_0 = \emptyset$ and Q_0 is the initial value of Q.

(i) The loops of the algorithm satisfy invariants $\mathrm{qatom}(\Lambda_i) = Q_0 - Q_i$ and $Q_i = Q_0 - \mathrm{qatom}(\Lambda_i)$. To see this, notice that the invariants hold initially as $\Lambda_0 = \emptyset$. Whenever an update is made a quasi-atom $L\chi$ is removed from Q and either $L\chi$ or $\neg L\chi$ is added to Λ. Therefore the invariants remains intact. The first invariant implies that Λ is finite because Q_0 is.

(ii) Let ϕ_i denote the quasi-literal (either $L\chi_i$ or $\neg L\chi_i$) that is added to Λ_i when the i^{th} update is made. Let us prove by induction on i that $\phi_1, ..., \phi_i$ forms a GL-proof from Σ. The base case is clear because the corresponding set of quasi-literals is empty. Induction step follows. By the inductive hypothesis $\phi_1, ..., \phi_{i-1}$ forms a GL-proof from Σ. Notice that $\Lambda_{i-1} = \{\phi_1, ..., \phi_{i-1}\}$. Thus a positive quasi-literal $\phi_i = L\chi_i$ is added to Λ only when $\Sigma \cup \{\phi_1, ..., \phi_{i-1}\} \models_{\mathrm{ae}} \chi_i$ holds. Similarly, a negative quasi-literal $\phi_i = \neg L\chi_i$ is added Λ only when $\langle \Sigma, \{\phi_1, ..., \phi_{i-1}\}\rangle \dashv_{\mathrm{ae}} \chi_i$ holds. It follows that $\phi_1, ..., \phi_i$ satisfies the requirements of a GL-proof from Σ.

(iii) The **Repeat**-loop exits if no updates are made in the **For**-loop. Let j give the total number of updates made. If $j = |Q_0|$, Q is empty and the the third requirement of the lemma holds vacuously. Otherwise $0 \leq j < |Q_0|$. Then the set of quasi-literals Λ_j will be returned by the algorithm, i.e. $\Lambda = \Lambda_j$. The last execution of the **For**-loop guarantees that for every $L\chi \in Q_j$ it holds that $\Sigma \cup \Lambda_j \not\models_{\mathrm{ae}} \chi$ and $\langle \Sigma, \Lambda_j \rangle \not\dashv_{\mathrm{ae}} \chi$. By the second loop invariant $Q_j = Q_0 - \mathrm{qatom}(\Lambda_j)$. This implies that the claim (iii) holds for Λ as Λ_j is returned by the function Gl.

(iv) The **Repeat**-loop is executed at most $|Q_0|+1$ times. During each execution of the **For**-loop, the oracle for **CN&IR** is called at most $2|Q_i| \leq 2|Q_0|$ times. Hence the number of calls to an oracle for **CN&IR** is quadratic in $|Q_0|$. □

Proposition 4.12 *The set* **GCAE** *is polynomial time Turing-reducible to the set* **CN&IR**.

Proof. The proof is based on the deterministic Turing-machine Gcae given in Figure 10. We show that (i) the machine Gcae accepts its input if and only if $\langle \Sigma, \phi \rangle \in$ **GCAE** and (ii) it performs its computation in a time polynomial to $|\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})|$.

(i) ($\Rightarrow$) Assume that Gcae accepts its input Σ and ϕ. Then there is a GL-proof Λ from Σ satisfying the properties of Lemma 4.11. Furthermore, ϕ is a logical consequence of $\Sigma \cup \Lambda$. This implies that ϕ has a GL-proof from Σ, i.e. $\Sigma \vdash_{\mathsf{C}} \phi$. Thus $\langle \Sigma, \phi \rangle \in$ **GCAE**. ($\Leftarrow$) Let us then assume that $\langle \Sigma, \phi \rangle \in$ **GCAE**, i.e. $\Sigma \vdash_{\mathsf{C}} \phi$. By lemma 3.10 the minimal GL-proofs of ϕ from Σ contain only members of $\mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) \cup \neg \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\})$. By lemma 3.11 ϕ is a logical consequence of every set of quasi-literals Λ' that forms a GL-proof from Σ and satisfies $\Sigma \cup \Lambda' \not\models_{\mathrm{ae}} \chi$ and $\langle \Sigma, \Lambda' \rangle \not\models_{\mathrm{ae}} \chi$ for every $L\chi \in \mathrm{Sf}^{\mathrm{L}}(\Sigma \cup \{\phi\}) - \mathrm{qatom}(\Lambda')$. The set of quasi-literals Λ satisfies these conditions implying that the machine Gcae has an accepting computation on Σ and ϕ. $\square$

Example 4.13 Let $\Sigma = \{\neg Lp \rightarrow q, \neg Lq \rightarrow p, \neg Lr \rightarrow p\}$ be the set of premises. The call $\mathsf{Gcae}(\Sigma, q)$ to the oracle machine Gcae leads to following computations. First of all, the function call $\mathsf{Gl}(\Sigma, \{Lp, Lq, Lr\})$ is made. We assume that the quasi-atoms Lp, Lq and Lr are accessed in this order. The table below describes the operation of Gl in detail.

Λ	Q	Results of oracle calls
$\emptyset$	$\{Lp, Lq, Lr\}$	$\Sigma \not\models_{\mathrm{ae}} p$ $\langle \Sigma, \emptyset \rangle \not\dashv_{\mathrm{ae}} p$ $\Sigma \not\models_{\mathrm{ae}} q$ $\langle \Sigma, \emptyset \rangle \not\dashv_{\mathrm{ae}} q$ $\Sigma \not\models_{\mathrm{ae}} r$ $\langle \Sigma, \emptyset \rangle \dashv_{\mathrm{ae}} r$
$\{\neg Lr\}$	$\{Lp, Lq\}$	$\Sigma \cup \{\neg Lr\} \models_{\mathrm{ae}} p$
$\{\neg Lr, Lp\}$	$\{Lq\}$	$\Sigma \cup \{\neg Lr, Lp\} \not\models_{\mathrm{ae}} q$ $\langle \Sigma, \{\neg Lr, Lp\} \rangle \dashv_{\mathrm{ae}} q$
$\{\neg Lr, Lp, \neg Lq\}$	$\emptyset$	

Therefore $\mathsf{Gl}(\Sigma, \{Lp, Lq, Lr\})$ returns the set of quasi-literals $\{\neg Lr, Lp, \neg Lq\}$. Then a test $\Sigma \cup \{\neg Lr, Lp, Lq\} \models_{\mathrm{ae}} q$ is made. As this does not hold, the input $\langle \Sigma, q \rangle$ is rejected by the machine Gcae. This is justified because q is disbelieved in the least expansion of Σ.

Propositions 4.14 and 4.15 give reducibility results for the brave reasoning problems. First of all, the set **BAE** is polynomial time nondeterministic Turing-reducible to **CNL&CFULL**. By adding the **N**-groundedness requirement, we get reducibility result for **GBAE**. It is polynomial time nondeterministic Turing-reducible to **CNL&CFULL&GND**.

Proposition 4.14 *The set* **BAE** *is polynomial time nondeterministic Turing-reducible to the set* **CNL&CFULL**.

Machine Bae(Σ, ϕ);
Var
  Σ: a finite subset of $\mathcal{L}_{ae}$;
  Λ: a set of quasi-literals belonging to Cand(Σ);
  ϕ: a sentence of $\mathcal{L}_{ae}$;
Begin
  Choose Λ **From** Cand(Σ);
  If $\langle 2, \langle \Sigma, \Lambda \rangle \rangle \notin$ **CNL&CFULL Then**
    Reject;
  If $\langle 1, \langle \Sigma, \Lambda, \phi \rangle \rangle \in$ **CNL&CFULL Then**
    Accept
  Else
    Reject
End

Figure 11: An oracle machine accepting **BAE**

Proof. The set **BAE** is accepted by the nondeterministic Turing-machine given in Figure 11. To see this, consider the situation when the machine Bae halts in an accepting state. Then Λ is a Σ-full set and $\langle \Sigma, \Lambda \rangle \models_{\mathrm{L}} \phi$ holds. By Theorem 2.10 there is a cautious expansion to which ϕ belongs and $\langle \Sigma, \phi \rangle \in$ **BAE**. Conversely, if Σ has a cautious expansion $\langle \Delta^+, \Delta^- \rangle$ such that $\phi \in \Delta^+$. As a consequence, there is a Σ-full set such that $\langle \Sigma, \Lambda \rangle \models_{\mathrm{L}} \phi$ holds. This implies that Bae has an accepting computation given $\langle \Sigma, \phi \rangle$ as its input. Only two queries to an oracle for **CNL&CFULL** are needed. So **BAE** is polynomial time nondeterministic Turing-reducible to **CNL&CFULL**. □

Proposition 4.15 *The set* **GBAE** *is polynomial time nondeterministic Turing-reducible to the set* **CNL&CFULL&GND**.

Proof. Again, by slightly modifying the Turing-machine for **BAE** an accepting machine for **GBAE** is obtained. All that is needed, is the groundedness condition of Proposition 4.10. The rest follows as in the proof of Proposition 4.14. □

Theorem 4.16 gives upper bounds on the complexity by relating the various problems to the polynomial time hierarchy (PH). We have provided completeness results for the sets **CN** and **IR**. The rest of completeness results remain open. Note that the main decision problem in CAEL, **GCAE**, is contained in Δ_3^{p}. Similar reasoning problems in Moore style autoepistemic logic are complete problems on the second level of PH [14, 26].

