HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A:** Research Reports ISSN 0783–5396
No. **28**; December 1993 ISBN 951-22-1915-8

PRIORITIES AND NONMONOTONIC REASONING

Jussi Rintanen

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 — ISSN 0783–5396
No. **28**; December 1993 — ISBN 951-22-1915-8

Priorities and Nonmonotonic Reasoning

JUSSI RINTANEN

Abstract: Many knowledge-based systems use some kind of priorities for controlling their reasoning. In this work, autoepistemic logic is extended with priorities. Priorities are formalized as partial orders of formulae that express that some formulae are preferred to others. Priorities increase the set of conclusions that can be inferred from a set of formulae, because the number of cases where conflicts between defaults are left unresolved decreases. A decision procedure for prioritized autoepistemic logic is given in terms of a decision procedure for a classical logic on which the prioritized autoepistemic logic is based. The complexity of the logic is analyzed with respect to the underlying classical logic. When the classical logic is the propositional logic, the decision problems of prioritized autoepistemic logic are located on the third level of the polynomial hierarchy. The results of this work carry to other formalisms as well because of the existence of translations of other formalisms to autoepistemic logic. Possible application areas are e.g. default logic, logic programs, deductive databases, and theories of diagnosis. It is shown that prioritized autoepistemic logic is in the propositional case more general than other related formalisms like preferred subtheories of Brewka, structured theories of Ryan, and a propositional version of Lifschitz's prioritized circumscription. Simple polynomial-time translations of these formalisms to prioritized autoepistemic logic are given. Finally, it is demonstrated how automated theorem-proving in prioritized autoepistemic logic benefits from priorities. An algorithm for automated reasoning is given and proved correct.
Keywords: Nonmonotonic logics, preferences, priorities, automated reasoning.

Printing: TKK Monistamo; Otaniemi 1993

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

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

Contents

List of symbols and notations

$\times$	cartesian product of sets
$\cup$	union of sets
$\cap$	intersection of sets
$\setminus$	difference of sets
$\in$	set membership
$\subseteq$	set inclusion
$\subset$	proper set inclusion
2^S	the power set, i.e. set of all subsets of S
$\sum_{i=a}^{b} E(i)$	the sum $E(a) + E(a+1) + \cdots + E(b-1) + E(b)$
$\lfloor r \rfloor$	the greatest integer less than the real number r
$n!$	the factorial $1 \cdot 2 \cdot \ldots \cdot n$ of n
$\left(\begin{array}{c} n \\ m \end{array} \right)$	binomial coefficient, i.e. $\frac{n!}{m!(n-m)!}$
$\mathcal{O}(P)$	upper bound of complexity
$\leftrightarrow$	equivalence
$\rightarrow$	implication
$\wedge$	conjunction
$\vee$	disjunction
$\neg$	negation
$\perp$	constant *false*
$\top$	constant *true*
$\forall$	universal quantification
$\exists$	existential quantification
$\models$	logical consequence, truth in a model
L	the monadic *"is believed"* operator (autoepistemic logic)
$Sf^L(\phi)$	subformulae of ϕ beginning with L
$Sf^{qL}(\phi)$	quasisubformulae of ϕ beginning with L
$SE_{\Sigma}(\Lambda)$	stable expansion of Σ determined by Λ
$\models_L$	L-consequence relation

1 Introduction

This work investigates the expression of preferences between beliefs and reasoning according to them. Preferences – whatever they concern – are important when there is a choice to be made. For example, whether to choose product A or product B in a grocery store, whether to obey the orders of some authority C or conflicting orders of another authority D. Preferences about beliefs are also important. If one has some reason to maintain both of two mutually contradictory beliefs, a choice has to be made between them. In computer science the problem of such conflicts shows up in software systems that support the expression of complex data and knowledge.

Two statements can be contradictory if at least one of them has been acquired by some invalid means. Ordinary deductive inference produces true statements from true statements. In complex application areas the knowledge that is available is usually incomplete and it is necessary to resort to default inference, i.e. to inferring a statement because it seems likely on the basis of other statements known. This kind of inference is not of a deductive nature. Default inference shows up in extensions of current data base management systems and in systems developed by the artificial intelligence community. Examples of principles of default inference are negation as unprovability or failure (forms of the closed world assumption) and inheritance reasoning. Default information in databases differs from non-default information in its defeasibility: default information may be overridden if contradicted by non-default information. A problem arises when two defaults are mutually contradictory and neither one is overridden by a fact. Should both of the defaults be ignored, or should one of the defaults override the other? In many cases the problem of contradictory defaults can be avoided by restricting the expressivity of default rules. Stratification in logic programs and deductive databases that have a negation operator is such a restriction. Similarly, by restricting to single-inheritance (as opposed to multiple-inheritance) in object-oriented databases and programming languages, problematic inheritance conflicts are avoided.

There seems to be a reason for studying conflicting defaults from an abstract point of view, in separation from details of actual systems that implement default reasoning. For this investigation the natural theoretical framework is that of nonmonotonic logics. We present an extension of a particular nonmonotonic logic, autoepistemic logic [Moore, 1985], in which conflicts between defaults are resolved using priorities. Priorities are a form of information distinct from the formulae of the logic. They are expressed as partial orders of formulae. When two defaults are in conflict one of the defaults is chosen according to the priorities. Conflict resolution needed in default reasoning has been formalized earlier by other researchers, for example in [Konolige, 1989;

Brewka, 1989; Ryan, 1991; Lifschitz, 1985]. We hope to demonstrate that our logic is more general and uniform, and that our notion of priorities is realistic in having a natural computational interpretation.

1.1 Representation of knowledge and data

Typical systems for representing knowledge and data are relational databases and their extensions like deductive databases [Kanellakis, 1990], object-oriented databases, and from the area of knowledge-engineering and artificial intelligence, rule systems, frame systems and semantic networks [Kramer and Mylopoulos, 1987]. All these systems provide facilities for representing facts and making queries about them.

Deductive databases extend relational databases with more expressive query languages based on simple deductive rules that are similar to the program clauses in logic programming languages like Prolog. Datalog – the most common query language for deductive databases – can be considered a subset Prolog, and hence a subset of Horn clause logic. The transitive closure of a relation cannot be expressed in relational algebra, but it is easily expressible in Datalog by the following rules [1].

t(X,Y) ← r(X,Y)
t(X,Z) ← r(X,Y), t(Y,Z)

However, Datalog without a negation operator cannot express the difference of relations which in turn is easily expressible in relational algebra. The language Datalog$^{(\neg)}$ extends Datalog with the negation operator *not*. The difference of relations $d = s - r$ can then be expressed simply as follows.

d(X) ← s(X), not(r(X))

The introduction of the *not* operator makes the semantic definition of Datalog$^{(\neg)}$ significantly more complicated. For example, given the following rules and the fact

p(X) ← not(q(X)), r(X)
q(X) ← not(p(X)), r(X)
r(a)

[1] We use the notational conventions of Prolog: relation (predicate) symbols and constants are in lower case, variables in upper case. Prolog uses :- instead of the implication sign ←.

it is not clear how to answer the queries $p(a)$ and $q(a)$. Which one of them should be answered "yes"? The answers to other queries may depend on this choice. Hence by case analysis, inferring one of $p(a)$ and $q(a)$, two separate sets of answers are obtained. The answers for Datalog$^{(\neg)}$ queries are unambiguous for the subset of *stratified* sets of rules in which cyclic use of *not* is disallowed by requiring that the rule set can be partitioned into a number of strata so that *not* refers only to relations the definitions of which are in earlier strata. Essentially, the stratification condition requires that there are no conflicting defaults.

Relational databases even with expressive deductive query languages like stratified Datalog$^{(\neg)}$ are insufficient for many representational purposes. In knowledge-based applications there is a demand for an expressivity going beyond Horn clause logic, e.g. unrestricted use of disjunctions and less restricted use of quantifiers $\forall$ and $\exists$ would be desirable. Furthermore, in many application areas conflicting defaults are inherently present: the knowledge-bases cannot be divided into strata as required by the stratification condition. This kind of problems have been faced by artificial intelligence researchers who have tried to represent and process computationally several kinds of complex knowledge. Production rules, frame systems and semantic networks have provided some of the capabilities needed. The theory of relational and deductive database systems is based on relational algebra, relational calculus, and various extensions of first-order logic. Similarly there is a need for providing an exact foundation for more complex systems of knowledge representation. Nonmonotonic logics are one such foundation. Unlike first-order predicate logic and other monotonic logics, nonmonotonic logics properly support the expression of default rules and the *closed world assumption* (CWA) "assume false if not known true", both of which are very frequently used in representing knowledge.

1.2 Priorities and preferences in knowledge-based systems

In this section we mention some natural situations where priorities are used. Not all uses of priorities are immediately recognizable as resolution of conflicts caused by defaults, but may be formalized as such. In databases, some parts of the data are by its nature more permanent than others. For example the integrity constraints of a database are of the highest priority: removing an integrity constraint just to preserve the consistency of a database during an update is forbidden. Updating a view of a database poses a problem as the update operation itself does not uniquely determine how it should be realized [Fagin *et al.*, 1983]. A view can be seen as a special case of a deductive rule. Consider the following database.

p(a)
q(a,b)
r(X,Y) ← p(X), q(X,Y)

If a user now requests the removal of the fact $r(a, b)$ from the database, the request does not uniquely determine what operation to apply because there are several ones that lead to the desired result: removal of either of $p(a)$ and $q(a, b)$ causes the fact $r(a, b)$ disappear. Fagin et al. [1983] suggest that the uniqueness of the actual update operation can be achieved by associating numeric (or other kind) priorities to the data: among alternative update operations choose the one that retains as many of the high-priority formulae as possible. Database updates are a simple special case of the problem of updating arbitrary sets of e.g. first-order formulae. This problem is considered in the research of belief revision [Gärdenfors, 1988; Nebel, 1991].

In diagnostic reasoning, for example in expert systems for medical diagnosis or diagnosing faults in electronic equipment, the possible causes of a fault can be ordered on the basis how likely they are. When such priority information is available, it is often sensible to compute only the most likely diagnoses. The computation of this subset of diagnoses may be more efficient than the computation of all diagnoses. Ideally this kind of priority information would be based on known probabilities of possible faults and the dependencies between them, and the diagnoses computed would be assigned probabilities accordingly. This is not possible in general because exact probabilities of faults are often difficult or impossible to obtain. Next we give a simple example to illustrate the use of priorities in diagnostic reasoning. The diagnostic problem is formalized as a set of formulae describing the causal relations between the parts of the system under consideration and a set of formulae describing the observations about the state of the system [Reiter, 1987]. The diagnostic problem is to compute minimal subsets of components that have to be assumed to be working incorrectly. In general, there can be several different such sets (diagnoses). If some of the faults are more likely to occur than others, the computation of some of the more likely ones could suffice. Our example diagnostic problem is about cars that do not start.

Observations: $\neg m, r$
Components: b, s, g

$b \leftrightarrow r$
$(b \wedge s \wedge g) \leftrightarrow m$

The atomic formulae can be read as follows.

m = "motor starts", r = "radio and cd-player work",
b = "battery ok", s = "starting motor ok", g = "tank not empty".

With the given observations, not all components can be assumed to work correctly because the observation $\neg m$ and the hypotheses b, s, g together with $b \wedge s \wedge g \leftrightarrow m$ are inconsistent. The minimal subsets Δ of $H = \{\neg b, \neg s, \neg g\}$ which are consistent with $\{\phi | \neg\phi \in H \backslash \Delta\}$, the observations, and the system description are $\{\neg g\}$ and $\{\neg s\}$. Diagnoses containing $\neg b$ cannot be accepted because an inconsistency would be obtained from $b \leftrightarrow r$, $\neg b \in \Delta$ and the observation r. The two diagnoses $\{\neg s\}$ and $\{\neg g\}$ are equally acceptable. Priorities could be applied to select between these two, if it was known that the failure of one of s or g is much more likely than the failure of the other.

Inheritance is used in programming languages, object-oriented databases, and knowledge-based systems to make the use and maintenance of large knowledge bases more efficient. The basic form of inheritance, single-inheritance, stores data in a tree with classes as nodes and instances of the classes as leaves. When inheriting a value of an attribute for an instance, the value lowest in the tree is used. We can say that the attribute values attached to a subclass have a higher priority than a value attached to a superclass. Single inheritance is easy because the conflicts between attribute values can always be solved simply by choosing information in subclasses instead of superclasses. In multiple-inheritance, where the class hierarchy is not a tree any more but a directed acyclic graph, inheritance is not as easy because there is the possibility that an attribute value can be inherited from two classes such that neither of them is a subclass of the other. In practical systems this kind of conflicts are often solved by ordering the superclasses of each class to some total order, and in conflict situations inheriting an attribute value from the first one in the ordering. One of the most popular examples of multiple-inheritance in literature is the Nixon diamond (named after the former U.S. president Richard M. Nixon) in Figure 1 [2]. Nixon is both a republican and a quaker. Republicans typically are not pacifists while quakers typically are. Is Nixon a pacifist or not? Again, the conflict between *quaker* and *republican* can be resolved by prioritizing one over the other. Many practical systems of inheritance reasoning work this way. The Nixon diamond can be expressed in Prolog or Datalog$^{(\neg)}$ as follows.

quaker(nixon)
republican(nixon)
pacifist(X) $\leftarrow$ quaker(X), not(nonpacifist(X))
nonpacifist(X) $\leftarrow$ republican(X), not(pacifist(X))

[2] Having value *true* for attribute *pacifist* is identified with the membership in the class Pacifist.

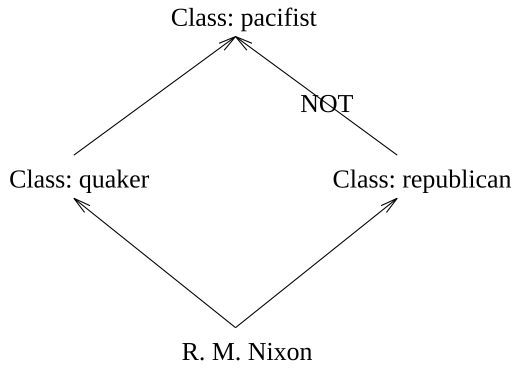


Figure 1: The Nixon diamond

This set is not stratified because of the cycle involving *pacifist*, *nonpacifist* and the *not* operator.

1.3 Priorities and preferences in nonmonotonic logics

Nonmonotonic logics are formalisms that were designed for expressing defeasible reasoning. Reasoning based on assumptions is difficult to express in classical logics (like propositional logic or predicate logic), because the conclusions of a set of formula monotonically increase: further information that cancels an assumption, cannot retract the conclusions from the assumption. The logical consequence relations $\models$ of classical logics obey the monotonicity property: $\Sigma \models \phi, \Sigma \subseteq \Delta$ imply $\Delta \models \phi$. In nonmonotonic logics old conclusions can be defeated on the arrival of new information.

The leading formalisms of nonmonotonic reasoning are default logic [Reiter, 1980], autoepistemic logic [Moore, 1985], and circumscription [McCarthy, 1980]. Autoepistemic logic extends the language of a classical logic (e.g. propositional logic) with the modal operator L. This operator can be intuitively read as "is believed". Default logic introduces a separate syntactic category of default rules $\frac{\alpha:\beta}{\gamma}$. A formula γ can be derived using such a rule if the prerequisite α is derivable and the negation $\neg\beta$ of the justification β is not derivable. Because the derivability of formulae may depend on the derivability and underivability of other formulae in complex ways, the sets of consequences of a set of formulae in these logics are defined as fixpoints of a nonmonotonic operator instead of simply as the deductive closure of some inference rules (like in classical logics). Circumscription defines the logical consequences of a theory as formulae that are true in all models of the theory that are minimal in a specific sense.

In autoepistemic logic the set of consequences Δ of a set of formulae Σ is defined using the fixpoint equation

$$\Delta = \{\Sigma \cup \{L\phi | \phi \in \Delta\} \cup \{\neg L\phi | \phi \notin \Delta\} \models \phi\}.$$

This equation can have any number of solutions depending on the set Σ. This is in contrast with ordinary monotonic logics, where each set of formulae has exactly one set of consequences. Similarly default logic may sanction several extensions. Consider the following default theory [3].

$$p \qquad\qquad p \rightarrow b \qquad\qquad \frac{p : \neg f}{\neg f} \qquad\qquad \frac{b : f}{f}$$

This theory has two extensions. The first is $\{\phi | \{p, b, \neg f\} \models \phi\}$, and the second is $\{\phi | \{p, b, f\} \models \phi\}$. When formalizing knowledge using these logics, sometimes only a subset of all extensions are desired. The disambiguation information to single out a subset of the extensions can be brought to these logics in several ways. Brewka [1991] and MacNish [1991] extend default logic with priorities that impose a total ordering on default rules of a default theory. For example, in the example above, the first default rule could be prioritized over the second one. Hence, first try to apply the rule $\frac{p : \neg f}{\neg f}$. The prerequisite p is derivable, and the negation of $\neg f$ is not derivable. Hence conclude $\neg f$. Then try to apply the second rule. Because the negation of the justification f is derivable, the rule cannot be applied. The resulting extension is therefore $\{\phi | \{p, b, \neg f\} \models \phi\}$. As we will demonstrate in Section 4.2.1, in general, not all orderings of default rules determine an extension in this simplistic manner.

Circumscription [McCarthy, 1980] is based on the idea that when representing knowledge as formulae of logic, e.g. classical predicate logic, not all models of the formulae are intended ones. In the intended models the extents of some of the predicates are as small as possible. For example, when a knowledge-engineer writes the knowledge-base

$$\begin{array}{lll} human(alice) & human(peter) & human(joe) \\ doctor(alice) & & \end{array}$$

he means that Alice is a doctor, but Peter and Joe are not. It is exceptional that someone is a doctor, and it is assumed that no-one is unless otherwise stated. The intended models are the ones in which the extents of predicates are as small as possible. The exact meaning of the term "as small as possible"

[3] In this example it is useful to read the proposition p as "X is a penguin", the proposition b as "X is a bird", and f as "X flies".

$$\begin{array}{l} penguin(Tweety) \\ \forall x\ penguin(x) \wedge \rightarrow bird(x) \\ \forall x\ penguin(x) \wedge \neg ab_{penguinfly}(x) \rightarrow \neg flies(x) \\ \forall x\ bird(x) \wedge \neg ab_{birdfly}(x) \rightarrow flies(x) \end{array}$$

$$\begin{array}{rcl} penguin & = & \{Tweety\} \\ bird & = & \{Tweety\} \\ flies & = & \{Tweety\} \\ ab_{birdflies} & = & \{\} \\ ab_{penguinflies} & = & \{Tweety\} \end{array} \qquad \begin{array}{rcl} penguin & = & \{Tweety\} \\ bird & = & \{Tweety\} \\ flies & = & \{\} \\ ab_{birdflies} & = & \{Tweety\} \\ ab_{penguinflies} & = & \{\} \end{array}$$

Figure 2: A knowledge base and its two minimal models

depends on the version of circumscription. Consider again the inheritance example, for which a predicate logic formalization is shown in Figure 2. The knowledge base says that "Tweety is a penguin", "all penguins are birds", "all penguins, unless they are somehow abnormal, are unable to fly", and "all birds, unless they are somehow abnormal, can fly". The predicates $ab_{birdfly}$ and $ab_{penguinfly}$ indicate what individuals are exceptional in some respect. Because exceptionality is to be minimized, the extents of these predicates are in the intended models as small as possible. The models of the set of formulae in which the extents of these predicates are set-inclusion minimal (i.e., there are no models in which the extents of all predicates would be subset) are shown in Figure 2. The first of the models in Figure 2 is not intuitively correct. Because Tweety is a penguin and penguins do not fly, Tweety does not fly. The fact that birds fly should be given a lower priority than the fact that penguins do not fly. In prioritized circumscription [McCarthy, 1986; Lifschitz, 1985] predicates can be minimized so that the minimality of some predicates is given a priority over other predicates. Prioritized circumscription, if $ab_{penguinfly}$ is given a priority higher than $ab_{birdfly}$, would accept as minimal only the second of the models in Figure 2.

1.4 The contributions of the work

In this work, autoepistemic logic is extended with priorities and the properties of the resulting formalism are investigated. We have chosen autoepistemic logic as the basis of our work because many other nonmonotonic formalisms can be seen as special cases of it. For example the default logic of Reiter can be translated to autoepistemic logic using a simple translation. This generality of autoepistemic logic can be utilized in bringing our results concerning prioritized autoepistemic logic to corresponding prioritized versions of other nonmonotonic formalisms. Prioritized autoepistemic logic is seen as a general framework for formalizing nonmonotonic reasoning with priorities. Compared to other work on the subject there are several points in which our logic differs from other similar formalisms.

- Our priorities are independent of the rules and formulae expressing the defaults. Hence they can be given a semantics in terms of the stable expansions in addition to a characterization based on more proof-theoretic terms. Both kinds of characterizations are given: on one hand we define what are the most preferred stable expansions among all expansions given a prioritization, and on the other we point out the proof-theoretic uses of the priorities in the computation of stable expansions.

- Our prioritized autoepistemic logic is (at least in the propositional case) more general than the logics and systems presented in the literature earlier, e.g. preferred subtheories of Brewka and propositional prioritized circumscription. These inclusions are based on a simple polynomial time translations. Konolige's hierarchic autoepistemic logic and multi-agent autoepistemic logic of Toyama et al. can be embedded in ordinary autoepistemic logic, which is a special case of prioritized autoepistemic logic. The maximality notion of Ryan's structured theories [Ryan, 1991] differs from our preferredness notion slightly. If this difference is ignored, Ryan's system can be embedded in ours. The translation however is not polynomial time, because we have to compute a subset of the so-called *natural consequences* for each formula.

- Our priorities have in many cases a favourable effect on the efficiency of decision procedures. We give a theorem-proving algorithm for prioritized autoepistemic logic. For an important subclass of prioritizations, the more priorities are available, the more the procedure can prune its search tree. In restricted cases we can have asymptotically more efficient theorem-proving than without priorities (Section 3).

  Automated nonmonotonic reasoning with priorities has been earlier investigated for example in [Baker and Ginsberg, 1989; Junker, 1991; Junker and Brewka, 1991]. Junker and Brewka devise algorithms for Brewka's preferred subtheories [1989] using truth-maintenance systems. In this approach, non-layered partial orders pose problems that are however partially solved by Junker and Brewka [1991]. They introduce an ordering check in their TMS-based approach using which the computation of one preferred subtheory is straightforward. Our procedures work without modification for any strict partial order, and we can compute all preferred stable expansions – not just the first one – uniformly.

1.5 Outline of the work

In Section 2 prioritized autoepistemic logic is formally defined, its complexity and other properties are analyzed. In Section 3 the possibility of achieving tractable, i.e. polynomial time reasoning using priorities is investigated. Restricted intractable and tractable classes are identified and directions to possible tractable classes are given. Section 4 compares prioritized autoepistemic logic to other formalizations of priorities within nonmonotonic formalisms. Section 5 presents an alternative notion of most preferred expansions, analyzes its properties and compares to the notion presented first. In Section 6 we present a theorem-proving algorithm for a subclass of prioritized autoepistemic logic. We also briefly discuss an implementation of an automatic theorem-prover for several non-monotonic logics. Section 7 applies prioritized autoepistemic logic to inheritance hierarchies and diagnosis. Representations of theories of inheritance are embedded in prioritized autoepistemic logic. Priorities can be used for expressing notions of specificity as well as resolving ambiguities caused by multiple inheritance. Section 8 concludes the work by briefly evaluating its results and pointing directions for further research.