Theorem 4.16

(i)	**CN** *is* Π_1^{p}*-complete*
(ii)	**IR** *is* Π_2^{p}*-complete*
(iii)	**CNL** $\in \Delta_3^{\mathrm{p}}$
(iv)	**IRL** $\in \Delta_3^{\mathrm{p}}$.
(v)	**CFULL** $\in \Delta_3^{\mathrm{p}}$
(vi)	**BAE** $\in \Sigma_3^{\mathrm{p}}$
(vii)	**CAE** $\in \Pi_3^{\mathrm{p}}$
(viii)	**CNN** $\in \Delta_2^{\mathrm{p}}$
(ix)	**GND** $\in \Delta_2^{\mathrm{p}}$
(x)	**GBAE** $\in \Sigma_3^{\mathrm{p}}$
(xi)	**GCAE** $\in \Delta_3^{\mathrm{p}}$

Proof. (i) The satisfiability problem **SAT** belongs to Σ_1^{p}. The problem $\overline{\mathbf{CN}}$ is polynomial time many-to-one reducible to **SAT** because it can be shown that $\phi \not\models_{\mathrm{ae}} \psi$ if and only if $(\phi\wedge\neg\psi)'$ is satisfiable where $(\phi\wedge\neg\psi)'$ is the propositional sentence obtained by replacing the quasi-L-subformulae $\mathrm{Sf}^{\,\mathrm{qL}}(\phi,\psi)$ of $\phi\wedge\neg\psi$ by new and distinct atomic propositions. Hence $\overline{\mathbf{CN}} \in \Sigma_1^{\mathrm{p}}$ and $\mathbf{CN} \in \Pi_1^{\mathrm{p}}$. For the Σ_1^{p}-hardness of $\overline{\mathbf{CN}}$, notice that the satisfiability of a propositional sentence ϕ can be decided in terms of $\models_{\mathrm{ae}}$ (more precisely $\not\models_{\mathrm{ae}} \neg\phi$). Consequently, **CN** is Π_1^{p}-complete.

(ii) By Proposition 4.3 $\overline{\mathbf{IR}}$ is polynomial time nondeterministic Turing-reducible to **CN** which is Π_1^{p}-complete. Thus **IR** is in Π_2^{p}. By proposition 4.4 $\mathbf{QBF}_{2,\exists}$ is polynomial time many-to-one reducible to $\overline{\mathbf{IR}}$ implying that $\overline{\mathbf{IR}}$ is Σ_2^{p}-complete. Hence, **IR** is Π_2^{p}-complete.

(iii)-(v) The problems **CNL**, **IRL**, and **CFULL** are all polynomial time Turing-reducible to **CN&IR** by Propositions 4.6 and 4.7. As **CN** is Π_1^{p}-complete, **IR** is in Π_2^{p} and $\Pi_1^{\mathrm{p}} \subseteq \Pi_2^{\mathrm{p}}$, **CN&IR** is in Π_2^{p} (here **IR** dominates computational complexity). Thus **CNL**, **IRL**, and **CFULL** are in Δ_3^{p}.

(vi) The problem **BAE** is polynomial time nondeterministic Turing-reducible to the problem **CNL&CFULL** by Proposition 4.14. As **CNL** and **CFULL** are in Δ_3^{p}, **CNL&CFULL** is in Δ_3^{p} and **BAE** is in Σ_3^{p}.

(vii) By proposition 4.8 the problem $\overline{\mathbf{CAE}}$ is polynomial time nondeterministic Turing-reducible to **CNL&CFULL**. Because **CNL&CFULL** is in Δ_3^{p}, **CAE** is in Π_3^{p}.

(viii) Gottlob [14] proves that **CNN** belongs to Δ_2^{p}.

(ix) The problem **GND** is polynomial time Turing-reducible to **CNN** by Proposition 4.9. As **CNN** is in Δ_2^{p}, **GND** belongs to Δ_2^{p} as well.

(x) By Proposition 4.15 the problem **GBAE** is polynomial time nondeterministic Turing-reducible to **CNL&CFULL&GND**. As **CNL&CFULL** belongs to Δ_3^p and **GND** belongs to Δ_2^p, **CNL&CFULL&GND** is in Δ_3^p. Thus **GBAE** belongs to Σ_3^p.

(xi) By Proposition 4.15 the problem $\overline{\mathbf{GBAE}}$ is polynomial time nondeterministic Turing-reducible to **CNL&CFULL&GND**. As **CNL&CFULL&GND** is in Δ_3^p, **GCAE** belongs to Π_3^p. In Proposition 4.12 this result is improved to Δ_3^p using the iterative construction of the least expansion. □

5 Cautious Models for General Logic Programs

As discussed in the introduction, a variety of semantics has been proposed for *general logic programs* (GLPs) that implement negation as failure. However, many of the proposed semantics leave the meaning of some GLPs open. For instance, there are GLPs that do not have a stable model [12]. The well-founded model of a general logic program P [32] can be seen as three-valued approximation of the stable models of P. Such an approximation can always be computed implying that well-founded models exist for every GLP. Therefore the relationship between stable models of a GLP and its well-founded model resembles the one between stable expansions and the least expansion. This suggests that the least expansion can be used to provide a model for GLPs. In this section, we propose a *cautious model* for general logic programs and study its relationship to the well-founded model. We begin by introducing GLPs and their well-founded models.

5.1 General Logic Programs

General logic programs are based on a standard *first-order language* [19]. Such a language contains among others *variable*, *constant*, *function* and *predicate* symbols. *Terms* are either *variables* $x, y, z, \ldots$, *constants* $a, b, c, \ldots$ or expressions of the form $f(t_1, \ldots, t_n)$ where f is an n-ary function symbol and $t_1, \ldots, t_n$ are terms. *Atomic formulae* (or just *atoms* for short) are of the form $p(t_1, \ldots, t_n)$ where p is an n-ary predicate symbol and $t_1, \ldots, t_n$ terms. *Formulae* are constructed from atoms using *connectives* and *quantifiers*. Every set Σ of first-order formulae has the Herbrand-universe and Herbrand-base associated with it. The Herbrand-universe of Σ contains all *ground* (variable-free) terms that are constructible from the constant and function symbols of Σ. Similarly, the Herbrand-base contains all ground atoms that are constructible from the predicate symbols and the terms of the Herbrand-universe of Σ. Having introduced some general terminology of first-order theories, let us turn our attention to the syntax of GLPs. We call atomic formulae *positive literals*. The symbol $\sim$ presents negation as failure. *Negative literals* are of the form $\sim p$ where p is a positive literal. *Literals* are just positive or negative literals. The *complement* of a positive literal p is the negative literal $\sim p$ and vice versa. For a set of positive literals L, $\sim L$ denotes the set $\{\sim p \mid p \in L\}$. *General logic programs* consist of rules of the form

$$p \leftarrow q_1, \ldots, q_n, \sim r_1, \ldots, \sim r_m \qquad (n \geq 0 \text{ and } m \geq 0)$$

where the positive literals $q_1, \ldots, q_n$ and the negative literals $\sim r_1, \ldots, \sim r_m$ are called the *positive and negative subgoals* of the rule, respectively. The subgoals altogether form the *body* of the rule. Because the body is understood as a conjunction of the subgoals, the order of the subgoals is irrelevant and we assume without loss of generality that positive subgoals precede the negative ones. The positive literal p is the *head* of the rule. The *definite logic programs* mentioned in the introduction are negation-free GLPs ($m = 0$). The variables of a GLP are understood to be universally quantified over its Herbrand-universe $\mathcal{U}_P$. Given a general logic program P, the corresponding *instantiated* program is constructed by replacing each rule of P that contains variables by the rules obtained by assigning the variables to all possible combinations of ground terms in $\mathcal{U}_P$. A logic program is *propositional* if it is variable free. If this is the case, all atoms in the program are ground, i.e. members Herbrand base $\mathcal{B}_P$. Thus instantiated GLPs can be seen as propositional logic programs.

5.2 Well-founded Models

The meaning of a definite logic program P can be characterized in terms of Herbrand-models that are subsets of $\mathcal{B}_P$. In fact, a definite logic program has a unique (minimal) Herbrand-model. This makes its declarative meaning clear. Such a unique model ceases to exist when the negation as failure is introduced. Let us now take a closer look at the *well-founded model* of general logic programs [32] which is one of the leading formal models for general logic programs. It is based on a three-valued logic and partial models of general logic programs. A *partial interpretation* I of a general logic program P contains a positive literal $p \in \mathcal{B}_P$ if p is true in I and a negative literal $\sim p$ if p is false in I. Otherwise, the truth value of p is considered undetermined in I. In addition, partial interpretations are required to be consistent, i.e. no partial interpretation contains a literal and its complement simultaneously. A partial interpretation of P is *total* if for every $p \in \mathcal{B}_P$ either p or $\sim p$ belongs to it. A ground literal is true in a partial interpretation I if it belongs to I and false in I if its complement belong to I. A conjunction of literals is true in a partial interpretation I if all the conjuncts are true I. Similarly, a conjunction of literals is false in I if some of the conjuncts is false in I. An instantiated rule is satisfied in I if the head is true in I or the body (i.e. the conjunction of the subgoals) is false in I. A total interpretation I of P is a total model of P if every instantiated rule of P is satisfied in I. A *partial model* of P is a partial interpretation that can be extended to a total model of P.