2 Prioritized autoepistemic logic

We take a declarative approach to priorities. That is, our notion of priorities is not a description of a mechanism how reasoning with priorities is performed. This is in contrast with most implementations of priorities in knowledge-based systems and also with some approaches within the research area of nonmonotonic logics. Our thesis is that priorities should have a clear declarative meaning. Priorities in systems of default reasoning that are of a probabilistic nature, would have a probabilistic interpretation. Default rules expressed in autoepistemic logic and other nonmonotonic logics based on the notion of consistency, do not immediately lend to a probabilistic interpretation. In such systems priorities could be interpreted as some kind of ordering on formulae or conclusions. In addition questions concerning their meaning also the implementation aspects of priorities, for instance the algorithmic treatment of them and their computational complexity, are important. However, we think that priority notions with no declarative meaning at all or with an unsatisfactory declarative meaning – even if they are efficient – are not very useful. In Section 4.2 we discuss approaches to priorities within default logic. Useful default priorities are easy to define when the default rules are of sufficiently simple form, e.g. prerequisite-free and normal. Using priorities with more complex default rules is more difficult. It seems that in the general case, construction methods for default extensions that are local in the sense that at each construction step the next applicable default rule to be applied is deterministically chosen on the basis of priorities and the formulae currently derivable, are likely to produce unintuitive results (note that the possibility of reasoning in such simplistic manner would imply that prioritized reasoning in default logic is tractable whenever the underlying classical logic has a polynomial time decision procedure). As an example of such unintuitive behaviour, consider the default theory

$$a, b, \frac{c : d}{d}, \frac{b : \neg d}{\neg d}, \frac{a : c}{c}.$$

Let the three rules be prioritized in the order indicated, i.e. the rule $\frac{c : d}{d}$ has the highest priority. Priorities can be used in reasoning in default logic as follows: always apply the highest priority rule among all applicable rules that have not been applied before. With the above default theory, the second and third default rules are both initially applicable. Hence apply the second one and obtain $\neg d$. Now only the third rule is applicable and c is obtained. The unintuitive thing is that whatever other default rules are applied, the lowest priority rule can always be applied, and hence the prerequisite of the highest priority rule $\frac{c : d}{d}$ belongs to all extensions of the theory. Despite this fact, the highest priority rule does not become applicable because of the rule $\frac{b:\neg d}{\neg d}$. It seems that this kind of procedural uses of priorities do not implement the principle "apply the highest-priority default if possible" properly. The priorities should choose the extensions in which the highest priority rules have been applied.

2.1 Design rationale

The goal of this research is the construction of a version of autoepistemic logic with a useful and general mechanism for expressing priorities. The notion of stable expansions – the sets of consequences – in autoepistemic logic is perfectly satisfying from our point of view, and it seems that all that is needed is to be able to choose using priorities a subset of all stable expansions of a set of formulae, the intended ones. There does not seem to be any reason why the priority mechanism should not be very loosely coupled from the mechanism of defining the stable expansions. Loose coupling also implies that the same priority mechanism can be used with minimum effort in conjunction with different versions of autoepistemic logic. Autoepistemic logic describes the reasoning of an ideal agent that is logically omniscient, that is, it believes all logical consequences of its beliefs, and it is completely aware of its beliefs and disbeliefs. The stable expansions are identified with the states of belief of the agent. Because in autoepistemic logic for a given set of formulae there can in general be several stable expansions, the most natural decision problem for it is the membership of a formula in the intersection of all stable expansions. As has been suggested in the introduction, not all stable expansions are intended ones. To reject some of the unintended ones we propose an abstract formalization of the internal criteria on which an agent bases its choice of its beliefs. When more criteria are available, the set of possible states of belief is smaller, and hence their intersection is larger and more formulae can be inferred.

An immediate formalization of preferences between states of belief are as partial orders with sets of beliefs as elements. The maximal elements of these orderings are the most preferred sets of beliefs, and the agent adopts one of them. Giving a preference relation directly on belief states or stable expansions is not possible nor desirable in general. The set of all stable expansions is not necessarily known and its computation may be expensive. When the set of premises is modified a completely new preference relation has to be constructed because the set of stable expansions is not the same as before. And last but not least, if it is known what the minimal – the desired – stable expansions are, why bother with the priorities.

From a computational point of view it is more economic to devise a representation for the criteria on which preference relations between stable expansions are based. It would be useful if the representation could be computationally easy and at least in some cases finite. In the following we consider some alternatives for methods using which the most preferable stable expansions are determined. They are based on the idea that for each set of formulae on which an autoepistemic agent bases its beliefs there is a "basis", some set of formulae Φ, using which we can differentiate between the possible states of belief of the agent. It would be useful if there were an injection (a partial function) $f : 2^{\Phi} \rightarrow E$ from subsets of the basis to the set of stable expansions E so that each stable expansion could be determined uniquely by its intersection with the basis. This however is not necessary.

Given a set of formulae Φ, each stable expansion E could be associated with a number $c(E) = |\Phi \cap E|$. The most preferable stable expansions would be the ones for which the number $c(E)$ is the smallest (or largest). In other words, stable expansions are compared in terms of the number of formulae in Φ they contain. This definition does not support the expression of differences in the significance of formulae.

Given a set of formulae Φ, the stable expansion E_1 is most preferable if there are no stable expansions E_2 such that $(E_1 \cap \Phi) \supset (E_2 \cap \Phi)$ (or $\subset$). The set of most preferred stable expansions defined in this way is a superset of the most preferred stable expansions according both to the previous and the next definition. Hence this definition is more cautious than either of the others, i.e. it is more reluctant to order a pair of stable expansions than the others.

The first definition could be generalized by associating each formula $\phi \in \Phi$ with a number $f(\phi)$, e.g. a real number in the range $[0, 1]$, that indicates the plausibility of the formula. Then define $c(E) = \sum_{\phi \in \Phi} f(\phi)$ and $E_1 \sqsupset E_2$ if and only if $c(E_1) > c(E_2)$. Different kinds of preference relations are obtained by using other methods of combining the numbers. One possibility is the product $\Pi_{\phi \in \{p_1, \ldots, p_n\}} \phi = p_1 \cdot \ldots \cdot p_n$. Formulae with real-valued weights clearly makes it possible to state that one formula is more important than another, and that two formulae are equally important. However, it is not possible to express the incomparability of two formulae.

Instead of associating a number with each formula, associate an element in an arbitrary ordered set (or equivalently, define an ordering relation on Φ). Formulae lower in the ordering are adopted into the set of beliefs more reluctantly than higher formulae. Instead of using a sum function for ordering expansions, order them *lexicographically*, i.e. compare the membership of the most significant formula in two expansions, and if the expansions differ in this respect, order them accordingly. If they do not differ, consider the next formula in the order of significance. Minimization/maximization that is based on lexicographic orderings is widely used in the research area, e.g. [Lifschitz, 1985; Brewka, 1989; Ryan, 1991]. Although lexicographic orderings are not very flexible (the assumption that a small number of more significant formulae weight more in determining the ordering irrespective of the possibly very great number of less significant formulae) they seem to provide a certain kind of modularity that is useful in reasoning using priorities: when the less significant formulae are irrelevant, the decision procedures do not have to take them into account. Also from the viewpoint of a user of the prioritization mechanism, the amount of complex interactions between parts of the prioritization is reduced if the ordering mechanism is lexicographic.

If agreed that prioritization should be lexicographic, there still are several choices to be made in deciding how the most preferred stable expansions are defined. Either the presence or absence of formulae in Φ in the stable expansions can be preferred. Researchers in autoepistemic logic have identified a number of cases when a set of formulae has stable expansions that are not intuitive because of the presence of improperly justified or ungrounded formulae. Hence it seems sensible to prefer stable expansions with less formulae rather than ones with more formulae. Note that using minimization maximization can still be performed. A simple scheme for maximizing certain formulae $\phi_1, \ldots, \phi_n$ is to include the formulae $\{\neg L \neg \phi_i \rightarrow \phi_i | 1 \leq i \leq n\}$ in the set of formulae that generate stable expansions and then minimize the formulae $\neg \phi_i$.

As discussed above, the most preferred stable expansions can be defined as the minimal elements of a partial ordering on stable expansions, that is defined lexicographically on the basis of an ordering of the formulae in a set Φ. The definition of a lexicographic ordering is clear when the set of formulae Φ is totally ordered and there is a formula for which no more significant formulae exist. In many applications, however, it is difficult or impossible to impose an ordering between all pairs of formulae, and therefore we allow partiality in the prioritization ordering. The question arises, how should this partiality be interpreted. It seems to us that the only sensible alternatives are based on extensions of the partial order to total orders. Stable expansions can be related if the first one is preferred to the second one according to *all* total extensions of the partial order, or alternatively, if it is preferred according to *some* total extension of the partial order. The former alternative naturally leads to preference relations between stable expansions that are strict partial orders, and the latter to preference relations that are preorders. For example Ryan [1991] uses the latter. In Section 5 we shall use the former and show its connection to Ryan's definition.

There is one more possibility in defining the most preferred stable expansions. The lack of priorities represents our ignorance about the preferences used (by a reasoning agent, for example). Similarly the partiality of the partial orders is some kind of ignorance. A total ordering of all relevant formulae would represent complete knowledge about the preferences. In this way, the most preferred sets of conclusions could be defined directly, i.e. without the aid of an ordering relation between sets of conclusions (or stable expansions), as follows. We take a most preferred set of conclusions to be one that is lexicographically most preferred among all sets of conclusions according to some total extension of the partial order. This is roughly the definition we shall adopt. We also show that Brewka's definition in [1989] is similar, though expressed in a procedural way. The difference between our definition and the one in which a partial ordering of stable expansions is used does not seem very significant. Examples that demonstrate their difference are fairly complicated. We discuss the connection between the definitions in Section 5.

2.2 Preference aggregation and Arrow's theorem

Results about the general possibility of useful systems of defeasible reasoning with priorities are presented in [Doyle and Wellman, 1991]. Here we discuss what are the implications – if any – on our design decisions concerning priorities in autoepistemic logic. Doyle and Wellman interpret the results of Arrow [1963; 1967] in the context of constructing preference relations for nonmonotonic reasoning. Arrow's results are from the area of social choice theory and concern the possibility of constructing society's preferences given the preferences of individuals. Preference relations are preorders that indicate the desirability of different alternatives, e.g. candidates in elections or alternative social states.

Our presentation of Arrow's results are mainly based on [Doyle and Wellman, 1991]. A global preference order $\sqsubseteq$ is constructed on the basis of a number of individual preference orders $\sqsubseteq_i$. The process of constructing a global preference order may or may not have the following five properties.

1. *Collective rationality.* The global preorder is a function of the individual orders $\sqsubseteq_i$, which are unrestricted, possibly partial, preorders. That is, if Π denotes the set of all preorders of $\mathcal{M}$, an aggregation policy for criteria indexed by I is a function from Π^I to Π.

2. *Pareto principle (unanimity).* If $M_1 \sqsubset_i M_2$ for some $i \in I$ and for no $j \in I$ does $M_2 \sqsubset_j M_1$, then $M_1 \sqsubset M_2$.

3. *Independence of irrelevant alternatives.* The relation of M_1 and M_2 according to the global order depends only on how the individual orders rank those two candidates.

4. *Nondictatorship.* There is no $i \in I$ such that for every M_1 and M_2, $M_1 \sqsubseteq M_2$ whenever $M_1 \sqsubseteq_i M_2$, regardless of the $\sqsubseteq_j$ for $j \neq i$.

5. *Conflict resolution.* If $M_1 \sqsubseteq_i M_2$ for some $i \in I$, then $M_1 \sqsubseteq M_2$ or $M_2 \sqsubseteq M_1$.

The voting paradox of de Condorcet [Doyle and Wellman, 1991; de Condorcet, 1785] can be seen as an instance of Arrow's results. Condorcet noticed that majority voting is not a proper way of constructing a global preference order on candidates from the individual preference orders. Consider the following three individuals and their preferences.

$$\begin{array}{ccc}
I_1 & I_2 & I_3 \\
 & & \\
A & B & C \\
\uparrow & \uparrow & \uparrow \\
B & C & A \\
\uparrow & \uparrow & \uparrow \\
C & A & B
\end{array}$$

The diagram indicates that both I_1 and I_3 prefer A to B, and both I_1 and I_2 prefer B to C. However, the majority consisting of I_2 and I_3 prefers C to A contrary to what would be supposed on the basis of the majority decisions concerning the pairs A, B and B, C. Hence majority voting does not always produce transitive global preference orders from individual preference orders. Arrow showed *the inexistence* of methods for constructing global preference orders that fulfill properties 1,2,3 and 4 above from individual preference orders. The main result about aggregation is the following one (this way of stating the result is from [Doyle and Wellman, 1991]).

Theorem 2.1 (Arrow) *If there are at least three alternatives which the individuals may order in all possible ways, then no aggregation policy satisfies the properties (1)–(4) above.*

The question arises, whether some of the alternatives for priority mechanisms for autoepistemic logic discussed earlier would seem unfavourable in the light of Arrow's theorem. We proposed, that preference relations on stable expansions would be based on a set of formulae, a "basis". Each formula is used in evaluating the desirability of a stable expansion by examining whether the formulae is a member of that stable expansion or not. This way we can view each formula as an individual i that classifies the stable expansions to *good* and *bad* ones, denote these with sets G_i and B_i. This classification corresponds to an individual preference order $\sqsubseteq_i$ for which $E \sqsubseteq_i E'$ if either both E and E' belong to G_i or B_i, or $E \in B_i$ and $E' \in G_i$. Each of the alternatives for constructing a preference relation discussed earlier can be seen as an aggregation function that maps the individual preference orders (each of which corresponds to a formula) to a global preference order, i.e. a preference relation on stable expansions.

It turns out that Arrow's results are not directly applicable in this setting. The preconditions for applying Theorem 2.1 are only partially satisfied. Often there are more than two stable expansions for a set Σ, but the assumption that the alternatives can be ordered in all possible ways by the individuals is not fulfilled, that is, for candidates c, d and e, if $c \sqsubseteq_i d \sqsubseteq_i e$, then either $\{c, d\} \subseteq B_i$ or $\{d, e\} \subseteq G_i$, i.e. it is not possible that $c \sqsubset_i d \sqsubset_i e$.

As a "counter-example" to Arrow's theorem, consider the preference notion that uses formulae with real-valued weights.

Example 2.1 Let E be a set. Let each individual i divide the set E to two subsets, G_i and B_i such that $G_i \cap B_i = \emptyset$ and $E = G_i \cup B_i$. Let $b \sqsubset_i g$ if $b \in B_i$ and $g \in G_i$ and $c \sqsubseteq_i d$ if $\{c, d\} \subseteq B_i$ or $\{c, d\} \subseteq G_i$.

Let $f : I \rightarrow \Re$ be a function from the individuals to the real numbers ≥ 0. Now define $p \sqsubseteq q$ if and only if

$$\sum_{i \in \{j \in I | p \sqsubset_j q\}} f(i) \leq \sum_{i \in \{j \in I | q \sqsubset_j p\}} f(i)$$

Essentially, the set of criteria I is a multi-set of real numbers, and the candidates C to be ordered are subsets of I. The global preference order orders the candidates according to the sum $\sum_{c \in C} c$. Properties 1–3 are immediate. The inexistence of dictators is shown by the following example.

candidate	1	1	1.5	sum
A	•	•		2
B			•	1.5
C	•			1
D		•		1

There are four candidates A, B, C and D, and e.g. the first individual with weight 1 prefers both A and C over B and D. The first individual is not a dictator because it prefers C to B but the society prefers B to C. Similarly, the second individual prefers D to B but society prefers B to D, and the third individual prefers B to A but the society prefers A to B. ■

Hence in the case where individual preference orders divide the candidates to two layers, there are satisfactory aggregation functions that are not dictatorial. Like in the case with only two candidates, majority voting both with and without weights is an acceptable method of aggregation.

The extension of the preference notions mentioned to the case where each individual (formula) divides the candidates (stable expansions) to three layers fails: in those cases Arrow's theorem indicates that not all properties 1,2,3 and 4 can hold. As an example of such extension, consider preferring stable expansions E with $\phi \in E$ over stable expansions E' with $\phi \notin E', \neg\phi \notin E'$, and these in turn over stable expansions E'' with $\neg\phi \in E''$. Doyle and Wellman suggest that in such cases property 4 should be rejected. Hence in the general case, dictatorial aggregation functions – that correspond to lexicographic ordering mechanisms – seem to be the only serious alternative.

2.3 Definition

In this section we give the formal definition of our prioritized autoepistemic logic. This logic is directly based on autoepistemic logic. For the motivation and background of autoepistemic logic see [Moore, 1985]. Different versions of autoepistemic logic are obtained by basing it on different monotonic logics, like classical propositional logic or predicate logic, modal logics, and so on. Many of our definitions and results are presented without assuming properties of the monotonic logic other than the existence of the usual connectives like $\neg, \rightarrow, \vee$, and the monotonicity of the consequence relation $\models$. The symbols $\top$ and $\bot$ denote propositions that are respectively true and false in all models. The language of the classical logic on which an autoepistemic logic is based on is denoted by $\mathcal{L}$. To obtain the language $\mathcal{L}_{ae}$ of an autoepistemic logic, the language $\mathcal{L}$ is extended with the prefix operator L which is read "is believed". The definition of autoepistemic logic extends the relation $\models$ to cover also formulae beginning with the operator L. This kind of formulae are treated by this relation like propositional variables in classical propositional logic, i.e. their valuation is independent of all other atomic formulae.

Autoepistemic logics describe the reasoning capabilities of ideally rational agents. An agent is logically omniscient, i.e. it believes all logical consequences of its own beliefs, and it is also capable of introspection, i.e. for each proposition ϕ either $L\phi$ or $\neg L\phi$ is its belief depending on whether ϕ is its belief. The belief states of an autoepistemic agent are based on a set of initial premises. Possible belief states Δ of an autoepistemic agent given a set of initial premises Σ are the solutions of the following equation [Moore, 1985].

$$\Delta = \{\phi | \Sigma \cup \{L\phi | \phi \in \Delta\} \cup \{\neg L\phi | \phi \notin \Delta\} \models \phi\} \qquad (1)$$

The solutions Δ of this equation are *stable expansions* of Σ. A set Σ does not in general have a unique stable expansion. One could think that the operator related to Equation 1 would conveniently have a least solution that could be found by iteration (by the famous fixpoint theorem of Tarski). The operator, however, is not monotonic because of its appeal to formulae *not* in the set Δ. In general, the equation can have none, one, or several solutions depending on the set Σ.

Example 2.2 The set $\Sigma = \{\neg L\phi \rightarrow \phi\}$ has no stable expansions. If $\phi \notin \Delta$, then by the fixpoint equation $\neg L\phi \in \Delta$, and as a logical consequence of $\neg L\phi$ and Σ should be $\phi \in \Delta$, which is a contradiction. Suppose then that $\phi \in \Delta$. This can only be if ϕ is a logical consequence of Σ and some set of formulae $L\chi$ and $\neg L\psi$. This would require $\neg L\phi \in \Delta$ and then $\phi \notin \Delta$, again a contradiction.

The set $\{L\phi \rightarrow \phi\}$ has two stable expansions. There are two solutions Δ to the fixpoint equation, one contains ϕ and the other does not. ∎

Because of the possibility of several stable expansions, the question of what is meant by correct reasoning in autoepistemic logic arises. The notions of *cautious* and *brave* reasoning have been introduced in this context. A formula ϕ can be inferred by cautious reasoning from a set of formulae Σ if and only if ϕ is contained in every stable expansion of Σ. Similarly, a formula ϕ can be inferred by brave reasoning from Σ if and only if ϕ is contained in at least one stable expansion of Σ. Of these two, cautious reasoning is more important, because using it only such formulae are obtained that are for certain in the belief state of an autoepistemic agent. If at least some of the preferences of the autoepistemic agent are available, the set of stable expansions that correspond to possible states of belief can be reduced, and by cautious reasoning more formulae can be inferred. Prioritized autoepistemic logic is a formalization of autoepistemic reasoning with priorities.

Definition 2.2 *A pair $P = \langle \Phi, \mathcal{P} \rangle$ is a* prioritization, *if $\Phi \subseteq \mathcal{L}_{ae}$ is a set of formulae and $\mathcal{P} \subseteq \Phi \times \Phi$ is a relation on Φ that is transitive and asymmetric, and all strict total orders $\mathcal{T} \supseteq \mathcal{P}$ are well-orderings*[4].

The set Φ is a set of formulae whose presence or absence in a stable expansion (set of beliefs) is relevant in estimating whether it should be adopted or not. The formulae in Φ are to be minimized, i.e. the lower number of them in a stable expansion the better. Not all members of Φ are equal. The relation $\mathcal{P}$ describes the preferences between formulae. The meaning of $\sigma\mathcal{P}\tau$ is that σ is less believable than τ, i.e. when having to believe one of them the agent chooses τ. The minimal elements of a prioritization partial order are elements σ such that $\tau\mathcal{P}\sigma$ for no $\tau \in \Phi$. An element σ is the least element, if $\sigma\mathcal{P}\tau$ for all $\tau \in \Phi$. The prioritization relation is asymmetric and hence it

[4] A relation $\mathcal{P}$ on a set D is a well-ordering if every non-empty subset S of D has an element σ such that $\tau\mathcal{P}\sigma$ for no $\tau \in S$.

describes strict preferences. The transitivity property makes the prioritization an ordering relation. The well-ordering property rules out the existence of infinite sequences of less and less believable beliefs, and thereby eliminates the possibility that two stable expansions cannot be ordered because of the lack of a most significant element in the prioritization.

Example 2.3 Sometimes we visualize stable expansions with a given prioritization in mind. For example the diagram below on the left depicts the prioritization $\langle \Phi, \mathcal{P} \rangle = \langle \{\alpha, \beta, \gamma\}, \{(\alpha, \beta), (\gamma, \beta)\}\rangle$. Whenever $a\mathcal{P}b$, a is depicted below b in a diagram. The other two diagrams represent respectively stable expansions E_1 and E_2 such that $E_1 \cap \Phi = \{\beta, \gamma\}$ and $E_2 \cap \Phi = \{\alpha\}$.

$$\begin{tikzcd} & \beta & \\ \alpha \arrow[ur] & & \gamma \arrow[ul] \end{tikzcd} \quad \begin{tikzcd} & \bullet & \\ \circ \arrow[ur] & & \bullet \arrow[ul] \end{tikzcd} \quad \begin{tikzcd} & \circ & \\ \bullet \arrow[ur] & & \circ \arrow[ul] \end{tikzcd}$$

■

Definition 2.3 (P-preferred) *Let $P = \langle \Phi, \mathcal{P} \rangle$ be a prioritization and $\Sigma \subseteq \mathcal{L}_{ae}$ a set of formulae. Then E is a P-*preferred *stable expansion of Σ if and only if it is a stable expansion of Σ and the following condition is fulfilled. There is a strict total order $\mathcal{T}$ on Φ such that $\mathcal{T} \supseteq \mathcal{P}$ and for all stable expansions E' of Σ the following holds.*

For all $\phi \in \Phi(\phi \in E \backslash E'$ implies there is $\psi \in E' \backslash E$ such that $\psi \mathcal{T} \phi)$.