Define an operator W_P that operates on partial models of P. It is defined by the equation $\mathrm{W}_P(I) = \mathrm{T}_P(I) \cup \sim \mathrm{U}_P(I)$ where

- $\mathrm{T}_P(I)$ gives the heads of the rules whose bodies are true in I

- $\mathrm{U}_P(I)$ is the greatest *unfounded set* of P with respect to I.

A set A of positive literals is an unfounded set of P with respect to I if and only if every element p of A and every rule R of P that has p as its head satisfies at least one the following conditions:

(i) some positive subgoal of the body of R is false in I

(ii) some negative subgoal of the body of R is false in I

(iii) some positive subgoal of the body of R appears in A.

The set $\mathrm{U}_P(I)$ is the greatest unfounded set of P with respect to I. It is the union of the all unfounded sets of P with respect to I. Notice that $\mathrm{T}_P(I)$ and $\mathrm{U}_P(I)$ are disjoint implying that $\mathrm{W}_P(I)$ is consistent. The intuitive idea behind unfounded sets is that the elements of an unfounded set may be set false in the partial model being constructed, as the properties of unfounded sets guarantee that none of the elements of the unfounded set can be derived by the rules of P. Thus, there is an analogy to cautious autoepistemic logic, but the notions of an extension (of a partial model or an expansion) differ. The well-founded model WFM(P) of P is the fixed point of the sequence $I_0, I_1, \ldots$ of partial interpretations of P defined by $I_0 = \emptyset$ and $I_i = \mathrm{W}_P(I_{i-1})$ when $i > 0$. The fixed point in question is the least fixed point of W_P. The construction of the well-founded models is demonstrated in Examples 5.1 and 5.2.

Example 5.1 Let P be the general logic program $\{p \leftarrow\sim p; q \leftarrow p\}$. The set $\mathrm{T}_P(I_0)$ is empty, because the bodies of the rules are not satisfied under I_0. Consider the the set $A_1 = \{p, q\}$ and the rule $p \leftarrow\sim p$ whose head $p \in A_1$. The set A_1 is not an unfounded set of P (with respect to I_0) because I_0 does not falsify $\sim p$ and there are no positive occurrences of the atoms of A_1 in P. By similar analysis the sets $A_2 = \{p\}$ and $A_3 = \{q\}$ are rejected as unfounded sets. It follows that $\emptyset$ is the greatest unfounded set of P with respect to I_0. Consequently, $\mathrm{W}_P(I_0) = I_0$ implying that I_0 is the least fixed point of W_P. Thus, the well-founded model leaves the truth values of p, q and r undetermined.

Example 5.2 Let P be the general logic program $\{q \leftarrow\sim p; p \leftarrow\sim q; r \leftarrow\sim p, \sim q\}$. Let $I_0 = \emptyset$ as required by the construction. Now $\mathrm{T}_P(I_0) = \emptyset$, because I_0 does not satisfy any of the bodies of the clauses in P. Then consider any *nonempty* subset A of $\{p, q, r\}$. The partial interpretation I_0 does not falsify any subgoal of the program clauses in P, as required by the first two unfoundedness conditions. Moreover, there are no positive occurrences of the atoms of A in P, so that the third condition for unfounded sets cannot hold. Therefore A cannot be an unfounded set. Therefore $\mathrm{U}_P(I_0) = \emptyset$. Consequently, $\mathrm{WFM}(P)$ is $\emptyset$.

5.3 Cautious Models

We propose now an alternative model for general logic programs. To do this general logic programs need to be translated into autoepistemic theories. In Definition 5.3 a translation function for instantiated general logic programs into autoepistemic formulae is given. It was first introduced by Gelfond [11]. The intuitive idea behind the translation is to replace negation as failure by disbelief. Elkan [9] has shown that the stable expansions of the translated general logic program correspond exactly to its stable models.

Definition 5.3 *Let P be an instantiated general logic program. Its translation* $\mathrm{Tr}(P)$ *into autoepistemic logic is*

$$\{q_1 \wedge \ldots \wedge q_n \wedge \neg Lr_1 \wedge \ldots \wedge \neg Lr_m \rightarrow p \mid p \leftarrow q_1, \ldots, q_n, \sim r_1, \ldots, \sim r_m \in P\}.$$

Two assumptions about general logic programs are made: they are (i) fully instantiated and (ii) finite. The first assumption is necessary as cautious autoepistemic logic was defined in the propositional case. The second restriction is due to [15] where the results were based on set of premises having only finite number of L-subformulae. However, this restriction is probably unnecessary. It seems that the presented results generalize to the case when (ii) ceases to hold. In Definition 5.4 the alternative model for a general logic program P is given. The proposed *cautious model* $\mathrm{CM}(P)$ is a partial model of P (in accordance with the well-founded model) which is based on the least $\mathbf{N}$-grounded cautious expansion of $\mathrm{Tr}(P)$. The cautious model of a general logic program is total if the expansion in question is also a Moore style stable expansion.

Definition 5.4 *Let P be a general logic program. The cautious model* $\mathrm{CM}(P)$ *of P is determined by the least* $\mathbf{N}$*-grounded* $\mathrm{Tr}(P)$*-full set* Λ *in the following way:*

$$\mathrm{CM}(P) = \{p \in \mathcal{B}_P \mid \mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} p\} \cup \{\sim p \in \sim\mathcal{B}_P \mid \langle \mathrm{Tr}(P), \Lambda\rangle \dashv_{\mathrm{ae}} p\}.$$

Now we intend to prove that the cautious model of P subsumes its well-founded model. On the autoepistemic side, partial interpretations are naturally presented as beliefs of the agent, i.e. sets of quasi-literals. We define $\mathrm{Tr}_P(I)$ as the translation of the partial interpretation I into a set of quasi-literals $\{Lp \mid p \in \mathrm{R}_P(I)\} \cup \{\neg Lp \mid \sim p \in \mathrm{R}_P(I)\}$ where $\mathrm{R}_P(I)$ is the restriction of I to the atoms in the negative subgoals of the rules in P. If $p \in \mathrm{R}_P(I)$, Lp is included in $\mathrm{Tr}_P(I)$ as the agent is assumed to believe p. Similarly, if $\sim p \in \mathrm{R}_P(I)$, $\neg Lp$ is included in $\mathrm{Tr}_P(I)$ to indicate disbelief in p. It follows that $\mathrm{Tr}_P(I)$ is a consistent subset of $\mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P)) \cup \neg\mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P))$. We provide now some basic results that relate the two models. Lemmas 5.5-5.7 relate the positive and negative introspection principles of the well-founded model (i.e. the operators $\mathrm{T}_P(I)$ and $\mathrm{U}_P(I)$, respectively) to their counterparts in the cautious model. Lemma 5.7 essentially shows that the negative introspection in cautious model is at least as strong as in the well-founded model.

Lemma 5.5 *Let P be a general logic program. Let I be a partial model of P and Λ a consistent set of quasi-literals such that (i) $\mathrm{Tr}_P(I) \subseteq \Lambda$ and (ii) $q \in I$ implies $\mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} q$. Then $p \in \mathrm{T}_P(I)$ implies $\mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} p$.*

Proof. Let $p \in \mathrm{T}_P(I)$ hold. Then there is a rule $p \leftarrow q_1, ..., q_n, \sim r_1, ..., \sim r_m$ of P whose all positive and negative subgoals are true in I. Thus $q_1, ..., q_n$ and $\sim r_1, ..., \sim r_m$ are contained in the partial model I. Consequently, the quasi-literals $\neg Lr_1, ..., \neg Lr_m$ belong to $\mathrm{Tr}_P(I)$ and to Λ as well, because (i) holds. By the requirement (ii), each q_i satisfies $\mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} q_i$. Thus $\mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} q_1 \wedge ... \wedge q_n \wedge \neg Lr_1 \wedge ... \wedge \neg Lr_m$ holds. As the translated rule $q_1 \wedge ... \wedge q_n \wedge \neg Lr_1 \wedge ... \wedge \neg Lr_m \rightarrow p$ is in $\mathrm{Tr}(P)$, p is a logical consequence of $\mathrm{Tr}(P) \cup \Lambda$. $\square$

Lemma 5.6 *Let P be a general logic program. Let I be a partial model of P and Λ a consistent set of quasi-literals such that (i) $\mathrm{Tr}_P(I) \subseteq \Lambda$ and (ii) $\sim q \in I$ implies $\mathrm{Tr}(P) \cup \Lambda \not\models_{\mathrm{ae}} q$. Then $p \in \mathrm{U}_P(I)$ implies $\mathrm{Tr}(P) \cup \Lambda \not\models_{\mathrm{ae}} p$.*