Example 2.4 Let E_1 and E_2 be the stable expansions of a set of formulae Σ. Let the prioritization $P = \langle \Phi, \mathcal{P} \rangle$ be as depicted below on the left (i.e. $\Phi = \{\alpha_1, \alpha_2, \alpha_3\}, \mathcal{P} = \{\langle \alpha_1, \alpha_2 \rangle, \langle \alpha_1, \alpha_3 \rangle, \langle \alpha_2, \alpha_3 \rangle\}$) and the stable expansions E_1 and E_2 as below in the middle and on the right.

$$\begin{tikzcd} P & E_1 & E_2 \\ \alpha_3 & \bullet & \circ \\ \alpha_2 \arrow[u] & \circ \arrow[u] & \bullet \arrow[u] \\ \alpha_1 \arrow[u] & \circ \arrow[u] & \circ \arrow[u] \end{tikzcd}$$

The stable expansion E_1 is P-preferred because of the following. There is only one strict total order $\mathcal{T}$ of $\mathcal{P}$, namely $\mathcal{P}$ itself. When instantiating $E = E_1$ and $E' = E_2$, the stable expansion E_1 fulfills the condition in Definition 2.3. The formula α_3 is in $E_1 \backslash E_2$, but there is the formula $\alpha_2 \in E_2 \backslash E_1$ such that $\alpha_2 \mathcal{T} \alpha_3$. On the other hand, E_2 is not P-preferred because there is $\alpha_2 \in E_2 \backslash E_1$ but there is no ψ such that $\psi \mathcal{T} \alpha_2$ and $\psi \in E_1 \backslash E_2$. ■

Example 2.5 Let P be the prioritization and E_1 and E_2 the stable expansions in Example 2.3. There are two strict total orders that extend $\mathcal{P}$, call them $\mathcal{T}_1$ and $\mathcal{T}_2$, which are shown below.

$$
\begin{array}{ccc}
\mathcal{T}_1 & E_1 & E_2 \\
 & & \\
\beta & \bullet & \circ \\
\uparrow & \uparrow & \uparrow \\
\gamma & \bullet & \circ \\
\uparrow & \uparrow & \uparrow \\
\alpha & \circ & \bullet
\end{array}
\qquad\qquad
\begin{array}{ccc}
\mathcal{T}_2 & E_1 & E_2 \\
 & & \\
\beta & \bullet & \circ \\
\uparrow & \uparrow & \uparrow \\
\alpha & \circ & \bullet \\
\uparrow & \uparrow & \uparrow \\
\gamma & \bullet & \circ
\end{array}
$$

Because $\alpha \in E_2 \backslash E_1$ and there are no formulae ψ such that $\psi \mathcal{T}_1 \alpha$, E_1 is a P-preferred stable expansion. Similarly, because $\gamma \in E_1 \backslash E_2$ and there are no formulae ψ such that $\psi \mathcal{T}_2 \gamma$, also E_2 is a P-preferred stable expansion. ■

The last line of Definition 2.3 is an ordering condition for lexicographic orderings. As indicated by the examples, E_1 comes before E_2 according to some strict total order if the least element ϕ in the ordering such that either $\phi \in E_1 \backslash E_2$ or $\phi \in E_2 \backslash E_1$, belongs to $E_2 \backslash E_1$. This suggests an alternative equivalent definition of the comparison:

$$\text{for no } \phi \in (\Phi \cap E) \backslash E', \text{ for all } \psi, \psi \mathcal{T} \phi \text{ implies } (\psi \in E \text{ iff } \psi \in E').$$

For instance Baker and Ginsberg [1989] define minimal models in prioritized circumscription about this way. The difference in these definitions show up only when used with partial orderings.

2.4 Finite characterizations of stable expansions

Niemelä [1990] presents a useful characterization of stable expansions using the notion of full sets. The proofs in this report frequently use full sets. The usefullness of full sets stems from the fact that stable expansions contain large amounts of redundant information that does not have to be represented explicitly. The relevant subset of a stable expansion of a set Σ is the set $Sf^L(\Sigma)$ of subformulae of Σ that begin with the L operator. The set $Sf^{qL}(\Sigma)$ is the subset of $Sf^L(\Sigma)$ consisting of those subformulae of Σ beginning with L and are not in the scope of an L operator. The following definitions and theorems are from [Niemelä, 1990].

Definition 2.4 *A set of formulae Λ is Σ-full if it satisfies the following conditions.*

$$
\begin{array}{l}
\Lambda \subseteq Sf^L(\Sigma) \cup \{\neg L\chi | L\chi \in Sf^L(\Sigma)\} \\
\textit{for all } L\chi \in Sf^L(\Sigma),\ L\chi \in \Lambda \textit{ iff } \Sigma \cup \Lambda \models \chi \\
\textit{for all } L\chi \in Sf^L(\Sigma),\ \neg L\chi \in \Lambda \textit{ iff } \Sigma \cup \Lambda \not\models \chi
\end{array}
$$

Theorem 2.5 *For a set of formulae Σ there is a bijective mapping from the Σ-full sets to the stable expansions of Σ.*

If there are only finitely many different subformulae beginning with an L in Σ, then the respective Σ-full sets are finite. This is the case e.g. when Σ is finite. The following definition and theorem describe how the membership in a stable expansion of a set of premises Σ can be determined in terms of a Σ-full set Λ.

Definition 2.6 *Given a set of formulae Σ and a formulae ϕ,*

$$\Sigma \models_L \phi \textit{ iff } \Sigma \cup SB_\Sigma(\phi) \models \phi$$

where $SB_\Sigma(\phi) = \{L\chi \in Sf^{qL}(\phi) | \Sigma \models_L \chi\} \cup \{\neg L\chi | L\chi \in Sf^{qL}(\phi) \textit{ and } \Sigma \not\models_L \chi\}$.

Theorem 2.7 *Let Λ be a Σ-full set. Then $\Delta = \{\phi | \Sigma \cup \Lambda \models_L \phi\}$ is the unique stable expansion of Σ for which $\Lambda \subseteq \{L\phi | \phi \in \Delta\} \cup \{\neg L\phi | \phi \notin \Delta\}$.*

Definition 2.8 *Let Σ be a set of formulae and Λ a Σ-full set. Define $SE_\Sigma(\Lambda) = \{\phi | \Sigma \cup \Lambda \models_L \phi\}$.*

Example 2.6 Consider the set $\Sigma = \{\neg L\phi \rightarrow \psi, \neg L\psi \rightarrow \phi\}$. It has four candidate Σ-full sets, two of which are in fact Σ-full. The set $\Lambda_1 = \{\neg L\phi, L\psi\}$ is Σ-full because $\Sigma \cup \Lambda_1 \not\models \phi$ and $\Sigma \cup \Lambda_1 \models \psi$. By symmetry also $\Lambda_2 = \{L\phi, \neg L\psi\}$ is Σ-full. ■

2.5 Properties of preferred stable expansions

For prioritizations P that are orderings of infinitely many formulae the existence of P-preferred stable expansions is not guaranteed. If there is no greatest element, there is the possibility of an infinite sequence of better and better stable expansions, none of which is P-preferred. This is demonstrated in Example 2.8 below. If there is no least element, two stable expansions are not necessarily put to any order by lexicographic comparison. The well-ordering requirement in the definition of prioritizations is meant to prevent this situation. We motivate this requirement by an example.

Example 2.7 Let $\Sigma = \{\neg L\alpha \rightarrow \beta, \neg L\beta \rightarrow \alpha\} \cup \{\alpha \leftrightarrow \alpha_i | i \geq 1\} \cup \{\beta \leftrightarrow \beta_i | i \geq 1\}$. The illegal prioritization P and the two stable expansions of Σ, E_1 and E_2, are depicted in Figure 3 on the left. There is no least element in this illegal prioritization. There is only one strict total order $\mathcal{T}$ that extends $\mathcal{P}$, $\mathcal{P}$ itself. For each $\alpha_i \in E_2 \backslash E_1$ there is $\beta_j \in E_1 \backslash E_2$ (take any $j \geq i$) such that $\beta_j \mathcal{T} \alpha_i$, and for each $\beta_j \in E_1 \backslash E_2$ there is $\alpha_i \in E_2 \backslash E_1$ (take any $i > j$) such

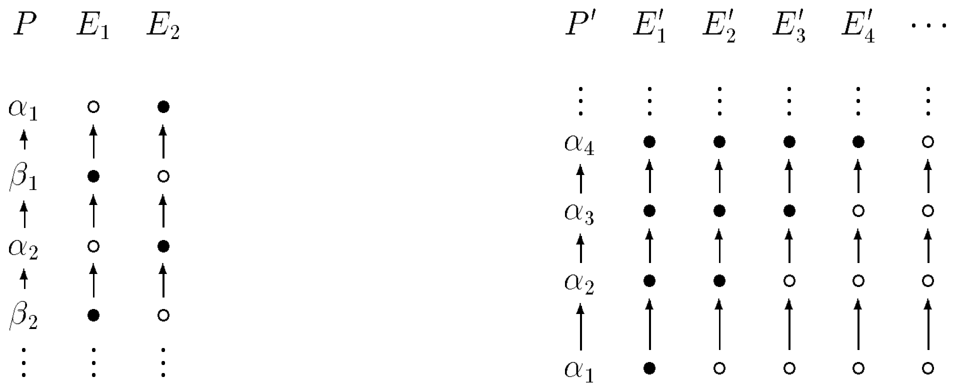


Figure 3: Two pathological prioritizations and sets of stable expansions

that $\alpha_i \mathcal{T} \beta_j$. Hence neither of the two stable expansions is P-preferred. Note that the infiniteness of Σ is not essential: let $\Sigma = \{\neg L\alpha \rightarrow \alpha, \neg L\beta \rightarrow \beta\}$ and replace each α_i and β_i in the prioritization with a conjunction of i αs and i βs. ■

Example 2.8 Let $\Sigma = \{L\alpha_i \rightarrow \alpha_i | i \geq 1\} \cup \{\alpha_{i-1} \rightarrow \alpha_i | i \geq 2\}$. The prioritization P' and the infinitely many stable expansions $E'_i, i \geq 1$ of Σ are depicted in Figure 3 on the right. In this prioritization, there is no greatest element. For each stable expansion E_i, there is stable expansion E_j (take any $j > i$) such that $\alpha_i \in E_i \backslash E_j$ and for no α_k, $\alpha_k \mathcal{T} \alpha_i$ is $\alpha_k \in E_j \backslash E_i$. Hence no stable expansion is P'-preferred. ■

If a prioritization P is finite and there are stable expansions of Σ, there necessarily are P-preferred stable expansions of Σ, as shown by the following theorem.

Theorem 2.9 (Existence) *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a set of formulae such that there is at least one stable expansion of Σ. Let $P = \langle \Phi, \mathcal{P} \rangle$ be a prioritization where Φ is finite, and $\mathcal{T} \supseteq \mathcal{P}$ is a strict total order on Φ. Then there is a stable expansion E of Σ such that for all stable expansions E' of Σ,*

$$\text{for all } \phi \in \Phi(\phi \in E \backslash E' \text{ implies there is } \psi \in E' \backslash E \text{ such that } \psi \mathcal{T} \phi).$$

Therefore E is a P-preferred stable expansion of Σ.

Proof: We prove by induction that there are preferred stable expansions for all finite cardinalities $n = |\Phi|$ of Φ.

Base case $n = 0$: Now $\Phi = \emptyset$, and for any given stable expansions E and E', obviously for all $\phi \in \Phi$ everything holds. Hence all stable expansions – by assumption there is at least one – are P-preferred.

Inductive case $n \geq 1$: Let $\mathcal{T} \supseteq \mathcal{P}$ be a strict total order on Φ. Then $P = \langle \Phi, \mathcal{T} \rangle$ is a prioritization. Let ψ be the greatest element of $\mathcal{T}$, and $\mathcal{T}' = \mathcal{T} \cap (\Phi \backslash \{\psi\} \times \Phi \backslash \{\psi\})$. Let X be the set of stable expansions E such that for all stable expansions E' of Σ,

$$\text{for all } \phi \in \Phi \backslash \{\psi\} (\phi \in E \backslash E' \text{ implies there is } \theta \in E' \backslash E \text{ such that } \theta \mathcal{T}' \phi).$$

By induction hypothesis this set is non-empty. Let $X_0 = \{E \in X | \psi \notin E\}$ and $X_1 = \{E \in X | \psi \in E\}$. Assume $X_0 \neq \emptyset$. Let E be a member of X_0 and let E' be any stable expansion of Σ. Assume there is $\phi \in (\Phi \cap E) \backslash E'$. Because $\psi \notin E$, it is not possible that $\phi = \psi$. Therefore by induction hypothesis, there is $\theta \in E' \backslash E$ such that $\theta \mathcal{T}' \phi$, and therefore $\theta \mathcal{T} \phi$. Hence E is P-preferred.

Assume $X_0 = \emptyset$. Now X_1 must be non-empty because $X = X_0 \cup X_1$ is non-empty. Let $E \in X_1$ and let E' be any stable expansion of Σ. Assume there is $\phi \in (\Phi \cap E) \backslash E'$. We have to show that there is $\theta \in E' \backslash E$ such that $\theta \mathcal{T} \phi$. We analyze two cases. Assume $\phi \neq \psi$. Now by induction hypothesis there is $\theta \in E' \backslash E$ such that $\theta \mathcal{T}' \phi$. Because $\mathcal{T} \supseteq \mathcal{T}'$, also $\theta \mathcal{T} \phi$. Assume $\phi = \psi$. Because $\psi = \phi \notin E'$ and $X_0 = \emptyset$, $E' \notin X$. Hence for some stable expansion E'' there is $\theta \in (\Phi \backslash \{\psi\} \cap E') \backslash E''$ and there is no $\gamma \in E'' \backslash E'$ such that $\gamma \mathcal{T}' \theta$. If $\theta \notin E$, then we have found θ such that $\theta \mathcal{T} \phi$ and $\theta \in E' \backslash E$. If $\theta \in E$, then $\theta \in E \backslash E''$ and by induction hypothesis there is $\theta' \in E'' \backslash E$ such that $\theta' \mathcal{T}' \theta$. Now $\theta' \mathcal{T} \phi$. Because $\theta' \mathcal{T}' \theta$ and $\theta' \in E''$, must $\theta' \in E'$ (because of the inexistence of γ as mentioned above). Hence $\theta' \in E' \backslash E$, and E fulfills the requirements for being P-preferred. □

The purpose of prioritizations is to select a subset of the stable expansions of a set. Sometimes exactly one stable expansion is desired. The next theorem gives a sufficient condition for the uniqueness of preferred stable expansions.

Theorem 2.10 *Let $\Sigma \in \mathcal{L}_{ae}$ be a set of formulae and let $\Phi = \{\phi | L\phi \in Sf^{qL}(\Sigma)\}$ or $\Phi = \{\phi | L\phi \in Sf^{L}(\Sigma)\}$. Then for any prioritization $P = \langle \Phi, \mathcal{T} \rangle$ where $\mathcal{T}$ is a strict total order on Φ, there is at most one P-preferred stable expansion of Σ.*

Proof: By the version of the full-set notion presented in [Niemelä, 1992a, Section 3], the intersection of the set $\{\phi | L\phi \in Sf^{qL}(\Sigma)\}$ with different stable expansions of Σ are different (the same holds for $\{\phi | L\phi \in Sf^{L}(\Sigma)\}$). Let E and E' be P-preferred stable expansions of Σ. There is only one strict total order that extends $\mathcal{T}$, the relation $\mathcal{T}$ itself. Now both

$$\text{for all } \phi \in \Phi (\phi \in E \backslash E' \text{ implies there is } \psi \in E' \backslash E \text{ such that } \psi \mathcal{T} \phi)$$

and

$$\text{for all } \phi \in \Phi (\phi \in E' \backslash E \text{ implies there is } \psi \in E \backslash E' \text{ such that } \psi \mathcal{T} \phi).$$

Assume E and E' are different stable expansions, i.e. $E \cap \Phi \neq E' \cap \Phi$. Hence there is an element of Φ that is in one but not in the other. Take the least element ϕ_1 of Φ such that $\phi_1 \in E \backslash E'$. There is a least such element because $\mathcal{T}$ is a well-ordering according to the definition prioritizations. Because E is P-preferred, there is ϕ_2 such that $\phi_2 \mathcal{T} \phi_1$ and $\phi_2 \in E' \backslash E$. Because E' is P-preferred, there is ϕ_3 such that $\phi_3 \mathcal{T} \phi_2$ and $\phi_3 \in E \backslash E'$. By the transitivity of $\mathcal{T}$ and the assumption that ϕ_1 is the least in $(\Phi \cap E) \backslash E'$, this is impossible. Hence there is at most one P-preferred stable expansion. □

Given a prioritization $P = \langle \Phi, \mathcal{P} \rangle$ and a set of formulae Σ, an interesting question is the number of P-preferred stable expansions Σ can have. For a set Σ and a prioritization P such that there are very many stable expansions for which the intersections with Φ are the same, the question is not very interesting. Therefore we give in the next theorem an upper bound for the number of preferred stable expansions assuming that the stable expansions differ in formulae in Φ. The upper bound is given in terms of the size of Φ only.

Theorem 2.11 *Let $\Sigma \in \mathcal{L}_{ae}$ be a set of formulae and $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization. Let $n = |\Phi|$. Suppose that for no two stable expansions E and E' of Σ, $E \cap \Phi = E' \cap \Phi$. Then there can be at most*

$$\binom{n}{\lfloor n/2 \rfloor}$$

P-preferred stable expansions of Σ.

Proof: Let $P' = \langle \Phi, \mathcal{P}' \rangle$, where $\mathcal{P}' \subseteq \mathcal{P}$. Because all strict total orders $\mathcal{T}$ such that $\mathcal{T} \supseteq \mathcal{P}$ also fulfill $\mathcal{T} \supseteq \mathcal{P}'$, the number of P'-preferred stable expansions cannot be smaller than that of P-preferred stable expansions. Hence for a fixed Φ, the greatest number of preferred stable expansions is obtained when the relation is empty, i.e. for the prioritization $P_f = \langle \Phi, \emptyset \rangle$. The set of P_f-preferred stable expansions consists of stable expansions E such that $E \cap \Phi$ is a set-inclusion minimal subset of Φ among the sets $\{E \cap \Phi | E \text{ is a stable expansion of } \Sigma\}$. A *Sperner system* of a set S is a collection $M_1, \ldots, M_n$ of subsets of S such that $M_i \subset M_j$ for no distinct i, j [Anderson, 1987]. The maximum possible number of preferred stable expansions is obtained by noticing that the sets $E \cap \Phi$ where E is P_f-preferred, form a Sperner system of Φ. Sperner systems can have at most $\binom{n}{\lfloor n/2 \rfloor}$ elements. Hence this is an upper bound for the number of preferred stable expansions of Σ. □

2.6 Complexity of the decision problems

In this section we evaluate the complexity of the decision problems of prioritized autoepistemic logic with respect to the complexity of the decision problems of the underlying (monotonic) logic. The major decision problems in autoepistemic logic are *cautious* and *brave* reasoning, corresponding to the membership of a formula in *all* stable expansions of a set of initial premises, and the membership in *some* stable expansion, respectively. The introduction of priorities modifies these to *prioritized cautious* reasoning and *prioritized brave* reasoning, which correspond to the membership in all P-preferred stable expansions and the membership in some P-preferred stable expansions, respectively. The languages corresponding to these decision problems are DERpc and DERpb, which are formally defined as follows.

Definition 2.12 *The language DERpc consists of tuples $\langle \Sigma, P, \phi \rangle$, where $\Sigma \subseteq \mathcal{L}_{ae}$ is a set of formulae, P is a prioritization, and $\phi \in \mathcal{L}_{ae}$ is a formula that belongs to all P-preferred stable expansions of Σ.*

The language DERpb consists of tuples $\langle \Sigma, P, \phi \rangle$, where $\Sigma \subseteq \mathcal{L}_{ae}$ is a set of formulae, P is a prioritization, and $\phi \in \mathcal{L}_{ae}$ is a formula that belongs to at least one P-preferred stable expansion of Σ.

To analyze the complexity of these languages, we first give some auxiliary definitions and results. The size of a prioritized autoepistemic theory $\langle P, \Sigma \rangle$ is the sum of the sizes of $P = \langle \Phi, \mathcal{P} \rangle$ and Σ. The size of Σ is the sum of the lengths of the formulae in it, and the size of P is the sum of the lengths of formulae in Φ and the cardinality of $\mathcal{P}$. Given a logic and its consequence relation $\models$, the language CN consists of pairs $\langle \Sigma, \phi \rangle \in 2^{\mathcal{L}_{ae}} \times \mathcal{L}_{ae}$ such that $\Sigma \models \phi$. CNL is the language that consists of pairs $\langle \Sigma, \phi \rangle \in 2^{\mathcal{L}_{ae}} \times \mathcal{L}_{ae}$ such that $\Sigma \models_L \phi$. This language is Turing reducible to CN in deterministic polynomial time [Niemelä, 1990]. In analyzing the complexity of the decision problems of prioritized autoepistemic logic we use the language FNC that consists of quadruples

- $\langle 0, \Sigma, \Lambda, P \rangle$, where $\Sigma \subseteq \mathcal{L}_{ae}$ is a set of formulae and Λ is a Σ-full set,
- $\langle 1, \Sigma, \Lambda, P \rangle$, where $\Sigma \subseteq \mathcal{L}_{ae}$ is a set of formulae, Λ a Σ-full set, $P = \langle \Phi, \mathcal{T} \rangle$ is a prioritization where $\mathcal{T}$ is a strict total order on Φ, and $SE_\Sigma(\Lambda)$ is *not* a P-preferred stable expansion of Σ, and
- $\langle 2, \Sigma, \Lambda, \phi \rangle$, such that $\langle \Sigma, \phi \rangle \in$ CNL.

Lemma 2.13 *FNC is Turing reducible to CN in nondeterministic polynomial time.*

Proof: This language is accepted by a Turing machine described by the procedure in Figure 4. The Turing machine uses an oracle for the language CN. The Turing machine runs in nondeterministic polynomial time for the following reasons. Questions concerning membership in the language CNL can be answered in deterministic polynomial time using the oracle for CN [Niemelä, 1990]. There are three other parts in the procedure that do not run in constant time. The first is the procedure *full* that runs in time proportional to the size of Λ. The second is the nondeterministic guessing of a Σ-full set Λ'. The size of Λ' is proportional to the size of Σ, and hence the amount of guessing is polynomial. The third is the procedure *compare*. This procedure considers each element of the total order $\mathcal{T}$ one at a time and clearly runs in polynomial time. The call compare($\Sigma, \Lambda, \Lambda', \langle \Phi, \mathcal{T} \rangle$) returns *true* if and only if for all $\phi \in \Phi$,

$$\phi \in SE_\Sigma(\Lambda) \backslash SE_\Sigma(\Lambda') \text{ implies there is } \psi \in SE_\Sigma(\Lambda') \backslash SE_\Sigma(\Lambda) \text{ such that } \psi \mathcal{T} \phi.$$

The purpose of the call in the body of the procedure FNC is to show that $SE_\Sigma(\Lambda)$ is a P-preferred stable expansion. If it were not, a full set Λ' could be guessed so that the call to *compare* would return *false*. □

Theorem 2.14 *DERpb is Turing reducible to FNC in nondeterministic polynomial time.*

Proof: The language DERpb is accepted by a Turing machine described by the procedure in Figure 5. The Turing machine uses an oracle for the language FNC. First the machine guesses a candidate Σ-full set Λ and a strict total order $\mathcal{T}$ that extends the relation $\mathcal{P}$ of the prioritization P. Then the machine verifies that Λ is Σ-full (the first consultation of oracle for FNC), that $SE_\Sigma(\Lambda)$ is P'-preferred stable expansion of Σ (the second consultation), and that $\Sigma \cup \Lambda \models_L \phi$ (the third consultation). That prioritization P' can be used instead of P is because $\mathcal{T}$ is an extension of $\mathcal{P}$ to a strict total order. Because $\mathcal{T}$ is guessed nondeterministically and a nondeterministic Turing machine accepts whenever there is *some* accepting computation, $\mathcal{T}$ is a strict total order using which $SE_\Sigma(\Lambda)$ can be shown P-preferred whenever such a strict total order exists.

The only parts of the procedure that do not run in constant time are the nondeterministic guessing of a Σ-full set, guessing of a strict total order $\mathcal{T}$, and verifying that it extends $\mathcal{P}$. The number of steps needed in the guessing of the full set is proportional to the size of Σ. At each step for $L\phi \in \Sigma$ either $L\phi$ or $\neg L\phi$ is included in the set. The number of steps in guessing a strict total order on Φ is polynomial on $|\Phi|$. Verification of $\mathcal{T} \supseteq \mathcal{P}$ is clearly polynomial time. □

```
procedure FNC(n, Σ, Λ, P);
  procedure full(Σ, Λ);
  begin
    if Sf^{qL}(Λ) ≠ Sf^L(Σ) then return false;
    for each ¬Lχ ∈ Λ do if ⟨Σ ∪ Λ, χ⟩ ∈ CN then return false;
    for each Lχ ∈ Λ do if ⟨Σ ∪ Λ, χ⟩ ∉ CN then return false;
    return true
  end
begin
  if n = 0 then if full(Σ, Λ) then accept else reject
  else if n = 1
  then
    guess nondeterministically a candidate Σ-full set Λ′;
    if not full(Σ, Λ′) then
      reject
    else
      if compare(Σ, Λ, Λ′, P) then reject
      else accept
  else
    if ⟨Σ, P⟩ ∈ CNL then accept
    else reject
end

procedure compare(Σ, Λ, Λ′, ⟨Φ, T⟩);
var i : integer;
let T be the ordering φ_0 T φ_1 T ··· T φ_n of Φ;
begin
  for i := 0 to n do
    if ⟨Σ ∪ Λ, φ_i⟩ ∈ CNL and ⟨Σ ∪ Λ′, φ_i⟩ ∉ CNL then return false;
    if ⟨Σ ∪ Λ, φ_i⟩ ∉ CNL and ⟨Σ ∪ Λ′, φ_i⟩ ∈ CNL then return true
  end for;
  return true
end
```

Figure 4: A decision procedure for FNC

procedure DERpb(Σ, P, ϕ);
begin
let Φ and $\mathcal{P}$ be such that $P = \langle \Phi, \mathcal{P} \rangle$;
guess nondeterministically a candidate Σ-full set Λ;
guess nondeterministically a strict total order $\mathcal{T}$ on Φ;
if $\mathcal{T} \not\supseteq \mathcal{P}$ **then reject**;
let $P' := \langle \Phi, \mathcal{T} \rangle$;
if $\langle 0, \Sigma, \Lambda, \emptyset \rangle \in$ FNC
and $\langle 1, \Sigma, \Lambda, P' \rangle \notin$ FNC
and $\langle 2, \Sigma \cup \Lambda, \emptyset, \phi \rangle \in$ FNC
then accept
else reject
end

Figure 5: A decision procedure for DERpb

Theorem 2.15 *The complement of DERpc is Turing reducible to FNC in nondeterministic polynomial time.*

Proof: Proof is similar to that of Theorem 2.14. The procedure is shown in Figure 6. □

For prioritized autoepistemic logic based on classical propositional logic, we give upper and lower bounds for the location of the decision problems in the polynomial hierarchy [Garey and Johnson, 1979]. This hierarchy is an infinite hierarchy of complexity classes Σ_i^p, Π_i^p, and $\Delta_i^p, i \geq 0$ defined using oracle Turing machines in the following way.

$$\begin{array}{lllllllll} \Sigma_0^p & = & P & \Pi_0^p & = & P & \Delta_0^p & = & P \\ \Sigma_{n+1}^p & = & \mathrm{NP}^{\Sigma_n^p} & \Pi_{n+1}^p & = & \text{co-}\Sigma_{n+1}^p & \Delta_{n+1}^p & = & P^{\Sigma_n^p} \end{array}$$

$C_1^{C_2}$ denotes the class of languages that belong to the class C_1 if given an oracle for a language in C_2, and co-C denotes the set of languages whose complements are in the class C. Note that Σ_1^p =NP.

When prioritized autoepistemic logic is based on the classical propositional logic, the decision problem CN of the associated language is in co-NP $= \Pi_1^p$ as shown in [Niemelä, 1990]. Because FNC is Turing reducible to CN in nondeterministic polynomial time we immediately get the following lemma.

Lemma 2.16 *When the underlying logic is the classical propositional logic, the language FNC is in Σ_2^p.*

procedure co-DERpc(Σ, P, ϕ);
begin
let Φ and $\mathcal{P}$ be such that $P = \langle \Phi, \mathcal{P} \rangle$;
guess nondeterministically a candidate Σ-full set Λ;
guess nondeterministically a strict total order $\mathcal{T}$ on Φ;
if $\mathcal{T} \not\supseteq \mathcal{P}$ **then reject**;
let $P' := \langle \Phi, \mathcal{T} \rangle$;
if $\langle 0, \Sigma, \Lambda, \emptyset \rangle \in$ FNC
and $\langle 1, \Sigma, \Lambda, P' \rangle \notin$ FNC
and $\langle 2, \Sigma \cup \Lambda, \emptyset, \phi \rangle \notin$ FNC
then accept
else reject
end

Figure 6: A decision procedure for co-DERpc

By Theorems 2.14 and 2.15 we get upper bounds for the location of the prioritized decision problems in the polynomial hierarchy. These results are immediate by the definition of the polynomial hierarchy.

Theorem 2.17 *When the underlying logic is the classical propositional logic, the language DERpb that corresponds to prioritized brave reasoning, is in Σ_3^p, and the language DERpc that corresponds to prioritized cautious reasoning, is in Π_3^p.*

The decision problems of ordinary autoepistemic logic can be reduced to the corresponding problems of prioritized autoepistemic logic by taking the prioritization to be the empty one $\langle \emptyset, \emptyset \rangle$. In this case the formula quantified over Φ in the definition of preferred stable expansions is always true whatever the stable expansions in question are. Thus every stable expansion is P-preferred. Thereby we get the theorem below as a direct corollary of results in [Gottlob, 1992]. Gottlob shows that cautious and brave reasoning in Moore's autoepistemic logic are respectively Π_2^p-hard and Σ_2^p-hard, i.e. all problems in Π_2^p and Σ_2^p can be respectively transformed to cautious and brave reasoning in polynomial time.

Theorem 2.18 *Prioritized brave and cautious reasoning are Σ_2^p-hard and Π_2^p-hard, respectively.*

For DERpb and DERpc the membership on the second level of the polynomial hierarchy Δ_3^p as well as their Σ_3^p-hardness and Π_3^p-hardness respectively remain open.

3 Tractability of nonmonotonic reasoning with priorities

The decision problems of nonmonotonic logics are very complex also in cases where the required classical reasoning can be performed very efficiently. For example, decision problems of default logic with Horn clauses and default rules of very restricted form are NP-hard [Stillman, 1990]. The complexity of nonmonotonic reasoning seems to be due to the possibility of an exponential number of candidate expansions. Sufficient priorities restrict the number of relevant stable expansions to the preferred ones, hence the possibility of efficient decision procedures for restricted subclasses of autoepistemic logic with priorities is worth investigating. All results in this section are about autoepistemic logic based on classical propositional logic.

3.1 Restricted intractable classes

A property of our notion of preferred stable expansions is that when prioritizations are finite, the existence of preferred stable expansions is equivalent to the existence of stable expansions of Σ in general (Theorem 2.9). Hence, the problem of the existence of preferred stable expansions is computationally as complex as the problem of the existence of stable expansions. In [Marek and Truszczyński, 1991] the class of K_1-theories that corresponds to propositional logic programs is introduced. The formulae in theories of this class are of the form $Lp_1 \wedge \cdots \wedge Lp_n \wedge \neg Lp_{n+1} \wedge \cdots \wedge \neg Lp_m \rightarrow q$, where p_i and q are propositional variables.

Theorem 3.1 *The problem of existence of preferred stable expansions is NP-complete.*

Proof: The problem of existence of stable expansions of K_1-theories is NP-complete by [Marek and Truszczyński, 1991, Theorem 6.4] [5]. This problem is equivalent to the existence of P-preferred stable expansions for K_1-theories. □

Corollary 3.2 *Prioritized cautious reasoning for K_1-theories is NP-hard.*

Proof: By reduction from the complement of the NP-complete problem in Theorem 3.1. First note that stable expansions of K_1-theories are always consistent. Let p be $\perp$ (or any other atomic formula that does not occur

[5] Actually Marek and Truszczyński prove the stronger result that for sets of formulae of the form $\neg Lp \rightarrow q$ the existence of stable expansions is NP-complete.

in the K_1-theory in question). The NP-complete existence question is Turing reducible in polynomial time to testing whether all preferred stable expansions contain p. Specifically, there are preferred stable expansions of a K_1-theory Σ if and only if p does not belong to all stable expansions of Σ. Hence the problem is NP-hard [6]. □

In a similar way prioritized brave reasoning can be proved NP-hard. Note that the proofs of the above theorem and corollary do not depend on any assumptions about the properties of prioritizations used. Hence tractability cannot be achieved merely by restricting to some subclass of prioritizations, e.g. prioritizations that totally order some set formulae. This remark directly hints at possible tractable classes: all tractable subclasses of prioritized autoepistemic logic are based on subclasses of autoepistemic theories for which the existence of stable expansions can be tested in polynomial time. Examples of such possibly tractable classes are autoepistemic counterparts of seminormal ordered and normal theories of default logic. Another, more radical way to obtain tractable classes of prioritized reasoning is to consider notions of most preferred stable expansions for which the existence does not coincide with existence of stable expansions.

3.2 Tractable classes

A source of complexity in the above intractable classes and in autoepistemic reasoning in general is the need to consider an exponential number of candidate stable expansions. Sufficient priorities restrict the number of interesting expansions to the P-preferred ones, but they are in general difficult to find. Below we give a tractable but very restricted class of prioritized autoepistemic theories for which the unique P-preferred stable expansion can be computed directly using the prioritization P. Exhaustive search is avoided by eliminating the need of retracting any decisions to include formulae $L\phi$ or $\neg L\phi$ in full sets.

Theorem 3.3 *For finite sets Σ of propositional Horn clauses and formulae of the form $\neg L\neg\phi \rightarrow \phi$ where ϕ is a propositional Horn clause, and prioritizations $P = \langle \Phi, \mathcal{P} \rangle$ where $\Phi = \{\psi | L\psi \in Sf^{qL}(\Sigma)\}$ and $\mathcal{P}$ is a strict total order on Φ, there is exactly one preferred stable expansion and the problem of testing membership in it for formulae χ that are negated Horn clauses, is solvable in polynomial time on the size of the set $\Sigma \cup \{\chi\}$.*

[6] We use the notion of NP-hardness of [Garey and Johnson, 1979]. A co-NP-hardness result using a notion of NP-hardness based on many-to-one reducibility [Balcázar *et al.*, 1988] is possible.

The same result holds if 2-literal clauses or other polynomial time subsets of classical propositional logic are used instead of Horn clauses.

Proof: By Theorem 2.10 there is at most one P-preferred stable expansion of Σ. A decision procedure is given in Figure 7. The correctness proof the procedure shows that there is at least one P-preferred stable expansion of Σ. The procedure computes a Σ-full set Λ and returns *true* if χ belongs to $SE_\Sigma(\Lambda)$. We first prove the correctness of the algorithm and then give a quadratic-time upper bound for the computation. Let $\neg\phi_1, \ldots, \neg\phi_n$ be the ordering $\mathcal{P}$ of the formulae of Φ. We prove by induction that for all $i, 0 \leq i \leq n$ the following equivalences hold for all $j, 1 \leq j \leq i$,

$$\begin{array}{rcl} L\neg\phi_j \in \Lambda_i & \text{iff} & \Sigma \cup \Lambda_i \models \neg\phi_j \\ \neg L\neg\phi_j \in \Lambda_i & \text{iff} & \Sigma \cup \Lambda_i \not\models \neg\phi_j \end{array}$$

The Σ-fullness of Λ is obtained when $i = n$.

Base case $i = 0$: Because $\Lambda_0 = \emptyset$, the claim is immediate.

Inductive case $i \geq 1$: For $\neg\phi_i$ the claim holds by construction. Let j be in $\{1, \ldots, i-1\}$. Either $L\neg\phi_j$ or $\neg L\neg\phi_j$ is in Λ_i.

$$\begin{array}{rlll} L\neg\phi_j \in \Lambda_i & \text{implies} & \Sigma \cup \Lambda_{j-1} \models \neg\phi_j & \text{By construction} \\ & \text{implies} & \Sigma \cup \Lambda_i \models \neg\phi_j & \text{By monotonicity of } \models \end{array}$$

When $\neg L\neg\phi_j \in \Lambda_i$ there are two cases to consider. First note that by the induction hypothesis $\Sigma \cup \Lambda_{i-1} \not\models \neg\phi_j$, i.e. $\Sigma \cup \Lambda_{i-1} \cup \{\phi_j\}$ is satisfiable. Because $\neg L\neg\phi_j \in \Lambda_{i-1}$ and $(\neg L\neg\phi_j \rightarrow \phi_j) \in \Sigma$, ϕ_j is true in every model of $\Sigma \cup \Lambda_{i-1}$.

Case 1: $\Lambda_i = \Lambda_{i-1} \cup \{\neg L\neg\phi_i\}$. By construction, the set $\Sigma \cup \Lambda_{i-1} \cup \{\phi_i\}$ is satisfiable. Let M be a model of the set. Modify M to falsify $L\neg\phi_i$. Now $M \models \Sigma \cup \Lambda_{i-1} \cup \{\phi_j, \neg L\neg\phi_i\}$. Hence $\Sigma \cup \Lambda_i \not\models \neg\phi_j$.

Case 2: $\Lambda_i = \Lambda_{i-1} \cup \{L\neg\phi_i\}$. By construction, the set $\Sigma \cup \Lambda_{i-1} \cup \{\phi_i\}$ is *not* satisfiable. Therefore ϕ_i is false in all models of $\Sigma \cup \Lambda_{i-1}$. Let M be a model of $\Sigma \cup \Lambda_{i-1}$. Modify M to satisfy $L\neg\phi_i$. Now $M \models \Sigma \cup \Lambda_{i-1} \cup \{\phi_j, L\neg\phi_i\}$. Hence $\Sigma \cup \Lambda_i \not\models \neg\phi_j$.

Now we show that the computation of Λ is $\mathcal{O}(s^2)$ where s is the sum of the lengths of formulae in Σ. By the satisfiability algorithm of [Dowling and Gallier, 1984] each satisfiability test can be performed in linear time on the size of the set to be tested. Let $s_1, \ldots, s_n$ be the sizes of $\phi_1, \ldots, \phi_n$. A quadratic upper bound for the time needed for the satisfiability tests is obtained as follows. An upper bound for the time taken by the satisfiability test at step i is $c + a(s + s_i + \sum_{j=1}^{i-1}(3 + s_j))$, i.e. the size of Σ, the size of ϕ_i and the

procedure decide(Σ, P, ϕ)
begin
  let $\neg\phi_1, \neg\phi_2, \ldots, \neg\phi_n$ be the ordering of Φ according to $\mathcal{P}$;
  $\Lambda_0 := \emptyset$;
  for $i := 1$ **to** n **do**
    if $\Sigma \cup \Lambda_{i-1} \cup \{\phi_i\}$ is not satisfiable (equivalently, $\Sigma \cup \Lambda_{i-1} \models \neg\phi_i$)
    then $\Lambda_i := \Lambda_{i-1} \cup \{L\neg\phi_i\}$
    else $\Lambda_i := \Lambda_{i-1} \cup \{\neg L\neg\phi_i\}$
  end for;
  $\Lambda := \Lambda_n$;
  if $\Sigma \cup \Lambda \models \phi$
  then return true **else return** false
end

Figure 7: A decision procedure for a tractable subset of PAEL

sum of the sizes of $L\neg\phi_j, 1 \leq j \leq i-1$ (possibly negated), and a, c are some constants. Now

$$\begin{aligned}
\sum_{i=1}^{n}(c + a(s + s_i + \sum_{j=1}^{i-1}(3 + s_j))) &\leq \sum_{i=1}^{n}(c + a(s + 3(i-1) + \sum_{j=1}^{n} s_j)) \\
&\leq nc + nas + 3an^2 + na\sum_{j=1}^{n} s_j \\
&\leq sc + sas + 3as^2 + sas \\
&= sc + (2a + 3a)s^2 \leq (c + 5a)s^2.
\end{aligned}$$

Finally, the consequence test $\Sigma \cup \Lambda \models \phi$ takes linear time on the size of $\Sigma \cup \Lambda \cup \{\neg\phi\}$. Hence the whole computation is $\mathcal{O}(n^2)$ where n is the size of $\Sigma \cup \{\phi\}$. Clearly, the proof could be reconstructed for other polynomial time subsets of propositional logic. The only part of the proof that has to be modified is the above chain of equations which depends on the linear-time assumption. □

Note that without priorities reasoning with prerequisite-free normal default rules and Horn clauses is NP-complete [Stillman, 1990]. A class of belief revision problems very closely related to the above class of formulae (without the restriction to Horn clauses) is one of the special cases investigated by [Nebel, 1991] in which his belief revision system satisfies all eight AGM postulates.

4 Prioritized autoepistemic logic and other non-monotonic formalisms

In this section we compare our prioritized autoepistemic logic to other approaches to nonmonotonic reasoning that either use some kind of preferences in their basic formulation (circumscription [McCarthy, 1980; Lifschitz, 1985], preferred subtheories [Brewka, 1989], or in general sanction several possible sets of conclusions and use preferences to single out only one or a subset (default logic [MacNish, 1991; Tan and Treur, 1992]).

4.1 Priorities in autoepistemic logic

4.1.1 Hierarchic autoepistemic logic

Konolige [1989] defines a layered version of autoepistemic logic. The language of this logic has n modal operators $L_0, L_1, \ldots, L_n$. The formula $L_i\phi$ refers to derivability of formula ϕ from layer i. The idea of this logic resembles the intuition behind stratification in autoepistemic logic [Gelfond, 1987; Marek and Truszczyński, 1991]. In hierarchic autoepistemic logic priorities between two conflicting defaults should always be expressed: if no priorities are given, the resulting expansion will be inconsistent. The assumption that priorities are always available seems unrealistic. Przymusinska [1989] reduces hierarchic autoepistemic logic to Moore's autoepistemic logic. The resulting theories have exactly one stable expansion.

4.1.2 Multi-agent autoepistemic logic

Toyama, Inagaki and Fukumura [1991] define a multi-agent version of autoepistemic logic. Instead of only one modal operator L the language $\mathcal{L}_{mae}$ of the logic has operators $L_1, \ldots, L_n$, one for each of n agents. Like hierarchic autoepistemic logic, multi-agent autoepistemic logic provides a mechanism of dividing sets of autoepistemic formulae into a hierarchy. Toyama et al. demonstrate how simple examples of single and multiple inheritance can be represented in their logic. In the paper, the use of this logic for inheritance reasoning is demonstrated. Each agent corresponds to a node in an inheritance hierarchy, and values in the hierarchy are propagated by "extended introspection" where an agents accesses the beliefs of other agents. Toyama et al. show that the lack of priorities between two defaults does not lead to an inconsistency – like in hierarchic autoepistemic logic – but to two distinct expansions like in Moore's autoepistemic logic.

The paper claims that using the logic it is easier to reduce the conflicts that give rise to multiple expansions. The definition of the logic does not restrict the connections between agents in any way, for example, each agent may access the beliefs of any other agent: the agents do not form a hierarchy. There does not seem to be any close connection between this logic and our prioritized autoepistemic logic (although their applications may be similar). We can, in fact, give a simple translation of the multi-agent autoepistemic logic to Moore's autoepistemic logic without priorities.

Definition 4.1 (Toyama *et al.*, 1991, Theorem 3.6) *Let $\mathcal{A} = \langle A_1, \ldots, A_n \rangle$ where each $A_i \subseteq \mathcal{L}_{mae}$. Then $\mathcal{T} = \langle T_1, \ldots, T_n \rangle$ is an* extension *of $\mathcal{A}$, if for all $i, 1 \leq i \leq n$,*

$$T_i \quad = \quad \{\phi | A_i \cup \bigcup_{j=1}^{n} \{L_j\psi | \psi \in T_j\} \cup \bigcup_{j=1}^{n} \{\neg L_j \psi | \psi \notin T_j\} \models \phi\}.$$

Definition 4.2 *Define the translation function $tr_{ma}(A_1, \ldots, A_n)$ as follows.*

$$\begin{array}{rcl} tr_{ma}(A_1, \ldots, A_n) & = & \{s_i(\phi) | \phi \in A_i, 1 \leq i \leq n\} \\ \textit{where} & & \\ s_i(p) & = & p_i, \ \textit{when } p \textit{ is a propositional variable} \\ s_i(L_j\phi) & = & L(s_j(\phi)) \\ s_i(\neg\phi) & = & \neg(s_i(\phi)) \\ s_i(\phi \otimes \phi') & = & s_i(\phi) \otimes s_i(\phi'), \ \textit{when } \otimes \ \textit{is a binary connective.} \end{array}$$

Here each s_j is a function from $\mathcal{L}_{mae}$ to $\mathcal{L}_{ae}$.

Lemma 4.3 *Let $\langle T_1, \ldots, T_n \rangle$ be an extension of some $\mathcal{A}$. Let $\psi \in \mathcal{L}_{mae}$ and $\psi' \in \mathcal{L}_{mae}$ be formulae. If $s_j(\psi) = s_j(\psi')$, then $\psi \in T_i$ if and only if $\psi' \in T_i$.*

Proof: Remark **A**: Let $\phi \in T_x$ if and only if $\phi' \in T_x$. Then $L_x\phi \leftrightarrow L_x\phi' \in T_i$. Proof: If $\phi \in T_x$, then $L_x\phi \in T_i$, and because T_i is closed under logical consequence, $L_x\phi' \rightarrow L_x\phi \in T_i$. In this case $L_x\phi' \in T_i$ and hence similarly $L_x\phi \rightarrow L_x\phi' \in T_i$. Because T_i is closed under logical consequence, $L_x\phi \leftrightarrow L_x\phi' \in T_i$. If $\phi \notin T_x$, then $\neg L_x\phi \in T_i$, and because T_i is closed under logical consequence $L_x\phi \rightarrow L_x\phi'$. Now $\phi' \notin T_x$, and similarly $L_x\phi' \rightarrow L_x\phi \in T_i$. Because T_i is is closed under logical consequence, $L_x\phi \leftrightarrow L_x\phi'$.

Define the level of nesting $\mathrm{LD}(\phi)$ of L in a formula ϕ as follows: if $Sf^{qL}(\phi) = \emptyset$, then $\mathrm{LD}(\phi) = 0$, otherwise $\mathrm{LD}(\phi) = 1 + \max_{L\chi \in Sf^{qL}(\phi)} \mathrm{LD}(\chi)$.

The proof of the lemma is by induction on $\mathrm{LD}(\psi)$. Note that when $s_j(\psi) = s_j(\psi')$, ψ and ψ' can be obtained from each other by renaming propositional variables and changing the subscripts of operators L.

Base case LD(ψ) = 0: If $s_j(\psi) = s_j(\psi')$, then ψ has to be the same formula as ψ' because s_j just changes propositional variables p to their p_j versions. Hence $\psi \in T_i$ iff $\psi' \in T_i$.

Inductive case LD(ψ) ≥ 1: Assume $s_j(\psi) = s_j(\psi')$. Again, the propositional variables in ψ and ψ' in corresponding locations and not inside an L_x operator, have to be the same. The formulae ψ and ψ' may differ only in the $L_x\phi$ formulae, i.e. where ψ has $L_x\phi$, ψ' has $L_y\phi'$ at the corresponding location. However, the indices x and y of L and the formulae ψ and ψ' are constrained by the requirement $s_j(L_x\phi) = s_j(L_y\phi')$. We analyze each occurrence of $L_x\phi$ in ψ. Let $L_y\phi'$ be the subformula of ψ' at the corresponding location where the subformula $L_x\phi$ is in ψ.

Case 1: Let ϕ contain propositional variables not inside an L_z operator. Because $s_j(L_x\phi) = s_j(L_y\phi')$, must $Ls_x(\phi)$ equal $Ls_y(\phi')$. Because there are propositional variables not inside L in ϕ, the equalities $x = y$ and $s_x(\phi) = s_y(\phi')$ hold. By induction hypothesis $\phi \in T_i$ iff $\phi' \in T_i$. Hence by Remark **A** and $x = y$, $(L_x\phi \leftrightarrow L_y\phi') \in T_i$.

Case 2: Let all propositional variables in ϕ be inside an L_z operator (for any z). Because in this case $s_u(\phi) = s_w(\phi)$ for all u, w, and $Ls_x(\phi) = s_j(L_x\phi) = s_j(L_y\phi') = Ls_y(\phi')$, $s_x(\phi) = s_x(\phi')$. By induction hypothesis $\phi \in T_i$ iff $\phi' \in T_i$. By Remark **A** $(L_x\phi \leftrightarrow L_y\phi') \in T_i$.

Because for each occurrence of a formula of the form $L_x\phi$ in ψ the equivalence $L_x\phi \leftrightarrow L_y\phi'$ belongs to T_i and T_i is closed under logical consequence, by substitution of equivalents we obtain: $\psi \in T_i$ if and only if $\psi' \in T_i$. □

Lemma 4.4 *Let $M_1, \ldots, M_n$ be models of the language $\mathcal{L}_{mae}$ and M a model of the language $\mathcal{L}_{ae}$ so that for propositional variables $p \in \mathcal{L}_{ae}$ and all $i \in \{1, \ldots, n\}$, $M \models p_i$ if and only if $M_i \models p$, and for all formulae $L_j\phi \in \mathcal{L}_{mae}$ and all $i \in \{1, \ldots, n\}$, $M \models Ls_j(\phi)$ if and only if $M_i \models L_j\phi$. Then for all $\chi \in \mathcal{L}_{mae}$, $M \models s_i(\chi)$ if and only if $M_i \models \chi$.*

Proof: We prove by induction on the structure of χ (starting from the model-theoretically atomic formulae $L\phi$ and p) that $M_i \models \chi$ iff $M \models s_i(\chi)$.