Proof. In the proof we construct an autoepistemic model $\langle J, B\rangle$ of $\mathrm{Tr}(P) \cup \Lambda$ and show that every $p \in \mathrm{U}_P(I)$ is false in $\langle J, B\rangle$. Here J is a standard propositional interpretation and B determines the truth values of quasi-literals in $\mathcal{L}_{ae}$ such that $L\chi$ is true in $\langle J, B\rangle$ if and only if $\chi \in B$. Let us define a sequence $\langle J_0, B\rangle, \langle J_1, B\rangle, ...$ of autoepistemic structures. First of all, let $B = \{\chi \mid L\chi \in \Lambda\}$ so that $\langle J_i, B\rangle \models_{\mathrm{ae}} \Lambda$ for all $i \geq 0$. Then define $J_0 = \emptyset$ and sets J_i ($i > 0$) as follows. An atom p belongs to J_i if and only if there is a translated rule $q_1 \wedge ... \wedge q_n \wedge \neg Lr_1 \wedge ... \wedge \neg Lr_m \rightarrow p$ in $\mathrm{Tr}(P)$ such that $q_1 \wedge ... \wedge q_n \wedge \neg Lr_1 \wedge ... \wedge \neg Lr_m$ is true in $\langle J_{i-1}, B\rangle$. By [19] this sequence is non-decreasing and the fixed point $\langle J, B\rangle$ of this sequence is a model for $\mathrm{Tr}(P) \cup \Lambda$. Furthermore $J = \{p \in \mathcal{B}_P \mid \mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} p\}$. (To establish these results, first reduce $\mathrm{Tr}(P)$ to a set of objective implications by assigning truth values to quasi-literals in $\mathrm{Tr}(P)$ according to B). We prove now that no $p \in \mathrm{U}_P(I)$ is a member of J_i by induction on i. Consequently, the sets J and $\mathrm{U}_P(I)$ are disjoint. So $\langle J, B\rangle$ is also a model for every $\neg p$ where $p \in \mathrm{U}_P(I)$ and $\mathrm{Tr}(P) \cup \Lambda \not\models_{\mathrm{ae}} p$ follows.

The base case is clear as $J_0 = \emptyset$. The induction step follows. Let p be a member of $\mathrm{U}_P(I)$. Assume that $p \in J_i$. Then there is a translated rule $q_1 \wedge ... \wedge q_n \wedge \neg Lr_1 \wedge ... \wedge \neg Lr_m \rightarrow p$ such that $q_1 \wedge ... \wedge q_n \wedge \neg Lr_1 \wedge ... \wedge \neg Lr_m$ is true in $\langle J_{i-1}, B \rangle$. It follows that $q_1 \in J_{i-1}, ..., q_n \in J_{i-1}$ and $r_1 \notin B, ..., r_m \notin B$. We also know that $p \leftarrow q_1, ..., q_n, \sim r_1, ..., \sim r_m$ is a rule of P. Because $p \in \mathrm{U}_P(I)$ there are three possible cases. (a) If some *positive* subgoal q_j of the rule is false in I, then $\sim q_j \in I$ and we get $\mathrm{Tr}(P) \cup \Lambda \not\models_{\mathrm{ae}} q_j$ by the requirement (ii). Because $J = \{q \in \mathcal{B}_P \mid \mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} q\}$ it follows that $q_j \notin J$. As $J_{i-1} \subseteq J$, $q_j \notin J$ implies $q_j \notin J_{i-1}$ i.e. a contradiction. (b) If some *negative* subgoal $\sim r_j$ is false in I, r_j belongs to I. As $\sim r_j$ is a negative subgoal of P, Lr_j is a member of $\mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P))$. Thus $r_j \in I$ yields that $Lr_j \in \mathrm{Tr}_P(I)$. By the requirement (i), Lr_j belongs to Λ. It follows that $r_j \in B$ which is a contradiction. (c) No positive subgoal q_j can be a member of $\mathrm{U}_P(I)$ because each q_j is a member of J_{i-1} and by the inductive hypothesis there is no $q \in \mathrm{U}_P(I)$ such that $q \in J_{i-1}$. By the above case analysis (a-c), $p \in J_i$ is contradictory. Hence $p \notin J_i$ for all $i \geq 0$. Thus no $p \in \mathrm{U}_P(I)$ is a member of the fixed point J. □

Lemma 5.7 *Let P be a general logic program. Let I be a partial model of P such that (i) $\sim q \in I$ implies $\langle \mathrm{Tr}(P), \mathrm{Tr}_P(I) \rangle \dashv_{\mathrm{ae}} q$. Then $p \in \mathrm{U}_P(I)$ implies $\langle \mathrm{Tr}(P), \mathrm{Tr}_P(I) \rangle \dashv_{\mathrm{ae}} p$.*

Proof. We write Λ for $\mathrm{Tr}_P(I)$. Let ε be any enumeration of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P)) - \mathrm{qatom}(\Lambda)\}.$$

Notice that $\varepsilon = \psi_1, ..., \psi_n$ consists of atoms only. Let us then verify the requirements of Lemma 5.6 for the set of quasi-literals $\Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda)$. The first condition is satisfied as $\mathrm{Tr}_P(I)$ is a subset of $\Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda)$. For the second condition, assume $\sim q \in I$. By the requirement (i) it follows that $\langle \mathrm{Tr}(P), \Lambda \rangle \dashv_{\mathrm{ae}} q$. Now assuming $\mathrm{Tr}(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda) \models_{\mathrm{ae}} q$ implies $\langle \mathrm{Tr}(P), \Lambda \rangle \not\dashv_{\mathrm{ae}} q$. But this is a contradiction. Hence

$$\mathrm{Tr}(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda) \not\models_{\mathrm{ae}} q.$$

Then by Lemma 5.6 $p \in \mathrm{U}_P(I)$ implies $\mathrm{Tr}(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda) \not\models_{\mathrm{ae}} p$.

Let $p \in \mathrm{U}_P(I)$ hold. Assuming that $\langle \mathrm{Tr}(P), \Lambda \rangle \not\dashv_{\mathrm{ae}} p$ implies that there is an enumeration ε of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P)) - \mathrm{qatom}(\Lambda)\}$ such that

$$\mathrm{Tr}(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda) \models_{\mathrm{ae}} p.$$

This is a contradiction, as $p \in \mathrm{U}_P(I)$ implies $\mathrm{Tr}(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}(P) \cup \Lambda) \not\models_{\mathrm{ae}} p$. Thus $\langle \mathrm{Tr}(P), \Lambda \rangle \dashv_{\mathrm{ae}} p$. □

Van Gelder and al. [32] have shown that the sequence $I_0, I_1, ...$ generated by W_P is both monotonic and consists of partial models of P. In Proposition 5.8 we provide a characterization of $\mathrm{Tr}_P(\mathrm{WFM}(P))$ in terms of a $\mathrm{Tr}(P)$-irrefutable generating enumeration (this implies in fact that $\mathrm{Tr}_P(\mathrm{WFM}(P))$

is N-grounded). Furthermore, the translation $\mathrm{Tr}_P(\mathrm{WFM}(P))$ is contained in the least N-grounded $\mathrm{Tr}(P)$-full set as shown in Proposition 5.9.

Proposition 5.8 *Let P be a general logic program. Then (i) the set of quasi-literals $\mathrm{Tr}_P(\mathrm{WFM}(P))$ has a $\mathrm{Tr}(P)$-irrefutable generating enumeration ε from $\mathrm{Tr}(P)$. Moreover, (ii) $q \in \mathrm{WFM}(P)$ implies $\mathrm{Tr}(P) \cup \mathrm{Tr}_P(\mathrm{WFM}(P)) \models_{\mathrm{ae}} q$ and (iii) $\sim q \in \mathrm{WFM}(P)$ implies $\langle \mathrm{Tr}(P), \mathrm{Tr}_P(\mathrm{WFM}(P))\rangle \dashv_{\mathrm{ae}} q$.*

Proof. We prove by induction on i that $\Lambda_i = \mathrm{Tr}_P(I_i)$ has a $\mathrm{Tr}(P)$-irrefutable generating enumeration ε_i from $\mathrm{Tr}(P)$ and I_i satisfies the conditions (ii) and (iii) where I_i is substituted for $\mathrm{WFM}(P)$. The base case is clear, because the empty enumeration λ generates $\Lambda_i = \emptyset$ from $\mathrm{Tr}(P)$. It is also $\mathrm{Tr}(P)$-irrefutable. The conditions (ii) and (iii) are also vacuously satisfied.

Induction step follows. By the inductive hypothesis the set of quasi-literals Λ_{i-1} has a $\mathrm{Tr}(P)$-irrefutable generating enumeration for ε_{i-1} from $\mathrm{Tr}(P)$. Furthermore, I_{i-1} satisfies the conditions (ii) and (iii).