Base case: For propositional variables p, $s_i(p) = p_i$. By definition $M \models p_i$ iff $M_i \models p$. For formulae $L_j\psi$, $s_i(L_j\psi) = Ls_j(\psi)$. By definition $M \models Ls_j(\psi)$ iff $M_i \models L_j\psi$.

Inductive case: The formula is built using one of the connectives. The result is immediate by induction hypothesis and the truth-definition of connectives. □

Lemma 4.5 *Given any model M of the language $\mathcal{L}_{ae}$, models $M_1, \ldots, M_n$ of the language $\mathcal{L}_{mae}$ can be constructed so that for all $\chi \in \mathcal{L}_{mae}$, $M \models s_i(\chi)$ if and only if $M_i \models \chi$.*

Given models $M_1, \ldots, M_n$ such that for an extension $\mathcal{T} = \langle T_1, \ldots, T_n \rangle$ of some premises $\mathcal{A}$, $M_i \models T_i$ for all $i, 1 \leq i \leq n$, a model M of the language $\mathcal{L}_{ae}$ can be constructed so that for all $\chi \in \mathcal{L}_{mae}$, $M \models s_i(\chi)$ if and only if $M_i \models \chi$.

Proof: The construction of $M_1, \ldots, M_n$ from M in the first part is as given in Lemma 4.4. The result immediately follows from Lemma 4.4.

To prove the second part, we construct a model M from $M_1, \ldots, M_n$ so that $M \models p_i$ iff $M_i \models p$, and $M \models Ls_i(\phi)$ iff $M_j \models L_i\phi$ (for all j and $\phi \in \mathcal{L}_{mae}$). The well-definedness of the requirement $M \models Ls_i(\phi)$ iff $M_j \models L_i\phi$ is not immediate. We show that for all $j, k \in \{1, \ldots, n\}$ and $\phi \in \mathcal{L}_{ae}$ and $\phi' \in \mathcal{L}_{ae}$ such that $s_i(\phi) = s_i(\phi')$, $M_j \models L_i\phi$ iff $M_k \models L_i\phi'$. So assume that $s_i(\phi) = s_i(\phi')$. By Lemma 4.3 $\phi \in T_i$ iff $\phi' \in T_i$. Using definition of T_j, T_k, therefore $L_i\phi \in T_j$ iff $L_i\phi' \in T_k$. Because for all i, $M_i \models T_i$, $M_j \models L_i\phi$ iff $M_k \models L_i\phi'$. Hence the truth-value in M for atomic formulae beginning with L is well-defined. That $M \models s_i(\chi)$ iff $M_i \models \chi$ for all $\chi \in \mathcal{L}_{mae}$ is immediate by Lemma 4.4. □

Lemma 4.6 *Let $\mathcal{A} = \langle A_1, \ldots, A_n \rangle$ where each $A_i \subseteq \mathcal{L}_{mae}$. Then for each consistent stable expansion E of $tr_{ma}(A_1, \ldots, A_n)$ there is an extension $\mathcal{T} = \langle T_1, \ldots, T_n \rangle$ of $\mathcal{A}$ such that*

$$E \supseteq \{s_i(\phi) | \phi \in T_i, 1 \leq i \leq n\} = tr_{ma}(T_1, \ldots, T_n)$$

and for formulae $\phi \in \mathcal{L}_{mae}$, $s_i(\phi) \in E$ if and only if $\phi \in T_i$.

Proof: We claim that $T_i = \{\phi \in \mathcal{L}_{mae} | s_i(\phi) \in E\}$ for all $i \in \{1, \ldots, n\}$. We instantiate this in the equation of Definition 4.1 for all i and show that the equation is fulfilled. Now $\bigcup_{i=1}^n \{L_j\psi | \psi \in T_j\} = \bigcup_{i=1}^n \{L_j\psi | \psi \in \{\phi \in \mathcal{L}_{mae} | s_j(\phi) \in E\}\} = \bigcup_{i=1}^n \{L_j\psi | s_j(\psi) \in E\}$. Similarly for $\bigcup_{i=1}^n \{\neg L_j\psi | \psi \notin T_j\}$. Hence we have to show that for all $i, 1 \leq i \leq n$ and for all $\phi \in \mathcal{L}_{mae}$,

$$s_i(\phi) \in E \quad \text{iff} \quad A_i \cup \bigcup_{j=1}^n \{L_j\psi | s_j(\psi) \in E\} \cup \bigcup_{j=1}^n \{\neg L_j\psi | s_j(\psi) \notin E\} \models \phi.$$

By definition

$$s_i(\phi) \in E \quad \text{iff} \quad tr_{ma}(A_1, \ldots, A_n) \cup \{L\psi | \psi \in E\} \cup \{\neg L\psi | \psi \notin E\} \models s_i(\phi).$$

Let $\phi \in \mathcal{L}_{mae}$ be a formulae and $i \in \{1, \ldots, n\}$. We show the equivalence of the right-hand sides of the above two equivalences. ($\Rightarrow$) Define

$\Delta = tr_{ma}(A_1,\ldots,A_n) \cup \{L\psi|\psi \in E\} \cup \{\neg L\psi|\psi \notin E\} \cup \{\neg s_i(\phi)\}$. Now for all formulae χ in $\Gamma = A_i \cup \bigcup_{j=1}^n \{L_j\psi|s_j(\psi) \in E\} \cup \bigcup_{j=1}^n \{\neg L_j\psi|s_j(\psi) \notin E\} \cup \{\neg\phi\}$, $s_i(\chi) \in \Delta$. Assume there is a model M of Δ, i.e. $tr_{ma}(A_1,\ldots,A_n) \cup \{L\psi|\psi \in E\} \cup \{\neg L\psi|\psi \notin E\} \not\models s_i(\phi)$. Construct models $M_1,\ldots,M_n$ can as indicated in Lemma 4.5. Now $M_i \models \Gamma$. Hence $A_i \cup \bigcup_{j=1}^n \{L_j\psi|s_j(\psi) \in E\} \cup \bigcup_{j=1}^n \{\neg L_j\psi|s_j(\psi) \notin E\} \not\models \phi$. ($\Leftarrow$) Let M_i be a model of Γ, and M a model of $tr_{ma}(A_1,\ldots,A_n) \cup \{L\psi|\psi \in E\} \cup \{\neg L\psi|\psi \notin E\}$. Modify M so that $M \models p_i$ iff $M_i \models p$. Now $M \models \Delta$. Hence $tr_{ma}(A_1,\ldots,A_n) \cup \{L\psi|\psi \in E\} \cup \{\neg L\psi|\psi \notin E\} \models s_i(\phi)$ iff $A_i \cup \bigcup_{j=1}^n \{L_j\psi|s_j(\psi) \in E\} \cup \bigcup_{j=1}^n \{\neg L_j\psi|s_j(\psi) \notin E\} \models \phi$.

Hence for all formulae $\phi \in \mathcal{L}_{mae}$, $s_i(\phi) \in E$ if and only if $\phi \in T_i$. Note that E contains formulae χ such that $\chi = s_i(\phi)$ for no ϕ. For example, $(p_1 \vee \neg p_1) \wedge (p_2 \vee \neg p_2)$ is a tautology in E but it cannot be obtained from any formula in $\mathcal{L}_{mae}$ using s_i. □

Lemma 4.7 *Let $\mathcal{A} = \langle A_1,\ldots,A_n\rangle$ where each $A_i \subseteq \mathcal{L}_{mae}$. Then for each extension $\mathcal{T} = \langle T_1,\ldots,T_n\rangle$ of $\mathcal{A}$ such that each T_i is consistent, there is a stable expansion E of $tr_{ma}(A_1,\ldots,A_n)$ such that*

$$E \supseteq \{s_i(\phi)|\phi \in T_i, 1 \leq i \leq n\} = tr_{ma}(T_1,\ldots,T_n).$$

And for formulae $\phi \in \mathcal{L}_{mae}$, $s_i(\phi) \in E$ if and only if $\phi \in T_i$.

Proof: Define $\Lambda_{\mathcal{T}} = \{Ls_j(\phi)|\phi \in T_j, L_j\phi \in Sf^L(A_k), 1 \leq k \leq n\} \cup \{\neg Ls_j(\phi)|\phi \notin T_j, L_j\phi \in Sf^L(A_k), 1 \leq k \leq n\}$. We claim that $\Lambda_{\mathcal{T}}$ is $tr_{ma}(A_1,\ldots,A_n)$-full, i.e. for all $L\phi \in Sf^L(tr_{ma}(A_1,\ldots,A_n))$, $L\phi \in \Lambda_{\mathcal{T}}$ iff $tr_{ma}(A_1,\ldots,A_n) \cup \Lambda_{\mathcal{T}} \models \phi$ and $\neg L\phi \in \Lambda_{\mathcal{T}}$ iff $tr_{ma}(A_1,\ldots,A_n) \cup \Lambda_{\mathcal{T}} \not\models \phi$. Let $Ls_j(\phi) \in Sf^L(tr_{ma}(A_1,\ldots,A_n))$. Now

$$\begin{aligned} Ls_j(\phi) \in \Lambda_{\mathcal{T}} \quad &\text{iff} \quad \phi \in T_j, L_j\phi \in Sf^L(A_k), \text{ for some } k, 1 \leq k \leq n \\ &\text{iff} \quad A_j \cup \bigcup_{i=1}^n \{L_i\psi|\psi \in T_i\} \cup \bigcup_{i=1}^n \{\neg L_i\psi|\psi \notin T_i\} \models \phi, L_j\phi \in Sf^L(A_k) \\ &\text{iff} \quad tr_{ma}(A_1,\ldots,A_n) \cup \Lambda_{\mathcal{T}} \models s_j(\phi), L_j\phi \in Sf^L(A_k). \end{aligned}$$

First two equivalences are by definition. We prove the third one. Let $M_1,\ldots,M_n$ be models of $T_1,\ldots,T_n$. Let $L_j\phi \in Sf^L(A_k)$ for some $k \in \{1,\ldots,n\}$.

Let M_j be such that $M_j \models A_j \cup \bigcup_{i=1}^n \{L_i\psi|\psi \in T_i\} \cup \bigcup_{i=1}^n \{\neg L_i\psi|\psi \notin T_i\} \cup \{\neg\phi\}$, for all $j \in \{1,\ldots,n\}$. Let M be a model constructed from $M_1,\ldots,M_n$ as stated in Lemma 4.5. Because all formulae in $tr_{ma}(A_1,\ldots,A_n) \cup \Lambda_{\mathcal{T}} \cup \{\neg s_j(\phi)\}$ are $s_i(\psi)$ for some $i \in \{1,\ldots,n\}$ and $\psi \in T_i$, by Lemma 4.4 $M \models tr_{ma}(A_1,\ldots,A_n) \cup \Lambda_{\mathcal{T}} \cup \{\neg s_j(\phi)\}$, and hence $tr_{ma}(A_1,\ldots,A_n) \cup \Lambda_{\mathcal{T}} \not\models s_j(\phi)$.

Assume $M \models tr_{ma}(A_1, \ldots, A_n) \cup \Lambda_{\mathcal{T}} \cup \{\neg s_j(\phi)\}$. Modify model M_j so that $M_j \models p$ iff $M \models p_j$. All formulae in $\bigcup_{i=1}^n \{L_i\psi | \psi \in T_i\} \cup \bigcup_{i=1}^n \{\neg L_i\psi | \psi \notin T_i\}$ are true in M_j by definition. All formulae χ in $A_j \cup \{\neg\phi\}$ are such that $s_j(\chi) \in tr_{ma}(A_1, \ldots, A_n) \cup \{s_j(\phi)\}$, hence by Lemma 4.4 $M_j \models A_j \cup \{\neg\phi\}$, and $A_j \cup \bigcup_{i=1}^n \{L_i\psi | \psi \in T_i\} \cup \bigcup_{i=1}^n \{\neg L_i\psi | \psi \notin T_i\} \not\models \phi$.

Hence $\Lambda_{\mathcal{T}}$ is $tr_{ma}(A_1, \ldots, A_n)$-full. By Lemma 4.6 $E = SE_{tr_{ma}(A_1,\ldots,A_n)}(\Lambda_{\mathcal{T}}) \supseteq \{s_i(\phi) | \phi \in T_i, 1 \leq i \leq n\}$, and for all formulae $\phi \in \mathcal{L}_{mae}$ and all $i \in \{1, \ldots, n\}$, $s_i(\phi) \in E$ if and only if $\phi \in T_i$. □

Theorem 4.8 *Let $\phi \in \mathcal{L}_{mae}$ be a formula and $\mathcal{A} = \langle A_1, \ldots, A_n\rangle$. Assume that for all extensions $\mathcal{T} = \langle T_1, \ldots, T_n\rangle$ of $\mathcal{A}$ each T_i is consistent and that all stable expansions of $tr_{ma}(A_1, \ldots, A_n)$ are consistent. Then $\phi \in T_i$ for all extensions $\mathcal{T}$ of $\mathcal{A}$ if and only if $s_i(\phi) \in E$ for all stable expansions E of $tr_{ma}(A_1, \ldots, A_n)$.*

Proof: Immediate by Lemmata 4.6 and 4.7. □

4.2 Priorities in default logic

4.2.1 Hierarchical default logic

MacNish [1991] defines *transition relations* R_D, where D is a set of default rules, between sets of formulae as follows.

$$W_1 R_D W_2 \text{ iff } W_2 \text{ is an extension of } \langle D, W_1\rangle$$

A *hierarchical parameter set* is a tuple $H_D = \langle\{D_1, \ldots\}, \preceq\rangle$ where D_i are sets of defaults and $\preceq$ is a total ordering of $\{D_1, \ldots\}$. This can be seen as a prioritization of default rules in $D = \bigcup_{i \in \{1,\ldots\}} D_i$. The rules in D_{i+1} have a higher priority than the rules in D_i. Prioritized extensions E of $\langle D, W\rangle$ are obtained using the relation composition $R_D = \cdots \circ R_{D_3} \circ R_{D_2} \circ R_{D_1}$ by $W R_D E$.

MacNish's extensions are not extensions as defined by [Reiter, 1980]. Consider $\Delta = \langle D, W\rangle$ where $D = D_1 \cup D_2$, $D_1 = \{a : b/b\}$, $D_2 = \{b : c/c\}$, $W = \{a\}$ and a hierarchical parameter set that prioritizes D_2 over D_1 is $H_D = \langle\{D_1, D_2\}, \preceq\rangle$ where $D_1 \preceq D_2$. Define $Cn(W) = \{\phi \in \mathcal{L} | W \models \phi\}$. Now $W R_{D_2} W_1$, where $W_1 = W = Cn(\{a\})$, and $W_1 R_{D_1} W_2$, where $W_2 = Cn(\{a, b\})$. The composition of these relations R_D is $\{(Cn(\{a\}), Cn(\{a, b\}))\}$ and $W R_D W_2$. The set W_2 is not an extension of Δ because the rule $b : c/c$ – although applicable –

has not been applied. MacNish's extensions are Reiter's extensions when the default rules are prerequisite-free and normal.

4.2.2 Prioritized default logic

Brewka [1991] mentions the possibility of ordering default rules in layers, and applying default rules with a higher priority before rules with a lower priority. He, however, restricts his attention to normal defaults without prerequisites because of the reasons that were already mentioned in the discussion of MacNish's work. He states the equivalence of this subset to his preferred subtheories [Brewka, 1989] which we discuss in Section 4.3.

4.2.3 Constructive default logic

This logic, which does not make any syntactic restrictions to the default rules, is defined in [Tan and Treur, 1992]. Extensions of default theories are constructed iteratively using the semiconstructive definition of extensions given in [Reiter, 1980], by applying only a subset of those default rules for which the prerequisite is derivable and the negations of justifications are not. This subset is determined by a selection function. The selection function has to fulfill the property that none of the negations of justifications of rules that have been applied, later become derivable, i.e. the selection function has to be non-self-defeating, and that each applicable default rule will be selected in some step. In this approach the source of computational complexity is the construction of non-self-defeating selection functions.

We would like to show that for a given selection function σ and a default theory D we could construct a prioritization P so that the extension of D determined by σ would be a unique P-preferred stable expansion of a translation of D to autoepistemic logic. The definition of acceptable selection functions does not rule out the selection of certain default conclusions in several steps of the computation of an extension. Hence the construction of a prioritization from a selection function alone (without actually simulating the computation of the extension using the selection function) does not seem to be possible. Extensions of default theories can be uniquely determined by prioritizations where the ordering of formulae is a total order, as shown by Lemma 2.10. Tan and Treur motivate their definition of constructive default logic by proposing that it could be useful that the selection function could be defined dynamically during the reasoning process.

4.2.4 Using our priorities in default logic

The discussion of Brewka's prioritized default logic PDL and MacNish's hierarchical default logic has indicated that applying default rules exactly in the order determined by priorities is not a valid method for constructing extensions of default logic. Such a method works only in the special case of prerequisite-free normal defaults, and as suggested by the example in the beginning of Section 2, more sophisticated methods that are still based on the order of application of default rules, are likely to produce inintuitive results. The priorities for autoepistemic logic presented in Section 2 can be used in default logic as well. Default logic can be embedded in autoepistemic logic by translating default rules $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$ into formulae $L\alpha \wedge L\neg L\neg\beta_1 \wedge \cdots \wedge L\neg L\neg\beta_n \rightarrow \gamma$ (translation proposed by [Truszczyński, 1991]), and restricting to a suitable subclass of stable expansions, for example to the L-hierarchic expansions [Niemelä, 1991] or N-expansions [Truszczyński, 1991]. Our priority notion can be brought to default logic through this translation. Our priorities apply to formulae only, but we can also prioritize default rules directly. There are at least two distinct ideas what priorities on default rules could be. The first is that giving a default rule $\frac{\alpha:\beta}{\gamma}$ a high priority means that the extensions containing γ are preferred to extensions not containing it. This is directly expressible as our priorities. The second possibility (more of the flavour of priorities as the order of applying default rules) is to require that for a high-priority default rule $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$ the preferred extensions contain the formula α and do not contain the formulae β_i. This is also expressible as our priorities. Consider the rules $D_1 = \frac{a:b}{c}$ and $D_2 = \frac{d:e}{f}$. Give D_1 a higher priority than to D_2. This corresponds to a prioritization $\mathcal{P}$ for which $\neg La \vee L\neg b \mathcal{P} \neg Ld \vee L\neg e$, i.e. prefer stable expansions that do not contain $\neg La \vee L\neg b$ or equivalently, stable expansions that do contain both La and $\neg L\neg b$. In Section 6 we discuss automated reasoning in prioritized autoepistemic logic which can be applied to default logic as well using the translations mentioned in this section.

4.3 Preferred subtheories

Poole's [1988] framework for default reasoning is based on sets of premises and maximal consistent subsets of sets of default formulae. Brewka [1989] gives two generalizations of this framework by introducing priorities to select preferred maximal consistent subsets of defaults. The first generalization allows arbitrary many levels of formulae of different plausibility, while in Poole's system there were only two levels. The second further generalization does not require that the formulae are divided into levels, but arbitrary partial orders of formulae are allowed. We show that preferred subtheories can be exactly represented in prioritized autoepistemic logic.

Definition 4.9 ([Brewka, 1989]) *Let $T = (T_1, \ldots, T_n)$, where each $T_i \subseteq \mathcal{L}$ is a set of formulae, be a level default theory. The set $S = S_1 \cup \cdots \cup S_n$, where $S_i \subseteq T_i$ for all $i \in \{1, \ldots, n\}$, is a preferred subtheory of T if and only if for all k $(1 \leq k \leq n)$ $S_1 \cup \cdots \cup S_k$ is a maximal consistent subset of $T_1 \cup \cdots \cup T_k$.*

Brewka generalizes the above definition of preferred subtheories to arbitrary partial orders.

Definition 4.10 ([Brewka, 1989]) *Let $<$ be a strict partial order on a (finite) set of formulae $T \subseteq \mathcal{L}$. The set S is a preferred subtheory of $\langle T, < \rangle$ if and only if there exists a strict total ordering $t_1, t_2, \ldots, t_n$ of T respecting $<$ such that $S = S_n$ with*

$$
\begin{array}{rcl}
S_0 & = & \emptyset \\
S_{i+1} & = & \left\{ \begin{array}{ll} S_i \cup \{t_{i+1}\} & \textit{if } S_i \cup \{t_{i+1}\} \textit{ is consistent} \\ S_i & \textit{otherwise} \end{array} \right. \qquad \textit{for } i, 0 \leq i < n.
\end{array}
$$

That this definition really is a generalization is shown by Lemma 4.14. All results about the connection between prioritized autoepistemic logic and Brewka's preferred subtheories are through the more general version of Brewka's system.

Assumption 4.11 *The levels $T_1, \ldots, T_n$ of a level default theory $T = (T_1, \ldots, T_n)$ theory are disjoint.*

This assumption is not particularly restrictive, as shown by the following construction.

Lemma 4.12 *Let $T = (T_1, \ldots, T_n)$ be a level default theory. Let $T' = (T'_1, \ldots, T'_n)$ where each $T'_i, 1 \leq i \leq n$ is defined as $T'_i = T_i \setminus \bigcup_{j=1}^{i-1} T_j$. Each preferred subtheory of T is a preferred subtheory of T', and vice versa.*

Proof: Let $T = (T_1, \ldots, T_n)$ and $T' = (T'_1, \ldots, T'_n)$. We show that if $S = S_1 \cup \cdots \cup S_n$ where each $S_i \subseteq T_i$, is a preferred subtheory of T, then S is a preferred subtheory of T', and vice versa. The proof is by induction on $i, 1 \leq i \leq n$ that $S_1 \cup \cdots \cup S_i$ is maximal consistent subset of $\bigcup_{j=1}^{i} T'_i$ (or in the second case a maximal consistent subset of $\bigcup_{j=1}^{i} T_i$).

Base case $i = 1$: Result is immediate because $T_1 = T'_1$.

Inductive case $i \geq 1$: Let $S_1 \cup \cdots \cup S_i$ be a maximal consistent subset of $T_1 \cup \cdots \cup T_i$. Clearly $S_1 \cup \cdots \cup S_i$ must be a maximal subset of $T'_1 \cup \cdots \cup T'_i$ because $T'_1 \cup \cdots \cup T'_i = T_1 \cup \cdots \cup T_i$.

Let $S_1 \cup \cdots \cup S_i$ be a maximal consistent subset of $T'_1 \cup \cdots \cup T'_i$. Assume $S_1 \cup \cdots \cup S_i$ is not a maximal subset of $T_1 \cup \cdots \cup T_i$, i.e. for some $e \notin S_1 \cup \cdots \cup S_i$, $e \in T_i \backslash T'_i$ the set $S_1 \cup \cdots \cup S_i \cup \{e\}$ is consistent. Because $e \in T_i \backslash T'_i$, by definition of T'_i must $e \in \bigcup_{j=1}^{i-1} T_j$, and because $e \notin S_1 \cup \cdots \cup S_{i-1}$, should $S_1 \cup \cdots S_{i-1} \cup \{e\}$ by induction hypothesis be inconsistent. This is a contradiction because supersets of inconsistent sets are inconsistent. □

Lemma 4.13 *Let T be a set of formulae ordered as $t_1, \ldots, t_n$. In the recursion equations of Definition 4.10 each $S_i, 1 \leq i < n$ is a maximal consistent subset of $\{t_j | 1 \leq j \leq i\}$.*

Proof: Proof is by induction on $i, 1 \leq i \leq n$. Base case $i = 0$: The empty set is clearly a maximal consistent subset of itself.

Inductive case $i \geq 1$: By induction hypothesis S_{i-1} is a maximal consistent subset of $\{t_1, \ldots, t_{i-1}\}$. If $S_{i-1} \cup \{t_i\}$ is inconsistent, then $S_i = S_{i-1}$, and hence it is consistent. The set S_i is maximal because t_i clearly cannot be added to it without making it inconsistent and by induction hypothesis none of $t_1, \ldots, t_{i-1}$ cannot be added to $S_{i-1} = S_i$. If $S_{i-1} \cup \{t_i\}$ is consistent, then clearly $S_i = S_{i-1} \cup \{t_i\}$ is also consistent. By induction hypothesis none of $t_1, \ldots, t_{i-1}$ can be added to $S_i \supseteq S_{i-1}$ without making it inconsistent. □

The next lemma shows that Brewka's level default theories are a special case of his preferred subtheories with arbitrary strict partial orders.