Let us now show how the enumeration ε_{i-1} can be extended to a generating enumeration ε_i for Λ_i without breaking $\mathrm{Tr}(P)$-irrefutability. Let $\varepsilon^- = \psi_1, \ldots, \psi_n$ be any enumeration of $\{\psi \mid \neg L\psi \in \Lambda_i - \Lambda_{i-1}\}$. It is then proven by induction on j that $\{\neg L\psi_k \mid 0 < k \leq j\} = \mathrm{B}_j^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1})$ and

$$\langle \mathrm{Tr}(P), \Lambda_{i-1} \cup \mathrm{B}_{k-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1})\rangle \dashv_{\mathrm{ae}} \psi_k$$

holds for each $0 < k \leq j$. The base case clearly holds because $j = 0$ implies that both sets are empty (there is no k such that $0 < k \leq 0$). The induction step with respect to j follows. By the inductive hypothesis, $\{\neg L\psi_k \mid 0 < k \leq j-1\} = \mathrm{B}_{j-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1})$ and for each $0 < k \leq j-1$ it holds that

$$\langle \mathrm{Tr}(P), \Lambda_{i-1} \cup \mathrm{B}_{k-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1})\rangle \dashv_{\mathrm{ae}} \psi_k.$$

Assume then that $\langle \mathrm{Tr}(P), \Lambda_{i-1} \cup \mathrm{B}_{j-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1})\rangle \not\dashv_{\mathrm{ae}} \psi_j$ holds. Notice that ε^- contains only propositional atoms. Thus $\mathrm{Sf}^{\mathrm{L}}(\psi_j)$ is empty. So there is an enumeration ε' of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P)) - \mathrm{qatom}(\Lambda_{i-1} \cup \mathrm{B}_{j-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1}))\}$$

such that $\psi_j \in \mathrm{EB}^{\varepsilon'}(\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup \mathrm{B}_{j-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1}))$. Let ε_{j-1}^- denote the enumeration consisting of the first $j-1$ elements of ε^-. As ε_{j-1}^- generates $\mathrm{B}_{j-1}^{\varepsilon^-}(\mathrm{Tr}(P) \cup \Lambda_{i-1})$ from $\mathrm{Tr}(P) \cup \Lambda_{i-1}$, there is also an enumeration of $\{\psi \mid L\psi \in \mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P)) - \mathrm{qatom}(\Lambda_{i-1})\}$, namely $\varepsilon_{j-1}^- \cdot \varepsilon'$, such that $\psi_j \in \mathrm{EB}^{\varepsilon_{j-1}^- \cdot \varepsilon'}(\mathrm{Tr}(P) \cup \Lambda_{i-1})$. This implies $\langle \mathrm{Tr}(P), \Lambda_{i-1}\rangle \not\dashv_{\mathrm{ae}} \psi_j$. Notice that the requirement of Lemma 5.7. is satisfied by the inductive hypothesis (iii).

Thus we get $\langle \mathrm{Tr}(P), \Lambda_{i-1}\rangle \dashv_{\mathrm{ae}} \psi_j$ by Lemma 5.7. But this is a contradiction. Hence, we have proven that $\langle \mathrm{Tr}(P), \Lambda_{i-1} \cup \mathrm{B}^{\varepsilon^-}_{j-1}(\mathrm{Tr}(P) \cup \Lambda_{i-1})\rangle \dashv_{\mathrm{ae}} \psi_j$ holds and $\varepsilon_{i-1} \cdot \varepsilon^-_j$ is $\mathrm{Tr}(P)$-irrefutable. Furthermore, it follows that

$$\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup \mathrm{B}^{\varepsilon^-}_{j-1}(\mathrm{Tr}(P) \cup \Lambda_{i-1}) \not\models_{\mathrm{ae}} \psi_j$$

by the mutual exclusion of $\models_{\mathrm{ae}}$ and $\dashv_{\mathrm{ae}}$. Thus $\neg L\psi_j \in \mathrm{B}^{\varepsilon^-}_j(\mathrm{Tr}(P) \cup \Lambda_{i-1})$ and the set $\{\neg L\psi_k \mid 0 < k \leq j\}$ equals to $\mathrm{B}^{\varepsilon^-}_j(\mathrm{Tr}(P) \cup \Lambda_{i-1})$. So we have shown that the enumeration ε^- generates $\{\neg L\psi_1, ..., \neg L\psi_n\}$ from $\mathrm{Tr}(P) \cup \Lambda_{i-1}$ and $\varepsilon_{i-1} \cdot \varepsilon^-$ is $\mathrm{Tr}(P)$-irrefutable.

Then consider any enumeration $\varepsilon^+ = \psi'_1, ..., \psi'_m$ of $\{\psi \mid L\psi \in \Lambda_i - \Lambda_{i-1}\}$. Like ε^-, the enumeration ε^+ contains only propositional atoms. Let B^- be the set $\{\neg L\psi_1, ..., \neg L\psi_n\}$ generated by ε^-. Now, each ψ'_j is contained in $\mathrm{T}_P(I_{i-1})$. The set of quasi-literals $\Lambda_{i-1} \cup B^- \cup \mathrm{B}^{\varepsilon^+}_{j-1}(\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^-)$ satisfies the requirements of Lemma 5.5 because $\mathrm{Tr}_P(I_{i-1})$ is contained in it and $q \in I_{i-1}$ implies $\mathrm{Tr}(P) \cup \Lambda_{i-1} \models_{\mathrm{ae}} q$ by the inductive hypothesis. It follows that $\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^- \cup \mathrm{B}^{\varepsilon^+}_{j-1}(\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^-) \models_{\mathrm{ae}} q$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus we obtain

$$\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^- \cup \mathrm{B}^{\varepsilon^+}_{j-1}(\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^-) \models_{\mathrm{ae}} \psi'_j$$

by Lemma 5.5. It follows that each $L\psi'_j$ belongs to $\mathrm{B}^{\varepsilon^+}_j(\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^-)$ and the enumeration ε^+ generates $\{L\psi'_1, ..., L\psi'_m\}$ from $\mathrm{Tr}(P) \cup \Lambda_{i-1} \cup B^-$. The enumeration ε_i can now be formed as the concatenation of ε_{i-1}, ε^- and ε^+. It is $\mathrm{Tr}(P)$-irrefutable (because $\varepsilon_{i-1} \cdot \varepsilon^-$ is) and generates Λ_i from $\mathrm{Tr}(P)$.

What remains is the verification of conditions (ii) and (iii) for I_i. This is done as follows. Let $q \in I_i$ hold i.e. $q \in \mathrm{T}_P(I_{i-1})$. The conditions of Lemma 5.5 are fulfilled for I_{i-1} and Λ_i because $\mathrm{Tr}_P(I_{i-1}) \subseteq \Lambda_i$ by the monotonicity of the sequence. In addition, $q \in I_{i-1}$ implies $\Sigma \cup \Lambda_{i-1} \models_{\mathrm{ae}} q$ by the inductive hypothesis. As $\Lambda_{i-1} \subseteq \Lambda_i$ it follows that $\Sigma \cup \Lambda_i \models_{\mathrm{ae}} q$ by the monotonicity of $\models_{\mathrm{ae}}$. Thus by Lemma 5.5, $q \in \mathrm{T}_P(I_{i-1})$ implies $\mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} q$ as required by (ii). Consider then the case $\sim q \in I_i$ i.e. $q \in \mathrm{U}_P(I_{i-1})$. The condition of Lemma 5.7 for I_{i-1} is fulfilled because I_{i-1} satisfies the condition (iii) by inductive hypothesis. Thus $q \in \mathrm{U}_P(I_{i-1})$ implies $\langle \mathrm{Tr}(P), \Lambda_{i-1}\rangle \dashv_{\mathrm{ae}} q$. As $\varepsilon^- \cdot \varepsilon^+$ generates $\Lambda_i - \Lambda_{i-1}$ from $\mathrm{Tr}(P) \cup \Lambda_{i-1}$, it follows that $\langle \mathrm{Tr}(P), \Lambda_i\rangle \dashv_{\mathrm{ae}} q$ by Lemma 2.24 and I_i satisfies condition (iii). This completes the inductive proof.

Thus the translation of fixed point of the sequence i.e. $\mathrm{Tr}_P(\mathrm{WFM}(P))$ has a $\mathrm{Tr}(P)$-irrefutable generating enumeration from $\mathrm{Tr}(P)$ such that (i) is satisfied. Furthermore, $\mathrm{WFM}(P)$ satisfies the conditions (ii) and (iii). $\square$

Proposition 5.9 *Let P be a general logic program. Then $\mathrm{Tr}_P(\mathrm{WFM}(P))$ is contained in the least $\mathrm{Tr}(P)$-full set.*

Proof. Let Λ be the least **N**-grounded $\mathrm{Tr}(P)$-full set. By Proposition 5.8, the set of quasi-literals $\mathrm{Tr}_P(\mathrm{WFM}(P))$ has a $\mathrm{Tr}(P)$-irrefutable generating enumeration $\varepsilon = \psi_1, ..., \psi_n$ from $\mathrm{Tr}(P)$. Therefore, $\neg L\psi_i \in \mathrm{B}_i^\varepsilon(\mathrm{Tr}(P))$ implies $\langle \mathrm{Tr}(P), \mathrm{B}_{i-1}^\varepsilon(\mathrm{Tr}(P))\rangle \dashv_{\mathrm{ae}} \psi_i$ for every $0 < i \leq n$. We show by induction on i that $\mathrm{B}_i^\varepsilon(\mathrm{Tr}(P)) \subseteq \Lambda$. The base case holds as $\mathrm{B}_0^\varepsilon(\mathrm{Tr}(P)) = \emptyset$. The induction step follows. Let $L\psi_i \in \mathrm{B}_i^\varepsilon(\mathrm{Tr}(P))$, i.e. $\mathrm{Tr}(P) \cup \mathrm{B}_{i-1}^\varepsilon(\mathrm{Tr}(P)) \models_{\mathrm{ae}} \psi_i$. Then $\mathrm{Tr}(P) \cup \Lambda \models_{\mathrm{ae}} \psi_i$ follows by the inductive hypothesis and monotonicity of $\models_{\mathrm{ae}}$. As $L\psi_i$ belongs to $\mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P))$, it is a member of Λ as well. Consider next the case $\neg L\psi_i \in \mathrm{B}_i^\varepsilon(\mathrm{Tr}(P))$. It follows that $\langle \mathrm{Tr}(P), \mathrm{B}_{i-1}^\varepsilon(\mathrm{Tr}(P))\rangle \dashv_{\mathrm{ae}} \psi_i$ as ε is $\mathrm{Tr}(P)$-irrefutable. As Λ is an **N**-grounded $\mathrm{Tr}(P)$-full set, it has a generating enumeration from $\mathrm{Tr}(P)$. It also satisfies the consistency requirement $\neg L\chi \in \Lambda$ implies $\mathrm{Tr}(P) \cup \Lambda \not\models_{\mathrm{ae}} \chi$ of Lemma 2.22. Because $\mathrm{B}_{i-1}^\varepsilon(\mathrm{Tr}(P)) \subseteq \Lambda$ by the inductive hypothesis, there is an enumeration ε_d that generates the difference $\Lambda - \mathrm{B}_{i-1}^\varepsilon(\mathrm{Tr}(P))$ from $\mathrm{Tr}(P) \cup \mathrm{B}_{i-1}^\varepsilon(\mathrm{Tr}(P))$. It follows that $\langle \mathrm{Tr}(P), \Lambda\rangle \dashv_{\mathrm{ae}} \psi_i$ by Lemma 2.24. Because $L\psi_i$ belongs to $\mathrm{Sf}^{\mathrm{L}}(\mathrm{Tr}(P))$, $\neg L\psi_i \in \Lambda$. Thus $\mathrm{B}_i^\varepsilon(\mathrm{Tr}(P)) \subseteq \Lambda$. □