Lemma 4.14 *Let $(T_1, \ldots, T_n)$ be a level default theory. Define $tr_{pp}(T_1, \ldots, T_n) = \{(t, t') | t \in T_i, t' \in T_j, i < j\}$. Then Z is a preferred subtheory of $\langle T_1 \cup \cdots \cup T_n, tr_{pp}(T_1, \ldots, T_n)\rangle$ if and only if it is a preferred subtheory of the level default theory $(T_1, \ldots, T_n)$.*

Proof: By Assumption 4.11 $tr_{pp}(T_1, \ldots, T_n)$ is a strict partial order on $U = T_1 \cup \cdots \cup T_n$. Let Z be a preferred subtheory of $\langle U, tr_{pp}(T_1, \ldots, T_n)\rangle$. Hence there is a strict total ordering $t_1, \ldots, t_n$ of U that extends $tr_{pp}(T_1, \ldots, T_n)$, as defined in Definition 4.10. We show that for all $i, 0 \leq i \leq n$, $Z \cap (\bigcup_{j=1}^{i} T_j)$ is a maximal consistent subset of $\bigcup_{j=1}^{i} T_j$. The proof is by induction on i.

Base case $i = 0$: Immediate.

Inductive case $i > 0$: Let $t_1, \ldots, t_k$ be the prefix of the total ordering of U mentioned above restricted to $T_1 \cup \cdots \cup T_i$. By Lemma 4.13 the set S_k is a maximal consistent subset of $T_1 \cup \cdots \cup T_i$. This completes the first part of the proof.

Let $Z = Z_1 \cup \cdots \cup Z_n$ be a preferred subtheory of $T = (T_1, \ldots, T_n)$. Let $<'$ be a strict total order on U such that $t <' t'$ if $t \in T_i, t' \in T_j, i < j$, or $t \in T_i, t' \in T_i$ for some i and $t \in Z, t' \notin Z$. The existence of such an ordering is obvious. We show that the set S obtained from the equations of Definition 4.10 using $<'$ coincides with Z. We prove by induction on i that for all $i \geq 1, S_i \subseteq Z$ and $(\bigcup_{j=1}^{i}\{t_j\}\backslash S_i) \cap Z = \emptyset$.

Base case $i = 1$: Assume $t_1 \in Z$. Then the set $\{t_1\} = S_0 \cup \{t_1\}$ is consistent and hence $S_1 = \{t_1\}$. Assume $t_1 \in S_1$. Then $\{t_1\}$ is consistent. Because t_1 is the $<'$-least element of T_1 and it is consistent, it must – by the definition of $<'$ – belong to Z_1.

Inductive case $i > 1$: By induction hypothesis $S_{i-1} \subseteq Z$ and $(\bigcup_{j=1}^{i-1}\{t_j\}\backslash S_{i-1}) \cap Z = \emptyset$. Assume $t_i \in Z$. Because $S_{i-1} \cup \{t_i\} \subseteq Z$, and Z is consistent, $S_{i-1} \cup \{t_i\}$ is consistent and $S_i = S_{i-1} \cup \{t_i\}$. Assume $t_i \notin Z$. Let $k, 1 \leq k \leq n$ be such that $t_i \in T_k$. Now by induction hypothesis all $t \in Z \cap T_k$ also belong to S_{i-1}, because for such t by construction $t <' t_i$. Because $Z \cap \bigcup_{j=1}^{k} T_j$ is a maximal consistent subset of $\bigcup_{j=1}^{k} T_j$, it is inconsistent with t_i. Hence $S_{i-1} \cup \{t_i\}$ is inconsistent, and by definition of S_i, $t_i \notin S_i$. □

The decision problems for preferred subtheories are *strong provability* that means logical consequence from all preferred subtheories of a theory, and *weak provability* that means logical consequence from some preferred subtheory of a theory. These notions correspond to cautious and brave reasoning in autoepistemic logic, respectively. The following lemmata are for establishing the connection between preferred subtheories and prioritized autoepistemic logic. The first one is a special case of a lemma proved in [Niemelä and Rintanen, 1992].

Lemma 4.15 *Let Σ be set of formulae of the form $\neg L\chi \rightarrow \psi'$ where $\chi, \psi \in \mathcal{L}$, and let Λ be a set of formulae of the form $L\phi, \neg L\phi$ that contains exactly one of $L\chi$, $\neg L\chi$ for each $L\chi \in Sf^L(\Sigma)$. Define $Red(\Sigma, \Lambda) = \{\psi | (\neg L\chi \rightarrow \psi) \in \Sigma, \neg L\chi \in \Lambda\}$. Assume that for all $L\phi, \neg L\phi' \in \Lambda$, $\Sigma \cup \Lambda \models \phi$ and $\Sigma \cup \Lambda \not\models \phi'$. Then for all $\chi \in \mathcal{L}$,*

$$Red(\Sigma, \Lambda) \models \chi \textit{ iff } \Sigma \cup \Lambda \models \chi.$$

Lemma 4.16 *Let $T \subseteq \mathcal{L}$ be a set of formulae and $\Sigma = \{\neg L\neg\phi \rightarrow \phi | \phi \in T\}$. Then for all $\phi \in T$, and all stable expansions E of Σ, $\neg\phi \notin E$ if and only if $\phi \in E$.*

Proof: Let Λ be a Σ-full set such that $E = SE_\Sigma(\Lambda)$.

$\neg\phi \notin E$	iff	$\Sigma \cup \Lambda \not\models \neg\phi$	Properties of full sets
	iff	$\neg L\neg\phi \in \Lambda$	Properties of full sets, $L\neg\phi \in Sf^L(\Sigma)$
	implies	$\Sigma \cup \Lambda \models \phi$	Because $\neg L\neg\phi \to \phi \in \Sigma$
	iff	$\phi \in E$	By properties of full sets

The converse of the implication above can be shown as follows. Assume $\Sigma \cup \Lambda \models \phi$ and $\neg L\neg\phi \notin \Lambda$. Now $L\neg\phi \in \Lambda$ and $\Sigma \cup \Lambda \models \neg\phi$. Hence $\Sigma \cup \Lambda \models \bot$, i.e it is inconsistent, and for all $\phi \in T$ $L\phi \in \Lambda$. By Lemma 4.15 $\mathit{Red}(\Sigma, \Lambda) \models \bot$, i.e. $\emptyset \models \bot$. This is a contradiction. Hence the missing implication in the above chain of equivalences holds. □

Lemma 4.17 *Let S be a maximal consistent subset of $T \subseteq \mathcal{L}$. Then for $\Sigma = \{\neg L\neg\phi \to \phi | \phi \in T\}$ and $\Lambda = \{\neg L\neg\phi | \phi \in S\} \cup \{L\neg\phi | \phi \in T \backslash S\}$, Λ is Σ-full and $SE_\Sigma(\Lambda) \cap \mathcal{L} = \{\phi \in \mathcal{L} | S \models \phi\}$. And for all Σ-full sets Λ', $\{\phi | \neg L\neg\phi \in \Lambda'\}$ is a maximal consistent subset of T.*

Proof: First we show that Λ is Σ-full. That $\Lambda \subseteq Sf^L(\Sigma) \cup \{\neg L\phi | L\phi \in Sf^L(\Sigma)\}$ is immediate from the definition of Λ. The third condition on full sets is immediate after we show the second one, i.e. for all $L\chi \in Sf^L(\Sigma)$, $L\chi \in \Lambda$ iff $\Sigma \cup \Lambda \models \chi$. Suppose $L\neg\phi \in Sf^L(\Sigma)$.

$L\neg\phi \in \Lambda$	iff	$\phi \in T \backslash S$	By definition of Λ
	iff	$S \models \neg\phi$	By $\phi \in T$ and maximality of S
	iff	$\mathit{Red}(\Sigma, \Lambda) \models \neg\phi$	Because $\mathit{Red}(\Sigma, \Lambda) = S$
	iff	$\Sigma \cup \Lambda \models \neg\phi$	By Lemma 4.15

$SE_\Sigma(\Lambda) \cap \mathcal{L} = \{\phi \in \mathcal{L} | S \models \phi\}$ because $\mathit{Red}(\Sigma, \Lambda) = S$ and $SE_\Sigma(\Lambda) \cap \mathcal{L}$ is the closure of $\Sigma \cup \Lambda$ under logical consequence.

Given T, let $\Sigma = \{\neg L\neg\phi \to \phi | \phi \in T\}$ and let Λ be a Σ-full set. Let $S = \{\phi | \neg L\neg\phi \in \Lambda\}$. We show that S is a maximal consistent subset of T. Obviously $S \subseteq T$. Suppose S is inconsistent, i.e. $S \models \bot$. Then because for all $\phi \in S$, $\neg L\neg\phi \to \phi \in \Sigma$ and $\neg L\neg\phi \in \Lambda$, $\Sigma \cup \Lambda \models \phi$, and further $\Sigma \cup \Lambda \models \bot$. This however contradicts the assumption that Λ is Σ-full, because $\neg L\chi \in \Lambda$ iff $\Sigma \cup \Lambda \not\models \chi$. Suppose there is a consistent set S' such that $S \subset S' \subseteq T$. This means that there would be a formula $\phi \in S' \backslash S$ such that $S' \not\models \neg\phi$ and by monotonicity of $\models$ also $S \not\models \neg\phi$. Because $L\neg\phi \in \Lambda$ and $\mathit{Red}(\Sigma, \Lambda) = S$, by Σ-fullness of Λ, $\Sigma \cup \Lambda \models \neg\phi$ and $S \models \neg\phi$, which is a contradiction. □

Definition 4.18 *The translation of preferred subtheories to prioritized autoepistemic logic is based on the following functions.*

$$\begin{array}{rcl} tr_{ps}(T) & = & \{\neg L\neg\phi \rightarrow \phi | \phi \in T\} \\ p_{ps}(T,<) & = & \langle\{\neg\phi|\phi \in T\}, \{(\neg\phi,\neg\psi)|\phi \in T, \psi \in T, \phi < \psi\}\rangle \end{array}$$

Lemma 4.19 *If S is a preferred subtheory of $\langle T,<\rangle$, then there is a $p_{ps}(T,<)$-preferred stable expansion E of $tr_{ps}(T)$ such that $E \cap \mathcal{L} = \{\phi \in \mathcal{L}|S \models \phi\}$.*

Proof: Let S be a preferred subtheory of $\langle T,<\rangle$ and let $t_1,\ldots,t_n$ be the total ordering of T using which S is obtained. Let $\langle\Phi,\mathcal{P}\rangle = p_{ps}(T,<)$. By Assumption 4.11 it is a prioritization. Take $\mathcal{T}$ to be the total ordering $\neg t_1,\ldots,\neg t_n$. Clearly $\mathcal{T} \supseteq \mathcal{P}$. We claim that $E = SE_\Sigma(\Lambda)$, where $\Lambda = \{\neg L\neg\phi|\phi \in S\} \cup \{L\neg\phi|\phi \in T\backslash S\}$ and $\Sigma = tr_{ps}(T)$, is a $p_{ps}(T,<)$-preferred stable expansion, i.e. for all stable expansions E' of Σ,

$$\text{for all } \phi \in \Phi(\phi \in E\backslash E' \text{ implies there is } \psi \in E'\backslash E \text{ such that } \psi\mathcal{T}\phi).$$

Suppose this is not so, i.e. for some stable expansion E' there is $\neg t_k \in E\backslash E'$ and for no $l \leq k$ $t_l \in E'\backslash E$. Let Λ' be a Σ-full set such that $E' = SE_\Sigma(\Lambda')$. Now by Lemma 4.17 the set $S' = \{\phi|\neg L\neg\phi \in \Lambda'\}$ is a maximal consistent subset of T. Let $\neg t_k$ be the $\mathcal{T}$-least element of Φ such that $\neg t_k \in E\backslash E'$. By Lemma 4.16 $t_k \in E'\backslash E$. Because S' is consistent, its subset $\{t_i|i < l, t_i \in T\} \cup \{t_k\}$ should also be, but this contradicts the fact that t_k was not included in S due to the inconsistency of $\{t_i|i < l, t_i \in T\} \cup \{t_k\}$. Hence the assumption that E is not $p_{ps}(T,<)$-preferred is refuted. □

Lemma 4.20 *If E is a $p_{ps}(T,<)$-preferred stable expansion of $tr_{ps}(T)$, then there is a preferred subtheory S of $\langle T,<\rangle$ such that $E \cap \mathcal{L} = \{\phi \in \mathcal{L}|S \models \phi\}$.*

Proof: By Assumption 4.11 $p_{ps}(T,<)$ is a prioritization. Let E' be a $p_{ps}(T,<)$-preferred stable expansion of $tr_{ps}(T)$. We claim that $S' = E' \cap T$ is a preferred subtheory of $\langle T,<\rangle$. Let $\langle\Phi,\mathcal{P}\rangle = p_{ps}(T,<)$. Let $\neg t_1,\ldots,\neg t_n$ be the total ordering of Φ that respects $\mathcal{P}$ and selects E'. We claim that taking $t_1,\ldots,t_n$ to be the extension of $<$ to a total order yields S'. Suppose it does not, i.e., at some step k in the computation of equations in Definition 4.10, $t_k \in S'$ is not included in S_k or $t_k \notin S'$ is included in S_k. Take k be the first such step, i.e. $S_{k-1} = \{t_i|1 \leq i < k\} \cap S'$.

First consider the case $t_k \in S'\backslash S$. The set $(\{t_i|i < k\} \cap S') \cup \{t_k\}$ is consistent because it is a subset of S' and S' is consistent by Lemma 4.17. Hence by construction $S_k = S_{k-1} \cup \{t_k\} = (\{t_i|i < k\} \cap S') \cup \{t_k\}$, and as $S_k \subseteq S$ a contradiction is obtained: $t_k \in S$.

Then consider the case $t_k \in S \backslash S'$. By Lemma 4.13 S' is a maximal consistent subset of T. By Lemma 4.17 the set $\Lambda' = \{\neg L \neg t | t \in S\} \cup \{L \neg t | t \in T \backslash S\}$ is a $tr_{ps}(T)$-full set. We show that the existence of the stable expansion $E = SE_{tr_{ps}(T)}(\Lambda')$ contradicts the $p_{ps}(T, <)$-preferredness of E'. Because $t_k \in E \backslash E'$, by Lemma 4.16 $\neg t_k \in E' \backslash E$. As $\neg t_k \in \Phi$, the preferredness of E' would require the existence of a ψ such that $\psi \mathcal{T} \neg t_k$ and $\psi \in E \backslash E'$. This however is impossible because t_k was chosen to be the least element on which S and S' differ. Hence the assumption $S \neq S'$ is refuted.

Because S is a maximal consistent subset of T and $E' \cap T = S$, by Lemma 4.17 $E' \cap \mathcal{L} = \{\phi \in \mathcal{L} | S \models \phi\}$. □

Theorem 4.21 *A formula $\phi \in \mathcal{L}$ is strongly provable from $T \subseteq \mathcal{L}$ if and only if ϕ is in all $p_{ps}(T, <)$-preferred stable expansions of $tr_{ps}(T)$, and ϕ is weakly provable from T if and only if ϕ is in some $p_{ps}(T, <)$-preferred stable expansion of $tr_{ps}(T)$.*

Proof: By Assumption 4.11 $p_{ps}(T, <)$ is a prioritization. A formula $\phi \in \mathcal{L}$ is strongly provable from T if and only if $S \models \phi$ for all preferred subtheories S of T if and only if ϕ is in all $p_{ps}(T, <)$-preferred stable expansions of $tr_{ps}(T)$. The first *iff* is by the definition of strong provability. The second *iff* is because the intersection of preferred subtheories of T is the same set as the intersection of the objective parts of the $p_{ps}(T, <)$-preferred stable expansions of $tr_{ps}(T)$. This claim we show as follows. Let $\phi \in \mathcal{L}$ be a formula such that there is a preferred subtheory S of T and $S \not\models \phi$. Now by Lemma 4.19 there is a $p_{ps}(T, <)$-preferred stable expansion E such that $\phi \notin E$. Conversely, let $\phi \in \mathcal{L}$ be a formula such that there is a $p_{ps}(T, <)$-preferred stable expansion E and $\phi \notin E$. Now by Lemma 4.20 there is a preferred subtheory S of T such that $S \not\models \phi$. Hence the union of the complements of the sets $\{\phi \in \mathcal{L} | S \models \phi\}$ and $\{\phi \in \mathcal{L} | E \models \phi\}$ are equal. By elementary set theory the intersections therefore coincide.

A formula $\phi \in \mathcal{L}$ is weakly provable from T if and only if $S \models \phi$ for some preferred subtheory S of T if and only if ϕ is in some $p_{ps}(T, <)$-preferred stable expansion of $tr_{ps}(T)$. The first *iff* is by definition, and the second is directly by Lemmata 4.19 and 4.20. □

The computation of maximal consistent subsets of an inconsistent set of formulae is computationally fairly complex in the general case. It is obvious that to achieve tractability the classical reasoning has to be restricted to a tractable subset of propositional logic, e.g. Horn clauses or 2-literal clauses.

With this restriction it is easy (quadratic time) to find some maximal consistent subset (by repeatedly adding each formula in order of plausibility to the theory if consistency is preserved), but for example the problem of whether there is a preferred subtheory from which some formula ϕ follows, is still difficult. The easy procedure is the desired one if the order of adding formulae is uniquely determined by the prioritization. This is the case when the ordering on the formulae T is total.

Theorem 4.22 *For the class of theories $\langle <, T \rangle$, where $T \subseteq \mathcal{L}$ is a set of propositional Horn clauses and $<$ is a strict total order on T, there is exactly one preferred subtheory, and the subtheory can be computed in polynomial time on the size of T.*

Proof: By Theorem 4.21 the decision problems of preferred subtheories map to the decision problems of prioritized autoepistemic logic by the translation given in Definition 4.18. A set T consisting of Horn clauses and a total $<$ map to a set of formulae Σ of the form $\neg L \neg \phi \rightarrow \phi$ where ϕ is a Horn clause, and to a prioritization $P = \langle \Phi, \mathcal{T} \rangle$ where $\Phi = \{\phi | L\phi \in Sf^{qL}(\Sigma)\}$ and $\mathcal{T}$ is a total order. The translation is clearly polynomial time. By Theorem 3.3 the decision problems can be solved in polynomial time. $\square$

4.4 Prioritized circumscription

In this section we show how a propositional version of prioritized parallel circumscription can be embedded in our prioritized autoepistemic logic. A similar embedding for a version of propositional parallel circumscription (that is, without priorities) is presented in [Niemelä, 1992b]. Circumscription is based on the idea of minimal models. Applications involving defeasible reasoning can be formalized using a number of first-order formulae that may contain abnormality predicates. For example, the formula

$$\forall x \; bird(x) \wedge \neg ab_{fly}(x) \rightarrow fly(x)$$

states that birds fly unless they are abnormal, or, birds typically fly. In classical predicate logic we should explicitly state that for every typical bird x, x is not abnormal with respect to flying. To exclude models in which unnecessary abnormalities are present, consequence relation of circumscription does not consider all models of a set of formulae but only ones minimal with respect to the extents of some set of predicates; for example a designated set of abnormality predicates. Roughly, a model of a theory is minimal if there are no models for which the extents of the predicates to be minimized are smaller. The exact definition of this depends on the version of circumscription. Prioritized parallel circumscription [McCarthy, 1986; Lifschitz, 1985] was introduced to support the statement of specificity, which is important in formalizing inheritance reasoning. The following is a model-theoretic definition for a propositional version of prioritized circumscription inspired by [Lifschitz, 1985; Baker and Ginsberg, 1989].

Definition 4.23 *Let $\langle \Delta, (D_1, \ldots, D_n), F \rangle$ be a circumscriptive theory, where Δ is a consistent set of formulae and $D_1, \ldots, D_n, F$ are finite sets of propositional variables. Define a partial order $<$ on models of Δ as $M_1 < M_2$, if the following condition holds.*

> *For some i $(1 \leq i \leq n)$ $\{p \in D_i | M_1 \models p\} \subset \{p \in D_i | M_2 \models p\}$ and $\{p \in F | M_1 \models p\} = \{p \in F | M_2 \models p\}$, and for all j $(1 \leq j < i)$ $\{p \in D_j | M_1 \models p\} = \{p \in D_j | M_2 \models p\}$.*

The models that are minimal in the partial order $<$ are the minimal models using which consequences of a theory are defined.

Definition 4.24 *Define (following [Niemelä, 1992b])*

$$\begin{aligned} \neg P &= \{\neg\phi | \phi \in P\} \\ MIN(P) &= \{\neg L\phi \rightarrow \neg\phi | \phi \in P\} \\ FIX(P) &= MIN(P) \cup MIN(\neg P) \end{aligned}$$

and define the following functions for translating circumscriptive theories $T = \langle \Delta, (D_1, \ldots, D_n), F \rangle$ to prioritized autoepistemic logic.

$$\begin{aligned} tr_{circ}(T) &= \Delta \cup MIN(\bigcup_{i=1}^{n} D_i) \cup FIX(F) \\ p_{circ}(T) &= \langle \neg F \cup F \cup \bigcup_{i=1}^{n} D_i, \{(\delta, \gamma) | \delta \in D_i, \gamma \in D_j, i < j\} \\ &\quad \cup \{(\neg\delta, \gamma) | \delta \in F, \gamma \in \bigcup_{i=1}^{n} D_i\} \cup \{(\delta, \gamma) | \delta \in F, \gamma \in \bigcup_{i=1}^{n} D_i\}\rangle \end{aligned}$$

Theorem 4.25 *Let $T = \langle \Delta, (D_1, \ldots, D_n), F \rangle$ be a circumscriptive theory with priorities such that Δ is consistent, and let $\phi \in \mathcal{L}$ be a formula. Then ϕ is true in every $<$-minimal model of Δ if and only if ϕ belongs to every $p_{circ}(T)$-preferred stable expansion of $tr_{circ}(T)$.*

Proof: We prove the result indirectly by establishing a connection between prioritized circumscription and level default theories (Section 4.3). Define

$$tr_{cp}(\Delta, (D_1, \ldots, D_n), F) = (\Delta, \neg F \cup F, \neg D_1, \ldots, \neg D_n).$$

The translation $\{\phi \in T_1 \cup \cdots \cup T_{n+2} | (T_1, \ldots, T_{n+2}) = tr_{cp}(\Delta, (D_1, \ldots, D_n), F)\}$ of formulae and the translation $tr_{pp}(tr_{cp}(\Delta, (D_1, \ldots, D_n), F))$ of priorities from prioritized circumscription to prioritized autoepistemic logic via preferred subtheories are otherwise exactly the same as the translation of formulae $tr_{circ}(\Delta, (D_1, \ldots, D_n), F)$ and of priorities $p_{circ}(\langle \Delta, (D_1, \ldots, D_n), F \rangle)$ directly from prioritized circumscription to prioritized autoepistemic logic, except that members ϕ of Δ are taken as such instead of as implications $\neg L \neg \phi \rightarrow \phi$. This however does not make a difference as Δ, because it is consistent and the level with the highest priority, is always included in the preferred subtheory.

We show that for a preferred subtheory S of $L = tr_{cp}(\Delta, (D_1, \ldots, D_n), F)$, all models M such that $M \models S$, are $<$-minimal. Suppose that one such model M is not $<$-minimal, i.e. there is another model M' such that $M' \models \Delta$ and for some $k(1 \leq k \leq n)$ $\{p \in D_k | M' \models p\} \subset \{p \in D_k | M \models p\}$ and $\{p \in F | M' \models p\} = \{p \in F | M \models p\}$ and for all j $(1 \leq j < k)$ $\{p \in D_j | M' \models p\} = \{p \in D_j | M \models p\}$. Define

$$\begin{aligned} S' &= \Delta \cup \{f \in F | M' \models f\} \cup \{\neg f | f \in F, M' \not\models f\} \\ &\quad \cup \{\neg d | M' \not\models d, d \in D_i \text{ for some } i \text{ such that } 1 \leq i \leq k\}. \end{aligned}$$

Clearly S' is consistent, and $S \cap (\Delta \cup F \cup \neg F \cup \neg D_1 \cup \cdots \cup \neg D_k) \subset S' \cap (\Delta \cup F \cup \neg F \cup \neg D_1 \cup \cdots \cup \neg D_k)$. Hence S would not be a maximal subset, and consequently not a preferred subtheory of $L = tr_{cp}(\Delta, (D_1, \ldots, D_n), F)$. This contradiction shows that the assumption that M is not $<$-minimal is false.

We show that for a $<$-minimal model M, there is a preferred subtheory S of L such that $M \models S$. Define

$$\begin{aligned} S &= \Delta \cup \{f | f \in F, M \models f\} \cup \{\neg f | f \in F, M \not\models f\} \\ &\quad \cup \{\neg d | M \not\models d, d \in D_i \text{ for some } i \text{ such that } 1 \leq i \leq n\}. \end{aligned}$$

Clearly $M \models S$ and S is consistent. It remains to prove by induction on $i, 0 \leq i < n$ that each $S \cap (\Delta \cup F \cup \neg F \cup \neg D_1 \cup \cdots \cup \neg D_{i-1})$ is a maximal consistent subset of $\Delta \cup F \cup \neg F \cup \neg D_1 \cup \cdots \cup \neg D_{i-1}$.

Base case $i = 0$: S contains Δ like it should because Δ is consistent and the level with highest priority. It also contains exactly one of $f, \neg f$ for each $f \in F$; if it contained more, it would be inconsistent.

Inductive case $i \geq 1$: Let $H = S \cap (\Delta \cup F \cup \neg F \cup \neg D_1 \cup \cdots \cup \neg D_{i-1})$. By induction hypothesis H is a preferred subtheory of $\Delta \cup F \cup \neg F \cup \neg D_1 \cup \cdots \cup \neg D_{i-1}$. Suppose there is a set $T \subseteq \neg D_i$ such that $H \cup T$ is consistent and $H \cup T \supset H \cup (S \cap \neg D_i)$. Now $H \cup T$ has a model M' (that agrees with M on $F \cup D_1 \cup \cdots \cup D_{i-1}$). Because $T \supset (S \cap \neg D_i)$, $\{p \in D_i | M' \models p\} \subset \{p \in$

$D_i | M \models p\}$ and hence $M' < M$. This contradicts the $<$-minimality of M. Therefore S is a preferred subtheory of L.

To complete the proof, consider the following. If there is a $<$-minimal model M such that $M \not\models \phi$, there is a preferred subtheory S such that $M \models S$ as shown above. Clearly $S \not\models \phi$. Conversely, let there be a preferred subtheory S of L such that $S \not\models \phi$. Hence the set $S \cup \{\neg\phi\}$ is consistent, i.e. there is a model M such that $M \models S \cup \{\neg\phi\}$. As shown above, because $M \models S$, M is $<$-minimal. Clearly $M \not\models \phi$.

Lemmata 4.14 and 4.21 show the connection of level default theories to prioritized autoepistemic logic, i.e. the intersection of preferred subtheories S of a level default theory T is exactly the intersection of sets $E \cap \mathcal{L}$ for respective preferred stable expansions E of a translation of T to autoepistemic logic. $\Box$

Because our priorities for autoepistemic logic are not restricted to stratified prioritizations, prioritized circumscription as defined above can be generalized to arbitrary partial orders in a straightforward way. Grosof [1991] investigates this possibility among other generalizations of prioritized circumscription.

5 An alternative notion of most desirable stable expansions

The notion of most preferred stable expansions adopted in this work is only one from a repertoire of alternatives. The criteria for choosing it were its potential for efficient theorem-proving (see Section 6) and the natural interpretation of the partiality in the prioritization as ignorance; the prioritization used by the agent was thought of as total.

Another notion of most preferred stable expansions is based on the idea of inducing a partial ordering on stable expansions using the prioritization, and taking the maximal elements of the partial ordering as the most preferred ones. This notion could be justified by saying that the partiality of the prioritization partial order describes the lack of preferences from the agent's point of view. Hence the most preferred stable expansions do not have to correspond to any total extension of the partial ordering. The fact that Ryan uses an ordering principle of this flavour in his structured theories [Ryan, 1991; Ryan, 1992] suggests that it should be further investigated in the context of autoepistemic logic. A similar principle is used by Geffner [1990; 1992].

5.1 Definition

Given a prioritization, we define preference relations (partial orderings) on stable expansions, and define the most preferred stable expansions as the maximal elements of the preference relations. We write $E_1 \sqsupset_P E_2$ or if the prioritization is obvious then simply $E_1 \sqsupset E_2$, and read it "E_1 is preferred to E_2".

A prioritization P induces a relation $\sqsupset_P : 2^{\mathcal{L}_{ae}} \times 2^{\mathcal{L}_{ae}}$ on stable expansions as follows.

Definition 5.1 *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a set of formulae, and let E_1 and E_2 be stable expansions of Σ. Let $P = \langle \Phi, \mathcal{P} \rangle$ be a prioritization. Then $E_1 \sqsupset E_2$ if and only if there is $x \in (\Phi \cap E_2) \backslash E_1$ and for all $y \in \Phi, y \in E_1 \backslash E_2$ implies that there is z such that $z \mathcal{P} y$ and $z \in E_2 \backslash E_1$.*

Theorem 5.2 *The relation $\sqsupset$ is transitive and asymmetric.*

Proof: Transitivity: Suppose that $E_1 \sqsupset E_2$ and $E_2 \sqsupset E_3$. We have to show that $E_1 \sqsupset E_3$. For each $c \in (\Phi \cap E_1) \backslash E_3$ we find a $d \in E_3 \backslash E_1$ such that $d \mathcal{P} c$. We consider the following two cases.

Suppose $c \in (\Phi \cap E_1)\backslash E_3$ and $c \in E_2$. By the definition of $E_2 \sqsupset E_3$ there is an element $c' \in (\Phi \cap E_3)\backslash E_2$ such that $c'\mathcal{P}c$. Let c' be a minimal such element (because $\mathcal{P}$ is a well-ordering, there are minimal elements). If $c' \notin E_1$ then we have found an element $c' = d \in E_3\backslash E_1$. If $c' \in E_1$ then by the definition of $E_1 \sqsupset E_2$ there is an element $c'' \in E_2\backslash E_1$ such that $c''\mathcal{P}c'$. Now $c'' \in E_3$, because if $c'' \in E_2 \in E_3$ and $E_2 \sqsupset E_3$, there has to be an element $c''' \in (\Phi \cap E_3)\backslash E_2$ such that $c'''\mathcal{P}c''$, but this is impossible because $c'''\mathcal{P}c'$ by the transitivity of $\mathcal{P}$ and c' was chosen to be a minimal element such that $c'\mathcal{P}c$ and $c' \in E_3\backslash E_2$. Hence $c'' \in E_3\backslash E_1$.

Suppose $c \in (\Phi \cap E_1)\backslash E_3$ and $c \notin E_2$. If for all e such that $e\mathcal{P}c$, $e \in E_3$ iff $e \in E_2$ then by the definition of $E_1 \sqsupset E_2$ there is an element $d \in E_2\backslash E_1$ such that $d\mathcal{P}c$ and consequently $d \in E_3\backslash E_1$. Otherwise, take c' to be a minimal element in $E_3\backslash E_2$ such that $c'\mathcal{P}c$ and proceed as in case 1.

Asymmetry: Suppose $E_1 \sqsupset E_2$ and $E_2 \sqsupset E_1$ for some stable expansions E_1 and E_2 of Σ. Then by definition of $\sqsupset$ there are elements $\phi_0 \in (\Phi \cap E_1)\backslash E_2$ and $\psi_0 \in (\Phi \cap E_2)\backslash E_1$, and for all $\phi \in \Phi, \phi \in E_1\backslash E_2$ implies that there is ψ such that $\psi\mathcal{P}\phi$ and $\phi \in E_2\backslash E_1$, and similarly for all $\psi \in \Phi, \psi \in E_2\backslash E_1$ implies that there is $\phi, \phi\mathcal{P}\psi$ and $\phi \in E_1\backslash E_2$. Starting from ψ_0 we get an infinite sequence of elements $\psi_0, \phi_1, \psi_1, \phi_2, \ldots$ of Φ such that each $\psi_i \in (\Phi \cap E_2)\backslash E_1$, each $\phi_i \in (\Phi \cap E_1)\backslash E_2$, and $\phi_{i+1}\mathcal{P}\psi_i, \phi_{i+1}\mathcal{P}\psi_{i+1}$ for all $i \geq 0$. This kind of infinite descending sequence of elements of Φ is impossible by the well-ordering property of prioritizations. □

Definition 5.3 (P-maximality) *Let E be a stable expansion of a set of formulae Σ and P a prioritization. The stable expansion E is P-*maximal *if for no stable expansion E' of Σ, $E' \sqsupset_P E$.*

Lemma 5.4 (Right monotonicity) *Let $P = \langle \Phi, \mathcal{P} \rangle$ and $P' = \langle \Phi, \mathcal{P}' \rangle$ be prioritizations such that $\mathcal{P} \subseteq \mathcal{P}'$. Then $\sqsupset_P \subseteq \sqsupset_{P'}$.*

Proof: The result is immediate by the definition of $\sqsupset$: extending $\mathcal{P}$ does not cancel the existence of a y for which $y\mathcal{P}x$. □

Theorem 5.5 (Meaning of partiality) *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a set of formulae, $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization, and E a stable expansion of Σ. The stable expansion E is P-maximal if and only if for every stable expansion E' of Σ such that $(E' \cap \Phi) \neq (E \cap \Phi)$, there is a prioritization $T = \langle \Phi, \mathcal{T} \rangle$, where $\mathcal{T} \supseteq \mathcal{P}$ is a strict total order on Φ, such that $E \sqsupset_T E'$.*

Proof: ($\Leftarrow$) Suppose E is not P-maximal, i.e. $E' \sqsupset_P E$ for some E'. Let $\mathcal{T} \supseteq \mathcal{P}$ is a strict total order on Φ. By Lemma 5.4, $E' \sqsupset_T E$ for the prioritization $T = \langle \Phi, \mathcal{T} \rangle$. By the asymmetry of $\sqsupset_T$, $E \not\sqsupset_T E'$.

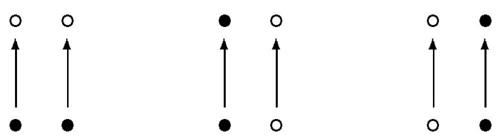
($\Rightarrow$) Let E be any given stable expansion of Σ. Suppose that for some stable expansion E' of Σ there is no prioritization $T = \langle \Phi, \mathcal{T} \rangle$ so that $\mathcal{T} \supseteq \mathcal{P}$ is a strict total order and $E \sqsupset_T E'$. That is, for all strict total orders $\mathcal{T} \supseteq \mathcal{P}$ on Φ, $E' \sqsupset_T E$, where $T = \langle \Phi, \mathcal{T} \rangle$. We show that $E' \sqsupset_P E$ and therefore E is not P-maximal. Because for all prioritizations T above $E' \sqsupset_T E$, there is a formula $\phi \in (\Phi \cap E) \backslash E'$. Hence the first requirement for $E' \sqsupset_P E$ is fulfilled. Let ϕ be in $(\Phi \cap E') \backslash E$. Assume there is no $\psi \in (\Phi \cap E) \backslash E'$ so that $\psi \mathcal{P} \phi$. Now we could construct a strict total order $\mathcal{T} \supseteq \mathcal{P}$ on Φ so that there is no $\psi \in (\Phi \cap E) \backslash E'$ such that $\psi \mathcal{T} \phi$ (first construct the strict partial order $\mathcal{R} = \mathcal{P} \cup \{(\chi, \chi') | \chi' = \phi$ or $\chi' \mathcal{P} \phi, \chi \neq \phi$ and not $\chi \mathcal{P} \phi\}$. Then extend $\mathcal{R}$ to any strict total order $\mathcal{T} \supseteq \mathcal{R}$ on Φ.) The existence of such $\mathcal{T}$ would contradict our assumption that for all $\mathcal{T}$, $E' \sqsupset_T E$. Hence there must be $\psi \in (\Phi \cap E) \backslash E'$ so that $\psi \mathcal{P} \phi$. Therefore $E' \sqsupset_P E$ and E is not P-maximal. $\square$

This theorem says that a stable expansion is maximal if it can be defended against all other stable expansions by extending the ordering to a total one, and using the obvious lexicographic comparison. Note that different defences against different stable expansions are allowed, and in fact there is in general no single defence that works against all other stable expansions at the same time. Hence the difference between this maximality notion and the one investigated earlier in this work is of the familiar kind: $\forall\exists$ versus $\exists\forall$.

Example 5.1 Let the prioritization $P = \langle \Phi, \mathcal{P} \rangle$ be the following one.

$$\begin{tikzcd} a & d \\ b \arrow[u] & c \arrow[u] \end{tikzcd}$$

Let $\Sigma = \{\neg L \neg \phi \rightarrow \phi | \phi \in \Phi\} \cup \{(\neg a \wedge b \wedge c \wedge \neg d) \vee (\neg a \wedge \neg b \wedge c \wedge d) \vee (a \wedge b \wedge \neg c \wedge \neg d)\}$. The stable expansion represented by the diagram below on the left is P-maximal. However, there are no total orderings of the partial order that would make it P-preferred (difference between the preference notions of Brewka [1989] and Geffner [1990] was demonstrated by Junker and Brewka [1991] using an example similar to this.



■

Theorem 5.6 (Inclusion) *Let Σ be a set of formulae and $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization. Every P-preferred stable expansion of Σ is P-maximal.*

Proof: Let E be a P-preferred stable expansion of Σ. Then there is a total order $\mathcal{T}$ on $\mathcal{P}$ such that for all stable expansions E' of Σ the following claim holds.

For all $\phi \in \Phi(\phi \in E \backslash E'$ implies there is $\psi \in (\Phi \cap E') \backslash E$ such that $\psi \mathcal{T} \phi)$.

Let E' be a stable expansion of Σ such that $E \cap \Phi \neq E' \cap \Phi$. Now either $(\Phi \cap E') \backslash E$ or $(\Phi \cap E) \backslash E'$ is non-empty. If $(\Phi \cap E') \backslash E \neq \emptyset$ then together with the above claim immediately $E \sqsupset_{\langle \Phi, \mathcal{T} \rangle} E'$. If $(\Phi \cap E) \backslash E' \neq \emptyset$ then by the above also $(\phi \cap E') \backslash E \neq \emptyset$ and consequently $E \sqsupset_{\langle \Phi, \mathcal{T} \rangle} E'$. By Theorem 5.5 E is P-maximal. □

Theorem 5.7 (Layered prioritizations) *Let Σ be a set of formulae and $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization such that for a partition $\Phi_1, \Phi_2, \ldots, \Phi_n$ of Φ, the following equality holds.*

$$\mathcal{P} = \{(\phi, \phi') | \phi \in \Phi_i, \phi' \in \Phi_j, i < j\}$$

Then a set of formulae $E \subseteq \mathcal{L}_{ae}$ is a P-maximal stable expansion of Σ if and only if E is a P-preferred stable expansion of Σ.

Proof: By Theorem 5.6 each P-preferred stable expansion of Σ is P-maximal. We have to show that each P-maximal stable expansion of Σ is P-preferred. For each P-maximal stable expansion E we construct a strict total order $\mathcal{T}$ on Φ such that

for all $\phi \in \Phi(\phi \in E \backslash E'$ implies there is $\psi \in (\Phi \cap E') \backslash E$ such that $\psi \mathcal{T} \phi)$.

Let E be a P-maximal stable expansion of Σ. Let $\mathcal{T}$ be a strict total order on Φ such that $\mathcal{T} \supseteq \mathcal{P}$ and for $\phi \in \Phi \backslash E$ and $\psi \in \Phi \cap E$, if neither $\phi \mathcal{P} \psi$ nor $\psi \mathcal{P} \phi$, then $\phi \mathcal{T} \psi$. That such an ordering exists can be easily seen: just take formulae of Φ layer by layer starting from the lowest one, and taking the ones not in E before the ones in E at the same layer. The maximality of E means that for every stable expansion E' of Σ, there is no $x \in (\Phi \cap E) \backslash E'$, or there is $y \in (\Phi \cap E') \backslash E$ and there is no $z \in E \backslash E'$ such that $z \mathcal{P} y$. Take any stable expansion E'. Assume there are no elements in $(\Phi \cap E) \backslash E'$. Now the implication in the definition of preferredness is true for every element of Φ. Assume there is $y \in (\Phi \cap E') \backslash E$ and for no $z \in E \backslash E'$, $z \mathcal{P} y$. Because y is not in E, all elements e of Φ that are incomparable to y according to $\mathcal{P}$ and belong to E fulfill $y\mathcal{T}e$ by construction. Hence we can strengthen the claim to: $y \in (\Phi \cap E')\backslash E$ and for no $z \in E\backslash E'$, $z\mathcal{T}y$. Now the formula ψ in the definition of preferredness can be taken to be y for all $\phi \in E\backslash E'$. Therefore E is a P-preferred stable expansion of Σ. □

5.2 Connection to structured theories

Ryan's [1991] structured theories view theories as partial orderings of formulae. The minimal elements in the orderings are the most plausible formulae. Each structured theory determines a class of models. In determining these models the formulae in the structured theory are taken into account in the order of plausibility. Formulae themselves are considered as well as their natural consequences, which are logical consequences of the formula that (as truth-functions) fulfill certain monotonicity conditions. Roughly, a logical consequence ψ of a formula ϕ is a natural consequence of ϕ, written $\phi \models^{\natural} \psi$, if ψ is monotonic in the propositional variables that occur in ϕ, i.e. if $M \models \psi$ and M' is a model that satisfies more (in the set-inclusion sense) propositional variables of ϕ than M, then also $M' \models \psi$. The set of natural consequences of a formula is infinite, but we shall demonstrate that in the propositional case it can be replaced by a finite set.

The next definition tells what structured theories are and how the consequence relation $\Gamma \models \phi$ for structured theories Γ and formulae ϕ is defined.

Definition 5.8 ([Ryan, 1991]) *Let X be a set, $\leq$ a partial order on X and $F : X \to \mathcal{L}$ a function that maps elements of X to formulae. Then $\Gamma = \langle X, \leq, F\rangle$ is a structured theory.*

Let M and N belong to $\mathcal{M}$ (the collection of all models). Then $M \sqsubseteq_\phi N$, if for each ψ,

$$\phi \models^{\natural} \psi \text{ implies } (M \models \psi \text{ implies } N \models \psi).$$

Let $\Gamma = \langle X, \leq, F\rangle$ be a structured theory, and let $M, N \in \mathcal{M}$. Then $M \sqsubseteq^\Gamma N$ if for each $x \in X$, if $M \not\sqsubseteq_{F(x)} N$ then there is a point $y \leq x$ such that $M \sqsubset_{F(y)} N$ ($M \sqsubset_{F(y)} N$ means that $M \sqsubseteq_{F(y)} N$ and not $N \sqsubseteq_{F(y)} M$.)

Let Γ be a structured theory, and M an element of $\mathcal{M}$. Then $M \models \Gamma$ if M is $\sqsubseteq^\Gamma$-maximal.

Let Γ be a structured theory. Then $\Gamma \models \phi$, if for each $M \in \mathcal{M}$, $M \models \Gamma$ implies $M \models \phi$.

The following translation of structured theories into prioritized autoepistemic logic resembles the translation of preferred subtheories in Section 4.3.

Definition 5.9 *Let $\Gamma = \langle X, \leq, F\rangle$ be a structured theory. The translation of Γ to prioritized autoepistemic logic is defined using the following functions.*

$$\begin{aligned} tr_{st}(\Gamma) &= \{\neg L\neg\phi \rightarrow \phi | F(x) \models^{\natural} \phi, x \in X\} \\ p_{st}(\Gamma) &= \langle\{\neg\phi | F(x) \models^{\natural} \phi, x \in X\}, \\ &\quad \{(\neg\phi, \neg\psi) | F(x) \models^{\natural} \phi, F(y) \models^{\natural} \psi, x \leq y \textit{ for some } x \in X, y \in X\}\rangle \end{aligned}$$

Because Ryan allows multiple occurrences of the same formula in a structured theory and distinct formulae may share natural consequences, there is a minor problem with our definition of prioritizations, where at most one occurrence of a formula is allowed. This can be solved by renaming each propositional variable p occuring in a natural consequence of a formula $F(x)$ to p_x and extending the translation of a structured theory with equivalences $p \leftrightarrow p_x$ (or equivalently implications $p \rightarrow p_x$ and $p_x \rightarrow p$).

This transformation causes a linear increase in the size of the translation. In practice a significantly smaller amount of renaming suffices, because natural consequences of different formulae are often distinct, and natural consequences that are duplicated higher in the partial order can be left out. Our discussion of structured theories from now on assumes that this transformation has been made.

Our main result about structured theories is the following theorem, which we prove later after giving some lemmata.

Theorem 5.10 *Let $\Gamma = \langle X, \leq, F\rangle$. Then $\Gamma \models \phi$ if and only if ϕ belongs to all $p_{st}(\Gamma)$-maximal stable expansions of $tr_{st}(\Gamma)$.*

Lemma 5.11 *Let $\Gamma = \langle X, \leq, F\rangle$ be a structured theory. Let E, E' be stable expansions of $tr_{st}(\Gamma)$ and let M, M' be models of $E\cap\mathcal{L}$ and $E'\cap\mathcal{L}$, respectively. Then $E \sqsupset_{p_{st}(\Gamma)} E'$ if and only if $M' \sqsubset^{\Gamma} M$.*

Proof: Remark **A**: Let E be a stable expansion of $tr_{st}(\Gamma)$. Let M be a model of $E\cap\mathcal{L}$. Let c be a natural consequence of $F(x), x \in X$. Then $\neg c \in E$ if and only if $M \models \neg c$, and $c \in E$ if and only if $M \models c$ (immediate by Lemma 4.16).

Define $\langle\Phi, \mathcal{P}\rangle = p_{st}(\Gamma)$.

($\Rightarrow$) Suppose $E \sqsupset_{p_{st}(\Gamma)} E'$. Suppose $M' \not\sqsubseteq_{F(x)} M$ for some $x \in X$. That is, for some c, $F(x) \models^{\natural} c$, $M' \models c$ but $M \not\models c$. Now $M' \not\models \neg c$ and $M \models \neg c$, and by **A** above $\neg c \in E\backslash E'$ because $\neg c \in \Phi$. Because $E \sqsupset_{p_{st}(\Gamma)} E'$ there is some

$\neg d \in \Phi$ such that $\neg d\mathcal{P}\neg c$ and $\neg d \in E'\backslash E$. Take $\neg d$ be a minimal element (according to $\mathcal{P}$) such that $\neg d\mathcal{P}\neg c$ and $\neg d \in E'\backslash E$. Now by **A** $M \models d$ and $M' \not\models d$, and there is $y \in X$ such that $F(y) \models^{\natural} d$ and $y \leq x$. To show that $M' \sqsubset_{F(y)} M$ assume the opposite, i.e. for some $e, F(y) \models^{\natural} e$ and $M' \models e$ and $M \not\models e$. This translates to $\neg e \in E\backslash E'$ by **A**. Now because $E \sqsupset_{p_{st}(\Gamma)} E'$ there should be $\neg f \in \Phi$ such that $\neg f\mathcal{P}\neg e, \neg f \in E'\backslash E$. That this contradicts the assumption that $\neg d$ is minimal element such that $\neg d\mathcal{P}\neg c, \neg d \in E'\backslash E$ is because of the following. There is some $z \in X$ such that $F(z) \models^{\natural} f$. Because $F(y) \models^{\natural} e$ and $\neg f\mathcal{P}\neg e$, $z \leq y$. Hence $\neg f\mathcal{P}\neg d\mathcal{P}\neg c$. This completes the proof of $M' \sqsubseteq^{\Gamma} M$.