The main result of this section is formulated in Theorem 5.11: the well-founded model of a general logic program P is subsumed by the cautious model of P. Thus, the cautious model provides at least as accurate set of conclusions as the well-founded model does. The general logic program $p \leftarrow\sim p$ is an example for which the two semantics coincide (they both leave p undetermined). However, given the general logic program $P = \{q \leftarrow\sim p; p \leftarrow\sim q; r \leftarrow\sim p, \sim q\}$ of Example 5.2 a difference arises. In the well-founded model of P, all the propositions p, q and r remain undetermined. However $\sim r$ is included in the cautious model as Example 5.10 indicates.

Example 5.10 Let P be the general logic program $\{q \leftarrow\sim p; p \leftarrow\sim q; r \leftarrow\sim p, \sim q\}$. Its translation $\mathrm{Tr}(P)$ into autoepistemic logic is $\Sigma = \{\neg Lp \rightarrow q, \neg Lq \rightarrow p, \neg Lp \wedge \neg Lq \rightarrow r\}$. Now both p and q are refutable from Σ (by the enumerations $\varepsilon_1 = q, p$ and $\varepsilon_2 = p, q$, respectively). However, these enumerations do not refute r implying that $\langle \Sigma, \emptyset\rangle \dashv_{\mathrm{ae}} r$ where $\emptyset$ is the least **N**-grounded Σ-full set. Consequently, $\sim r \in \mathrm{CM}(P)$.

Theorem 5.11 *Given a general logic program P,* $\mathrm{WFM}(P) \subseteq \mathrm{CM}(P)$.

Proof. Let Σ be the translation $\mathrm{Tr}(P)$ and Λ the least **N**-grounded Σ-full set. (i) Let p be an atom such that $p \in \mathrm{WFM}(P)$. Then $p \in \mathrm{T}_P(\mathrm{WFM}(P))$ and $\Sigma \cup \mathrm{Tr}_P(\mathrm{WFM}(P)) \models_{\mathrm{ae}} p$ follows by Proposition 5.8. Because $\mathrm{Tr}_P(\mathrm{WFM}(P))$ is a subset of Λ by Proposition 5.9, $\Sigma \cup \Lambda \models_{\mathrm{ae}} p$ follows by the monotonicity of $\models_{\mathrm{ae}}$ and thus $p \in \mathrm{CM}(P)$. (ii) Let p be an atom such that $\sim p \in \mathrm{WFM}(P)$. Then $\langle \Sigma, \mathrm{Tr}_P(\mathrm{WFM}(P))\rangle \dashv_{\mathrm{ae}} p$ follows by Proposition 5.8. Because Λ is an **N**-grounded Σ-full set, it has a generating enumeration and it satisfies the consistency requirement $\neg L\chi \in \Lambda$ implies $\Sigma \cup \Lambda \not\models_{\mathrm{ae}} \chi$. Because $\mathrm{Tr}_P(\mathrm{WFM}(P))$

is contained in Λ by Proposition 5.9, there is an enumeration ε_d that generates the difference $\Lambda - \mathrm{Tr}_P(\mathrm{WFM}(P))$ from $\Sigma \cup \mathrm{Tr}_P(\mathrm{WFM}(P))$ by Lemma 2.22. It follows that $\langle \Sigma, \Lambda \rangle \dashv_{\mathrm{ae}} p$ by Lemma 2.24 and so $\sim p$ belongs to $\mathrm{CM}(P)$. By the cases (i) and (ii), the theorem holds. □

5.4 An Alternative Translation for General Logic Programs

The cautious model was defined in terms of the translation function suggested by Gelfond. In Definition 5.12 an alternative translation for GLPs is given. This translation function was first introduced by Truszczyński [31] to embed GLPs into McDermott and Doyle style modal nonmonotonic logics. Again, the negation as failure is replaced by disbelief. In addition each subgoal of the rule is assumed to be believed by the agent. This forces the implications of the translation to act as one-directional inference rules. We shall next examine another cautious model based on this translation. As it is apparent from Definition 5.13, the formal definition of the model is not affected by the change of translation.

Definition 5.12 *Let P be an instantiated general logic program. Its translation* $\mathrm{Tr}'(P)$ *into autoepistemic logic is*

$$\{Lq_1 \wedge \ldots \wedge Lq_n \wedge L\neg Lr_1 \wedge \ldots \wedge L\neg Lr_m \rightarrow p \mid \\ p \leftarrow q_1, \ldots, q_n, \sim r_1, \ldots, \sim r_m \in P\}.$$

Definition 5.13 *Let P be a general logic program. The alternative cautious model* $\mathrm{CM}'(P)$ *of P is determined by the the least* $\mathbf{N}$*-grounded* $\mathrm{Tr}'(P)$*-full set* Λ *in the following way:*

$$\mathrm{CM}'(P) = \{p \in \mathcal{B}_P \mid \mathrm{Tr}'(P) \cup \Lambda \models_{\mathrm{ae}} p\} \cup \{\sim p \in \sim \mathcal{B}_P \mid \langle \mathrm{Tr}'(P), \Lambda \rangle \dashv_{\mathrm{ae}} p\}.$$

The translation of partial models of GLPs has to be reconsidered. We define $\mathrm{Tr}'_P(I)$ as the translation of the partial interpretation I into a set of quasi-literals $\{Lp \mid p \in \mathrm{R}_P(I)\} \cup \{\neg Lp \mid \sim p \in \mathrm{R}_P(I)\} \cup \{L\neg Lp \mid \sim p \in \mathrm{R}_P(I)\}$. Here $\mathrm{R}_P(I)$ denotes the restriction of I to the atoms that appear in the bodies of the rules in P. If $p \in \mathrm{R}_P(I)$, Lp is included in $\mathrm{Tr}'_P(P)$ as the agent is assumed to be believe p on the autoepistemic side. If $\sim p \in \mathrm{R}_P(I)$, $L\neg Lp$ is included in $\mathrm{Tr}'_P(P)$ as the agent is assumed to believe that p is not believed. The quasi-literal $\neg Lp$ is also included to keep $\mathrm{Tr}'_P(P)$ closed with respect to subbeliefs. In other words, the belief in the fact that "p is disbelieved" involves disbelief in p. In the following we present three lemmas that correspond to the three given for the former translation. There are slight differences due to differences in the translations.

Lemma 5.14 *Let P be a general logic program. Let I be a partial model of P and Λ a consistent set of quasi-literals such that $\mathrm{Tr}'_P(I) \subseteq \Lambda$. Then $p \in \mathrm{T}_P(I)$ implies $\mathrm{Tr}'(P) \cup \Lambda \models_{\mathrm{ae}} p$.*

Proof. Let $p \in \mathrm{T}_P(I)$ hold. Then there is a rule $p \leftarrow q_1, ..., q_n, \sim r_1, ..., \sim r_m$ of P whose positive and negative subgoals are true in I. Thus $q_1, ..., q_n$ and $\sim r_1, ..., \sim r_m$ belong to the partial interpretation I. Consequently, the quasi-literals $Lq_1, ..., Lq_n$, $\neg Lr_1, ..., \neg Lr_m$ and $L\neg Lr_1, ..., L\neg Lr_m$ are included in $\mathrm{Tr}'_P(I)$. The translation $\mathrm{Tr}'(P)$ contains the sentence $Lq_1 \wedge ... \wedge Lq_n \wedge L\neg Lr_1 \wedge ... \wedge L\neg Lr_m \rightarrow p$, implying that p is a logical consequence of $\mathrm{Tr}'(P) \cup \mathrm{Tr}'_P(I)$. As $\mathrm{Tr}'_P(I)$ is a subset of Λ, $\mathrm{Tr}'(P) \cup \Lambda \models_{\mathrm{ae}} p$ follows by the monotonicity of $\models_{\mathrm{ae}}$. □

Lemma 5.15 *Let P be a general logic program. Let I be a partial model of P and Λ a consistent set of quasi-literals such that (i) $\mathrm{Tr}'_P(I) \subseteq \Lambda$, (ii) $L\neg Lp \in \Lambda$ implies $\neg Lp \in \Lambda$ and (iii) there is no $Lq \in \Lambda$ such that $q \in \mathrm{U}_P(I)$. Then $p \in \mathrm{U}_P(I)$ implies $\mathrm{Tr}'(P) \cup \Lambda \not\models_{\mathrm{ae}} p$.*

Proof. Let us again construct an autoepistemic structure $\langle J, B\rangle$. The consistent set of quasi-literals Λ is satisfied by setting $B = \{\chi \mid L\chi \in \Lambda\}$. Then define the propositional interpretation J as follows. An atom p belongs to J if and only if there is a translated rule $Lq_1 \wedge ... \wedge Lq_n \wedge L\neg Lr_1 \wedge ... \wedge L\neg Lr_m \rightarrow p$ in $\mathrm{Tr}'(P)$ such that $Lq_1 \wedge ... \wedge Lq_n \wedge L\neg Lr_1 \wedge ... \wedge L\neg Lr_m$ is true under the current choice for B. This ensures that every sentence of $\mathrm{Tr}'(P)$ is true in $\langle J, B\rangle$ and thus $\langle J, B\rangle \models_{\mathrm{ae}} \mathrm{Tr}'(P) \cup \Lambda$ holds. We prove now that no $p \in \mathrm{U}_P(I)$ is a member of J. Thus $\langle J, B\rangle \models_{\mathrm{ae}} \neg p$ holds for every $p \in \mathrm{U}_P(I)$ and $\mathrm{Tr}'(P) \cup \Lambda \not\models_{\mathrm{ae}} p$ follows.

Let p be any member of $\mathrm{U}_P(I)$ and $R = p \leftarrow q_1, ..., q_n, \sim r_1, ..., \sim r_m$ any rule of P having p as its head. The definition of unfounded sets leaves us three possible choices to consider. (a) If some positive subgoal q_i of R is false in I, then $\sim q_i \in I$ and $\neg Lq_i \in \mathrm{Tr}'_P(I) \subseteq \Lambda$ implying that Lq_i is false in $\langle J, B\rangle$. Thus $Lq_1 \wedge ... \wedge Lq_n \wedge L\neg Lr_1 \wedge ... \wedge L\neg Lr_m$ is false in $\langle J, B\rangle$ and by the above construction p is not included in J. (b) Quite similarly, in the case of a negative subgoal $\sim r_i$ being false in I, $r_i \in I$ and $Lr_i \in \mathrm{Tr}'_P(I) \subseteq \Lambda$. As Λ is consistent, $\neg Lr_i \notin \Lambda$. Assuming $L\neg Lr_i \in \Lambda$ leads to contradiction by the requirement (ii). Thus $L\neg Lr_i \notin \Lambda$ and $L\neg Lr_i$ is false in $\langle J, B\rangle$. It follows that $Lq_1 \wedge ... \wedge Lq_n \wedge L\neg Lr_1 \wedge ... \wedge L\neg Lr_m$ is also false in $\langle J, B\rangle$ and p does not belong to J. (c) The R has a positive subgoal q_i that belongs to $\mathrm{U}_P(I)$. By the requirement (iii) only the quasi-literal $\neg Lq_i$ can appear in Λ. Consequently the definition of B implies that Lq_i evaluates to false in $\langle J, B\rangle$. Thus $Lq_1 \wedge ... \wedge Lq_n \wedge L\neg Lr_1 \wedge ... \wedge L\neg Lr_m$ is also false in $\langle J, B\rangle$ and p is not included in J. By the above case analysis (a-c), $p \notin J$ and $\langle J, B\rangle \models_{\mathrm{ae}} \neg p$. □

Lemma 5.16 *Let P be a general logic program. Let I be a partial model of P such that $I \subseteq \mathrm{W}_P(I)$. Then $p \in \mathrm{U}_P(I)$ implies $\langle \mathrm{Tr}'(P), \mathrm{Tr}'_P(I)\rangle \dashv_{\mathrm{ae}} p$.*

Proof. We write Λ for $\mathrm{Tr}'_P(I)$. Let ε be any enumeration of

$$\{\psi \mid L\psi \in \mathrm{Sf}^{\,\mathrm{L}}(\mathrm{Tr}'(P)) - \mathrm{qatom}(\Lambda)\}.$$

We prove by induction on i that $\Lambda \cup \mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$ satisfies the requirements (i-iii) of Lemma 5.15. Consider the base case $i = 0$. Now $\mathrm{B}^{\varepsilon}_0(\mathrm{Tr}'(P) \cup \Lambda)$ is empty and therefore the requirements should be verified for Λ. The first condition holds as $\Lambda = \mathrm{Tr}'_P(I)$. The second condition is satisfied by the definition of $\mathrm{Tr}'_P(I)$. Assume then that there is $q \in \mathrm{U}_P(I)$ such that $Lq \in \Lambda$. This implies that $q \in I \subseteq \mathrm{W}_P(I)$ and $q \in \mathrm{T}_P(I)$, a contradiction. Thus the third condition is satisfied.

Induction step follows. The set of quasi-literals $\Lambda \cup \mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$ is consistent and $\Lambda = \mathrm{Tr}'_P(I)$ is its subset. Thus the first requirement is met. By the second requirement, $L\neg Lq \in \Lambda \cup \mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$ should imply $\neg Lq \in \Lambda \cup \mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$. If $L\neg Lq$ belongs to Λ, $\neg Lq$ belongs to it as well, by the definition of $\mathrm{Tr}'_P(I)$. However, if $L\neg Lq$ belongs to $\mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$, then $\psi_j = \neg Lq$ for some j and $\mathrm{Tr}'(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}_{j-1}(\mathrm{Tr}'(P) \cup \Lambda) \models_{\mathrm{ae}} \neg Lq$ holds. As $\mathrm{Tr}'(P)$ does not contain quasi-literals and $\mathrm{Tr}'(P)$ is propositionally consistent with $\Lambda \cup \mathrm{B}^{\varepsilon}_{j-1}(\mathrm{Tr}'(P) \cup \Lambda)$, $\neg Lq$ necessarily belongs to $\Lambda \cup \mathrm{B}^{\varepsilon}_{j-1}(\mathrm{Tr}'(P) \cup \Lambda)$. This is a subset of $\Lambda \cup \mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$ as $0 < j < i$. The verification of the third requirement follows. It should be proven that there is no Lq in $\Lambda \cup \mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$ such that $q \in \mathrm{U}_P(I)$. Assuming that there is such an element Lq in Λ implies that $q \in I \subseteq \mathrm{W}_P(I)$ and $q \in \mathrm{T}_P(I)$, which is a contradiction. Thus there is no such element in Λ. By the inductive hypothesis such elements are not contained in $\mathrm{B}^{\varepsilon}_{i-1}(\mathrm{Tr}'(P) \cup \Lambda)$. The only possibility is therefore that $\psi_i = q$. By Lemma 5.15 it holds that $\mathrm{Tr}'(P) \cup \Lambda \cup \mathrm{B}^{\varepsilon}_{i-1}(\mathrm{Tr}'(P) \cup \Lambda) \not\models_{\mathrm{ae}} q$ and therefore Lq is not included in $\mathrm{B}^{\varepsilon}_i(\mathrm{Tr}'(P) \cup \Lambda)$. Therefore the requirements of Lemma 5.15 are satisfied. This completes the induction step.

Consequently, the set of quasi-literals $\Lambda \cup \mathrm{B}^{\varepsilon}(\mathrm{Tr}'(P) \cup \Lambda)$ satisfies the requirements of Lemma 5.15. As this is the case, assuming $\langle \mathrm{Tr}'(P), \Lambda\rangle \not\dashv_{\mathrm{ae}} p$ for some $p \in \mathrm{U}_P(I)$ leads to a contradiction. Hence, $\langle \mathrm{Tr}'(P), \Lambda\rangle \dashv_{\mathrm{ae}} p$. □

Let us then provide a $\mathrm{Tr}'(P)$-irrefutable generating enumeration for the set of quasi-literals $\mathrm{Tr}'_P(\mathrm{WFM}(P))$ from $\mathrm{Tr}'(P)$. This implies in fact that the translation $\mathrm{Tr}'_P(\mathrm{WFM}(P))$ is **N**-grounded.

Proposition 5.17 *Let P be a general logic program. Then (i) the set of quasi-literals $\mathrm{Tr}'_P(\mathrm{WFM}(P))$ has a $\mathrm{Tr}'(P)$-irrefutable generating enumeration ε from $\mathrm{Tr}'(P)$. Moreover,*

(ii) $q \in \mathrm{WFM}(P)$ *implies* $\mathrm{Tr}'(P) \cup \mathrm{Tr}'_P(\mathrm{WFM}(P)) \models_{\mathrm{ae}} q$ *and*
(iii) $\sim q \in \mathrm{WFM}(P)$ *implies* $\langle \mathrm{Tr}'(P), \mathrm{Tr}'_P(\mathrm{WFM}(P))\rangle \dashv_{\mathrm{ae}} q$.

Proof. The proof is obtained with slight modifications from the proof of Proposition 5.8. First of all, references to Lemmas 5.5 and 5.7 are replaced to their counterparts Lemmas 5.14 and 5.16. The latter Lemmas have weaker conditions than the previous. Thus the verification of these conditions can just be dropped from the proof. □

The proofs of Proposition 5.18 and Theorem 5.19 are obtained from the proofs of Proposition 5.9 and Theorem 5.11 by changing the translations. Thus the well-founded model of a general logic program P is also subsumed by the alternative cautious model. For instance, given the general logic program $P_1 = \{p \leftarrow\sim p; q \leftarrow p\}$, $\sim q$ is included in $\mathrm{CM}'(P_1)$. The details are found in Example 5.20. The well-founded model of P_1 is empty as indicated by Example 5.1.

Proposition 5.18 *Let P be a general logic program. Then* $\mathrm{Tr}'_P(\mathrm{WFM}(P))$ *is contained in the least* $\mathbf{N}$*-grounded* $\mathrm{Tr}'(P)$*-full set.*

Theorem 5.19 *Given a general logic program P,* $\mathrm{WFM}(P) \subseteq \mathrm{CM}'(P)$.

Example 5.20 Let P be the general logic program $\{p \leftarrow\sim p; q \leftarrow p\}$. Its alternative translation into autoepistemic sentences is $\Sigma = \{L\neg Lp \rightarrow p, Lp \rightarrow q\}$. Both p and $\neg Lp$ are refuted by the enumeration $\varepsilon = p, \neg Lp$ as $\mathrm{B}^{\varepsilon}(\Sigma) = \{\neg Lp, L\neg Lp\}$. This enumeration does not refute q from Σ. Neither does the enumeration $\varepsilon_2 = \neg Lp, p$. Therefore $\langle \Sigma, \emptyset\rangle \dashv_{\mathrm{ae}} q$ holds, as $\emptyset$ is the least $\mathbf{N}$-grounded Σ-full set.

The proposed cautious models are not identical. If the former translation of P_1 is used, c remains undetermined. This might suggest that the alternative cautious model would be more accurate. However, the roles change if we consider the cautious models of the general logic program $P_2 = \{p \leftarrow\sim q; q \leftarrow\sim p; r \leftarrow\sim p, \sim q\}$. In the former cautious model $\sim r$ is included, but not in the alternative cautious model.

6 Conclusions

This work focuses on analyzing structural properties and computational complexity of cautious autoepistemic reasoning. First of all, various properties of cautious autoepistemic reasoning are investigated. The basis of the analysis was laid in [15]. The main results presented there are reviewed in Section 2. In addition, we demonstrate how cautious autoepistemic logic provides a tool for analyzing applications of NMR. This study leads to introduction of a new semantic model for general logic programs.

A proof theory for cautious autoepistemic reasoning is developed in Section 3. The generalized L-proofs characterize reasoning with respect to the least cautious expansion as a step-by-step process. At each step either a positive introspection or a negative introspection decision is made. These decisions depend only on the premises and the previous introspection decisions. This kind of proof-theoretical characterization is exceptional within nonmonotonic logics. In Section 4, GL-proofs prove useful when a decision procedure for **GCAE** is devised.

The cumulativity of a reasoning system requires the system to satisfy the rules cut and cautious monotonicity. It is shown that cautious autoepistemic reasoning (the relation $\vdash_{\mathrm{C}}$) satisfies the cut rule. However, the rule of cautious monotonicity is satisfied only in restricted cases. Here the role of GL-proofs becomes important. The results indicate that the cautious monotonicity is satisfied if GL-proofs or propositional consequences of the premises are inserted to the premises. Our analysis concentrates on the relation $\vdash_{\mathrm{C}}$ but similar properties can be proven for the relation $\dashv_{\mathrm{C}}$. It remains open whether cautious autoepistemic reasoning is cumulative for subclasses of autoepistemic theories such as translations of general logic programs into autoepistemic theories.

The complexity analysis in Section 4 shows that the negative introspection principle of CAEL corresponds to a Π_2^{p}-complete problem. If we consider the major reasoning task **GCAE**, this gives a lower limit for its complexity (in fact we could devise a Π_2^{p}-hardness proof for **GCAE** from the Π_2^{p}-hardness proof of **IR**). An upper limit for the computational complexity of **GCAE** is obtained as we prove that **GCAE** belongs to Δ_3^{p}. Testing whether a sentence belongs to the intersection of stable expansions is Π_2^{p}-complete [26]. Recall that Δ_3^{p} is the class of sets that are polynomial time Turing-reducible to Σ_2^{p} (or alternatively Π_2^{p}). Hence the computational complexity of **GCAE** is only slightly greater than that of a similar problem in the original autoepistemic logic. This is good news from the point of view that CAEL is a generalization of AEL. However, the exact computational complexity of **GCAE** relative to the polynomial time hierarchy remains open.

In Section 5 we provide a new formal model for general logic programs in terms of cautious autoepistemic logic. It is proved that the cautious model of a general logic program P is at least as accurate as the well-founded model of P. Furthermore, we provide examples of general logic programs whose cautious models are more accurate than their well-founded models. The cost of these differences can be seen in the computational complexity of the required reasoning. By [32] the membership of a ground atom in the well-founded model can be determined in polynomial time. Because **GCAE** is on the second level of PH, it is evident that **GCAE** remains NP-hard even though we reduce the computational complexity of the underlying monotonic reasoning (e.g. by restricting to translations of general logic programs).

Further research

The use of cautious autoepistemic logic for approximating NMR seems promising. However, the negative introspection principle in cautious autoepistemic reasoning is rather strong as indicated by the complexity results. It seems that there is a tradeoff between the strength of negative introspection and the computational complexity of the corresponding reasoning. The study of this relationship is an interesting topic for further research.

Acknowledgements

This report contains results of my work in Digital Systems Laboratory at Helsinki University of Technology from February 1993 to November 1994. I am very grateful to professor Leo Ojala for arranging the opportunity to work in his laboratory. I am greatly indebted to Dr.Tech. Ilkka Niemelä for his advice. Ilkka's experience and insight into the research field have been indispensable. I also thank my colleague, Lic.Tech. Jussi Rintanen, for his help in implementing cautious autoepistemic reasoning. All my colleagues have contributed to the creative atmosphere of the laboratory.

This work has been funded by Academy of Finland. The financial support from Alfred Kordelin Foundation and Emil Aaltonen Foundation is also acknowledged with gratitude.

Otaniemi, December 1994

Tomi Janhunen

References

[1] Apt, K., Blair, H., and Walker, A.: Towards a theory of declarative knowledge. In Minker, J., editor, *Foundations of Deductive Databases and Logic Programming*, pages 89–148. Morgan Kaufmann Publishers, Los Altos, 1988.

[2] Balcázar, J., Díaz, I., and Gabarró, J.: *Structural Complexity I.* Springer-Verlag, Berlin, 1988.

[3] Brewka, G.: Cumulative default logic: in defense of nonmonotonic inference rules. *Artificial Intelligence*, pages 183–205, 1991.

[4] Clark, K.: Negation as failure. In Gallaire, H. and Minker, J., editors, *Logic and Data Bases*, pages 293–322. Plenum Press, New York, 1978.

[5] Dix, J.: Classifying semantics of logic programs. *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 160–180, 1991.

[6] Doyle, J.: A truth maintenance system. *Artificial Intelligence*, 12:231–272, 1979.

[7] Eiter, T. and Gottlob, G.: 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.

[8] Eiter, T. and Gottlob, G.: Propositional circumscription and extended closed world reasoning are Π_2^p-complete. *Theoretical Computer Science*, 114:231–245, 1993.

[9] Elkan, C.: A rational reconstruction of nonmonotonic truth maintenance systems. *Artificial Intelligence*, 43:219–234, 1990.

[10] Gabbay, D., Hogger, C., and Robinson, J., editors. *Nonmonotonic Reasoning and Uncertain Reasoning*, volume 3. Clarendon Press, Oxford, 1994.

[11] Gelfond, M.: 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.

[12] Gelfond, M. and Lifschitz, V.: 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.

[13] Gottlob, G.: 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.

[14] Gottlob, G.: Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425, June 1992.

[15] Janhunen, T.: Weakened negative introspection in autoepistemic reasoning. Series A: Research reports 21, Helsinki University of Technology, Digital Systems Laboratory, February 1993.

[16] Janhunen, T. and Niemelä, I.: A scheme for weakened negative introspection in autoepistemic reasoning. In *Proceedings of the 3rd Kurt Gödel Colloquium on Computational Logic and Proof Theory*, pages 211–222, Brno, Czech Republic, August 1993. Springer-Verlag.

[17] Johnson, D.: A catalog of complexity classes. In van Leeuwen, J., editor, *Handbook of Theoretical Computer Science*, volume A. Algorithms and Complexity, pages 67–161. Elsevier, 1990.

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

[19] Lloyd, J.: *Foundations of Logic Programming*. Springer-Verlag, Berlin, 1984.

[20] Marek, W. and Truszczyński, M.: Modal logic for default reasoning. *Annals of Mathematics and Artificial Intelligence*, 1:275–302, 1990.

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

[22] Moore, R.: 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.

[23] Moore, R.: 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.

[24] Niemelä, I.: 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.

[25] Niemelä, I.: 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.

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

[27] Niemelä, I. and Rintanen, J.: 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.

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

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

[30] Touretzky, D., Horty, J., and Thomason, R.: 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.

[31] Truszczyński, M.: 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.

[32] Van Gelder, A., Ross, K., and Schlipf, J.: The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, July 1991.