To prove that $M \not\sqsubseteq^{\Gamma} M'$, we show that there is $x \in X$ such that $M \not\sqsubseteq_{F(x)} M'$ and there is no $y \in X$ such that $y \leq x$ and $M \sqsubset_{F(y)} M'$. Because $E \sqsupset_{p_{st}(\Gamma)} E'$, there is some $\neg\phi \in (\Phi \cap E')\backslash E$. Take $\neg\phi$ to be such that $\neg\phi \in (\Phi \cap E')\backslash E$ and there are no formulae χ such that $\chi\mathcal{P}\neg\phi, \chi \in (\Phi \cap E')\backslash E$ (such a formula exists because $\mathcal{P}$ is a well-ordering). Now $M \not\models \neg\phi$ and $M' \models \neg\phi$, and hence $M \models \phi, M' \not\models \phi$. By definition of Φ there is $x \in X$ such that $F(x) \models^{\natural} \phi$, hence $M \not\sqsubseteq_{F(x)} M'$. Assume there is y such that $y \leq x$ and $M \sqsubset_{F(y)} M'$. Now there should be c such that $F(y) \models^{\natural} c$ and $M \not\models c$ and $M' \models c$. Now $\neg c$ should be in $(\Phi \cap E)\backslash E'$ by Remark **A** and $\neg c\mathcal{P}\neg\phi$. However, this contradicts our assumption that $\neg\phi$ is a $\mathcal{P}$-minimal formula in $(\Phi \cap E')\backslash E$.

($\Leftarrow$) Suppose $M' \sqsubset^{\Gamma} M$, i.e. $M' \sqsubseteq^{\Gamma} M$ and $M \not\sqsubseteq^{\Gamma} M'$. We show that $E \sqsupset_{p_{st}(\Gamma)} E'$, i.e. there is $e \in (\Phi \cap E')\backslash E$ and for all $\psi \in (\Phi \cap E)\backslash E'$ there is $\chi \in (\Phi \cap E')\backslash E$ such that $\chi\mathcal{P}\psi$. First we prove the existence of an $e \in (\Phi \cap E')\backslash E$. $M \not\sqsubseteq^{\Gamma} M'$ means that for some $x \in X, M \not\sqsubseteq_{F(x)} M'$ and there is no y such that $y \leq x$ and $M \sqsubset_{F(y)} M'$. $M \not\sqsubseteq_{F(x)} M'$ means that there is c such that $F(x) \models^{\natural} c$, $M \models c$ and $M' \not\models c$. Because $M \models E \cap \mathcal{L}$ and $M' \models E' \cap \mathcal{L}$ we from Remark **A** $\neg c \notin E$ and $\neg c \in E'$. Hence there is $\neg c = e \in (\Phi \cap E')\backslash E$. Then we prove that for all $\psi \in (\Phi \cap E)\backslash E'$ there is $\chi \in (\Phi \cap E')\backslash E$ such that $\chi\mathcal{P}\psi$. Assume $\psi \in (\Phi \cap E)\backslash E'$. Then $\psi = \neg\phi$ for some ϕ. Now $M \not\models \phi$, $M' \models \phi$ and for some $x \in X$, $F(x) \models^{\natural} \phi$. Because $M' \sqsubseteq^{\Gamma} M$, $M' \not\sqsubseteq_{F(x)} M$ implies the existence of $y \in X$ such that $y \leq x$ and $M' \sqsubset_{F(y)} M$. This implies that for some w, $F(y) \models^{\natural} w$, $M' \not\models w$ and $M \models w$. By Remark **A**, $\neg w \in E'$ and $\neg w \notin E$. Hence there is $\neg w = \chi \in (\Phi \cap E')\backslash E$ and $\neg w\mathcal{P}\neg\phi$. □

Lemma 5.12 *Let $\Gamma = \langle X, \leq, F\rangle$ be a structured theory, and $\phi \in \mathcal{L}$ a formula. If $\phi \notin E$ for some $p_{st}(\Gamma)$-maximal stable expansion E, then there is a $\sqsubseteq^{\Gamma}$-maximal model M such that $M \not\models \phi$.*

Proof: Let E be a $p_{st}(\Gamma)$-maximal stable expansion of $\Sigma = tr_{st}(\Gamma)$ such that $\phi \notin E$. Hence there is a model M such that $M \models (E \cap \mathcal{L}) \cup \{\neg\phi\}$. Assume there is N such that $M \sqsubset^{\Gamma} N$. The set $\Omega = \{\phi | N \models \phi, F(x) \models^{\natural} \phi$ for some $x \in X\}$ is not necessarily a maximal consistent subset of $T = \{\phi | F(x) \models^{\natural} \phi$ for some $x \in X\}$. Let Ω' be a maximal consistent set such that $\Omega \subseteq \Omega' \subseteq T$. Because Ω' is consistent, it has a model L. Because for all ϕ such that $F(x) \models^{\natural} \phi$ for some $x \in X$, $N \models \phi$ implies $L \models \phi$, for no $x \in X$ $N \not\sqsubseteq_{F(x)} L$. Therefore $N \sqsubseteq^{\Gamma} L$ and by the transitivity of $\sqsubseteq^{\Gamma}$, $M \sqsubset^{\Gamma} L$. Because Ω' is a maximal consistent subset of T, there is by Lemma 4.17 a corresponding stable expansion E' of $tr_{st}(\Gamma)$. By Lemma 5.11 $E' \sqsupset_{p_{st}(\Gamma)} E$. This contradicts the maximality of E, and hence M must be $\sqsubseteq^{\Gamma}$-maximal. □

Lemma 5.13 *Let $\Gamma = \langle X, \leq, F \rangle$ be a structured theory. Let $\phi \in \mathcal{L}$ be a formula. If $M \not\models \phi$ for some $\sqsubseteq^{\Gamma}$-maximal model M, then there is a $p_{st}(\Gamma)$-maximal stable expansion E of $tr_{st}(\Gamma)$ such that $\phi \notin E$.*

Proof: Let $\Sigma = tr_{st}(\Gamma)$ and $T = \{\phi | F(x) \models^{\natural} \phi, x \in X\}$. Let M be a $\sqsubseteq^{\Gamma}$-maximal model. Hence $\Pi = \{\psi \in T | M \models \psi\}$ is a consistent subset of T. Suppose Π is not maximal, i.e. there is a consistent set $\Omega \subseteq T$ such that $\Omega \supset \Pi$ Now $M' \models \Omega$ for some model M'. Clearly, for no $\chi \in T$, $M \models \chi$ and $M' \not\models \chi$. Hence $M \sqsubseteq^{\Gamma} M'$. Take any $\chi \in \Omega \backslash \Pi$. Now $M' \models \chi$ and $M \not\models \chi$. Hence $M' \not\sqsubseteq^{\Gamma} M$, and finally $M \sqsubset^{\Gamma} M'$. This contradicts the maximality of M, so Π must be a maximal consistent subset of T. Hence by Lemma 4.17 $\Lambda = \{\neg L \neg \phi | \phi \in \Pi\} \cup \{L \neg \phi | \phi \in T \backslash \Pi\}$ is a $tr_{st}(\Gamma)$-full set. Because $M \models \Pi$ and $M \not\models \phi$, $\Pi \not\models \phi$. $\Pi \not\models \phi$ implies $Red(\Sigma, \Lambda) \not\models \phi$ implies $\phi \notin SE_{\Sigma}(\Lambda) = E$. Suppose E is not $p_{st}(Gamma)$-maximal. By Lemma 5.11, there would be a model N for which $M \sqsubset^{\Gamma} N$. This would contradict the maximality of M. Hence E is $p_{st}(Gamma)$-maximal and $\phi \notin E$. □

Proof of Theorem 5.10:

$$\begin{array}{lll}
\Gamma \models \phi & \text{iff} & \text{for all } \sqsubseteq^{\Gamma}\text{-maximal models } M, M \models \phi \\
& \text{iff} & \text{there is no } \sqsubseteq^{\Gamma}\text{-maximal model } M \text{ such that } M \not\models \phi \\
& \text{iff} & \text{there is no } p_{st}(\Gamma)\text{-maximal stable expansion } E \text{ such that } \phi \notin E \\
& \text{iff} & \text{for all } p_{st}(\Gamma)\text{-maximal stable expansions } E, \phi \in E
\end{array}$$

The third equivalence is by Lemmata 5.12 and 5.13: $\phi \notin E$ for some $p_{st}(\Gamma)$-maximal stable expansion E of $tr_{st}(\Gamma)$ if and only if $M \not\models \phi$ for some $\sqsubseteq^{\Gamma}$-maximal model M. □

Proposition 5.14 *For structured theories based on classical propositional logic, the relation $\models^{\natural}$ in Ryan's definition can be replaced by a relation $\models^{\circ}$*

that is a subset of $\models^{\natural}$ *and fulfills the property that the set* $\{\psi|\phi \models^{\circ} \psi\}$ *is finite for all* $\phi \in \mathcal{L}$*. This change preserves all properties of* $\models$ *in* $\Gamma \models \phi$*, where* Γ *is a structured theory and* $\phi \in \mathcal{L}$ *a formula.*

Proof: For a formula ϕ, none of its natural consequences contains propositional variables not contained in ϕ, and the number of truth-functions of propositional variables in ϕ is finite. Because logically equivalent natural consequences are treated in Ryan's definitions equally, it suffices that we have only one natural consequence from each class of logically equivalent natural consequences. $\square$

As a special case consider disjunctions of literals. A disjunction of literals has only one relevant natural consequence, it itself. With this kind of disjunctions as formulae structured theories come very close to preferred subtheories because the only difference is the interpretation of partiality in the preference partial orders. From our point of view the differences between Ryan's structured theories and Brewka's subtheories are the use of the natural consequences of formulae in the former and the different notions of preferredness that correspond to P-maximality and P-preferredness, respectively.

6 An automatic theorem-prover

We describe a decision procedure for prioritized autoepistemic logic that is more efficient than the ones given in Section 2.6. A theorem-proving system that we have implemented as a part of this work is briefly discussed in the end of the section.

6.1 A decision procedure for prioritized autoepistemic logic

Section 2.6 contains a complexity analysis of prioritized autoepistemic logic. The complexity analysis is based on a decision procedure for the logic that reduces its decision problem to the decision problem of a classical logic. The decision procedure is not very efficient and can hardly be used for any practical application. In this section we give a decision procedure that works efficiently for an important subclass of prioritized autoepistemic theories. We use the full set notion 2.4 in the construction of preferred stable expansions. Hence prioritizations that are partial orders on a subset of $\{\phi | L\phi \in Sf^L(\Sigma)\}$ can be effectively used in the computation of preferred stable expansions. Such prioritizations naturally occur in many applications: for example prioritizations in examples in this report are such, and the translations from other formalisms to prioritized autoepistemic logic presented in this report, prioritize $\{\phi | L\phi \in Sf^L(\Sigma)\}$ or a subset of it. Our algorithms work for other kinds of priorities, too. By adding to Σ tautologies $L\phi \rightarrow L\phi$ for formulae $\phi \in \Phi$ for which $L\phi \notin Sf^L(\Sigma)$, the same mechanism for computing preferred stable expansions can be used for all prioritizations. However, if the number of such formulae ϕ is large, the computation of preferred stable expansions is not very efficient: it essentially reduces to an exhaustive traversal of all candidate full sets of Σ.

Our decision procedures for prioritized autoepistemic logic are based on the corresponding procedures of [Niemelä, 1993] for ordinary autoepistemic logic. Niemelä gives procedures for three different versions of autoepistemic logic, the original version defined by Moore, and two versions due to Niemelä that restrict to enumeration-based expansions and L-hierarchic expansions. These three procedures differ only in some details. In this section we give a decision procedure for the prioritized autoepistemic logic described in Section 2, i.e. the one based on Moore's definition of stable expansions but it is easy to modify the procedures for enumeration-based and L-hierarchic versions. The decision procedure is given in Figure 9 and two auxiliary procedures in Figure 8.

Niemelä's procedures for ordinary autoepistemic logic are improvements of the trivial procedures that use exhaustive search to find all stable expansions of a set Σ. This search is based on identifying the full sets of Σ among the

```
procedure extendible(P, Λ, Q)
begin
  ⟨Φ, 𝒫⟩ := P;
  if Φ = ∅ then return true
  else
    begin
      A := {φ ∈ Φ|φ minimal in Φ, Lφ ∈ Λ implies Lφ ∈ ⋂Q};
      if A = ∅ then return false
      else
        begin
          let φ be any member of A;
          P′ := ⟨Φ\{φ}, 𝒫 ∩ (Φ\{φ} × Φ\{φ})⟩;
          if Lφ ∈ Λ
          then return extendible(P′, Λ, Q)
          else return extendible(P′, Λ, {Λ′ ∈ Q|¬Lφ ∈ Λ′})
        end
    end
end

procedure next(P, Σ, Λ, Q)
begin
  ⟨Φ, 𝒫⟩ := P;
  if Φ = ∅ then
    begin
      F := {φ|Lφ ∈ Sf^L(Σ), Lφ ∉ Λ, ¬Lφ ∉ Λ};
      return any element of F
    end
  else
    begin
      A := {φ ∈ Φ|φ minimal in Φ, Lφ ∈ Λ implies Lφ ∈ ⋂Q};
      if A = ∅ then return any χ such that Lχ ∈ Sf^L(Σ)\Sf^qL(Λ);
      B := {φ ∈ A|Lφ ∉ Λ, ¬Lφ ∉ Λ};
      if B ≠ ∅ then return any element of B;
      let φ be any member of A;
      P′ := ⟨Φ\{φ}, 𝒫 ∩ (Φ\{φ} × Φ\{φ})⟩;
      if Lφ ∈ Λ
      then return next(P′, Σ, Λ, Q)
      else return next(P′, Σ, Λ, {Λ′ ∈ Q|¬Lφ ∈ Λ′})
    end
end
```

Figure 8: Subprocedures for the decision procedure for prioritized autoepistemic logic

```
procedure is_m(Σ, B, F)
begin
  repeat
    B' := B;
    for each Lχ ∈ Sf^L(Σ)\Sf^{qL}(B) do
      if Σ ∪ B ∪ F ⊨ χ
      then B := B ∪ {Lχ}
  until B = B';
  return (B, F)
end

procedure der_p(P, Σ, B, F, φ, Q)
begin
  (B, F) := is_m(Σ, B, F);
  if extendible(P, B ∪ F, Q) = false then return (∅,false);
  if for some ¬Lχ ∈ B, Σ ∪ B ∪ F ⊨ χ then return (∅,false)
  else
    if Sf^L(Σ) ⊆ Sf^{qL}(B) ∪ F then
      if F ⊆ B then return ({B},test(Σ, B, φ))
      else return (∅,false)
    else
      begin
        χ := next(P, Σ, B ∪ F, Q);
        (S, b) := der_p(P, Σ, B ∪ {¬Lχ}, F, φ, Q);
        (S', b') := der_p(P, Σ, B, F ∪ {Lχ}, φ, S ∪ Q);
        return (S ∪ S', b or b')
      end
end
```

Figure 9: The decision procedure for prioritized autoepistemic logic

$2^{|Sf^L(\Sigma)|}$ candidates. Niemelä's procedures use the monotonicity property of $\models$ to add $L\phi \in Sf^L(\Sigma)$ to a full set as soon as ϕ becomes derivable, and to detect that formulae ϕ for which $\neg L\phi$ is in the full set, become derivable. These improvements prune the search tree in several cases. Our procedures for prioritized autoepistemic logic directly build on these algorithms (they can be based on the trivial algorithm just as well as on Niemelä's more sophisticated algorithm). The idea is to select elements ϕ of $\Phi \subseteq \{\phi | L\phi \in Sf^L(\Sigma)\}$ in an order that conforms to the partial order $\mathcal{P}$ in the prioritization P, and include $\neg L\phi$ in the full set Λ to be constructed: always choose a $\phi \in \Phi$ such that for all $\psi\mathcal{P}\phi$ either $L\psi \in \Lambda$ or $\neg L\psi \in \Lambda$. This policy guarantees that the first Σ-full set found – maybe after some backtracking – is indeed P-preferred, i.e. corresponds to some strict total order $\mathcal{T} \supseteq \mathcal{P}$ on Φ. The rest of the P-preferred stable expansions are found by backtracking. When taking a different branch in the search tree, a call to the subprocedure *extendible* ensures that the full set corresponding to that branch really has an associated strict total order $\mathcal{T}$ on Φ under which all the stable expansions found earlier are "worse" than the current one. Note that the possibility of more than one preferred stable expansions for which the intersections with Φ differ, is due to the partiality in the partial order $\mathcal{P}$ in prioritizations.

The way in which new elements $\neg L\chi$ are chosen guarantees that the strict total order $\mathcal{T}$ can be used for defeating all stable expansions found later. The call to the procedure *next* returns a formula $\chi \in \{\phi | L\phi \in Sf^L(\Sigma) \backslash Sf^{qL}(B \cup F)\}$ such that $\chi\mathcal{T}\phi$ for all $\phi \in \{\phi | L\phi \in Sf^L(\Sigma) \backslash Sf^{qL}(B \cup F)\}$. On backtracking to a different branch of the search tree the formula $\neg L\chi$ in the full set is replaced by $L\chi$. Hence all stable expansions found later are "worse" according to $\mathcal{T}$. The procedure essentially considers candidate full sets in a lexicographically decreasing order: first it constructs the set $\{\neg L\phi_1, \ldots, \neg L\phi_{n-1}, \neg L\phi_n\}$, then the set $\{\neg L\phi_1, \ldots, \neg L\phi_{n-1}, L\phi_n\}$, the set $\{\neg L\phi_1, \ldots, L\phi_{n-1}, \neg L\phi_n\}$, and so on. The algorithm prunes the search tree as soon as it is certain that no total ordering can make the stable expansion of the current branch preferred. For example, by Theorem 2.10 for prioritizations P that totally order $\{\phi | L\phi \in Sf^L(\Sigma)\}$ there is at most one P-preferred stable expansion. Our algorithm uses backtracking search to find this stable expansion, and all subsequent computation is avoided thanks to the pruning mechanism. The number of nodes in the search tree is at most $2^{|Sf^L(\Sigma)|}$ although there may be $|Sf^L(\Sigma)|!$ strict total orders on Φ.

Next we give some lemmata for proving the correctness of the algorithm. We use the notation $SE_\Sigma(Q)$ for the set of stable expansions $\{SE_\Sigma(\Lambda) | \Lambda \in Q\}$, where Q is a set of Σ-full sets.

Definition 6.1 (Preferred in) *Let $P = \langle \Phi, \mathcal{P} \rangle$ be a prioritization and $\Sigma \subseteq \mathcal{L}_{ae}$ a set of formulae. A stable expansion E of Σ is P-*preferred in* a set of* *stable expansions* $X = \{E_1, \ldots, E_n\}$ *of* Σ *if and only if there is a strict total order* $\mathcal{T} \supseteq \mathcal{P}$ *on* Φ *such that for all* $E' \in X$,

$$\textit{for all } \phi \in \Phi(\phi \in E \backslash E' \textit{ implies there is } \psi \in E' \backslash E \textit{ such that } \psi \mathcal{T} \phi).$$

Lemma 6.2 *Let* $\Sigma \subseteq \mathcal{L}_{ae}$ *be a set of formulae and* $P = \langle \Phi, \mathcal{P} \rangle$ *a prioritization where* Φ *is finite. Let* X *be a set of stable expansions of* Σ *such that each* $E \in X$ *is* P*-preferred in* X*, and all* P*-preferred stable expansions of* Σ *are in* X*. Then* X *is exactly the set of* P*-preferred stable expansions of* Σ.

If a stable expansion E *of* Σ *is* P*-preferred, then* E *is* P*-preferred in any set of stable expansions of* Σ.

Proof: Let Σ be a set of formulae and $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization. The definition of P-preferred stable expansions of Σ is obtained from the definition of stable expansions of Σ P-preferred in some set X, if X is the set of all stable expansions of Σ.

To prove the first part it suffices to show that all members of X are preferred stable expansions of Σ. Assume there is $E \in X$ such that E is P-preferred in X but not P-preferred. Let $\mathcal{T} \supseteq \mathcal{P}$ be a strict total order using which E is P-preferred in X. By Theorem 2.9 there is a P-preferred stable expansion E' of Σ such that the associated strict total order is $\mathcal{T}$. Now both

$$\text{for all } \phi \in \Phi(\phi \in E \backslash E' \text{ implies there is } \psi \in E' \backslash E \text{ such that } \psi \mathcal{T} \phi)$$

and

$$\text{for all } \phi \in \Phi(\phi \in E' \backslash E \text{ implies there is } \psi \in E \backslash E' \text{ such that } \psi \mathcal{T} \phi).$$

Because E' is P-preferred and E is not, $E \cap \Phi \neq E' \cap \Phi$ must hold. Starting from any $\phi \in (\Phi \cap E) \backslash E'$ the above conditions induce an infinite sequence of formulae $\phi_1, \phi_2, \phi_3, \ldots$, where $\phi_1 = \phi$ and $\phi_{i+1} \mathcal{T} \phi_i$, and $\phi_i \in E \backslash E'$ for odd i and $\phi_j \in E' \backslash E$ for even j. This kind of infinite descending sequence of formulae is impossible because Φ is finite.

The second part of the lemma is by noticing that if E is preferred in some set X, it is preferred in any subset of X. $\square$

Lemma 6.3 *Let* $\{E\} \cup Q$ *be a set of stable expansions, and* $P = \langle \Phi, \mathcal{P} \rangle$ *a prioritization. Let* E *be* P*-preferred in* Q*. Let* ϕ *be one of the* $\mathcal{P}$*-minimal elements of* Φ *such that* $\phi \in E$ *implies* $\phi \in \bigcap Q$*. Then there is a strict total order* $\mathcal{T} \supseteq \mathcal{P}$ *on* Φ *with* ϕ *as the least element and that fulfills the condition in Definition 6.1 for* E.

Proof: That E is P-preferred in Q implies the existence of a strict total order $\mathcal{T}' \supseteq \mathcal{P}$ on Φ such that the condition in Definition 6.1 is fulfilled. The least element of the ordering $\mathcal{T}'$ is not necessarily ϕ. Let $\phi_1, \ldots, \phi, \ldots$ be the ordering $\mathcal{T}'$. We claim that the ordering $\mathcal{T} : \phi, \phi_1 \ldots$ that is like $\mathcal{T}'$ except that ϕ has been taken to be the least element, also fulfills the condition for E. Note that $\phi\mathcal{T}'\psi$ implies $\phi\mathcal{T}\psi$. Clearly $\mathcal{T} \supseteq \mathcal{P}$ because ϕ is one of the P-minimal elements of Φ. Let E' be a stable expansion in Q, and let there be $\psi \in (\Phi \cap E)\backslash E'$. By P-preferredness, there is $\psi' \in E'\backslash E$ such that $\psi'\mathcal{T}'\psi$. We show that also $\psi'\mathcal{T}\psi$. Note that all pairs of elements on which $\mathcal{T}$ and $\mathcal{T}'$ disagree are such that ϕ is one of the components of the pair. It is not possible that $\psi = \phi$, because either $\phi \notin E$ or $\phi \in E$ and $\phi \in \bigcap Q$. Because $\phi\mathcal{T}'\psi$ implies $\phi\mathcal{T}\psi$, also $\psi'\mathcal{T}\psi$. This establishes the result. □

Lemma 6.4 (Procedure *extendible*) *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a finite set of formulae, $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization such that $\Phi \subseteq \{\phi | L\phi \in Sf^L(\Sigma)\}$, Λ a Σ-full set, and Q a finite set of Σ-full sets. Let b be the value returned by the procedure call extendible(P, Λ, Q). Then b is* true *if and only if $SE_\Sigma(\Lambda)$ is P-preferred in $SE_\Sigma(Q)$. If b is not* true, *it is* false.

Let Λ' be a subset of a Σ-full set Λ and let $SE_\Sigma(\Lambda)$ be P-preferred in $SE_\Sigma(Q)$. Then the procedure call extendible(P, Λ', Q) returns true.

Proof: The proof is by induction on $|\Phi|$. The base case corresponds to the *then*-part of the outermost *if*-statement of the procedure, and the inductive case to the *else*-part.

Base case $|\Phi| = 0$: For the empty prioritization $P_e = \langle \emptyset, \emptyset \rangle$ every stable expansion is P_e-preferred, hence the return value b is *true*.

Inductive case $|\Phi| \geq 1$: We prove the equivalence concerning b by showing the implications to both directions.

($\Leftarrow$) Assume that $SE_\Sigma(\Lambda)$ is P-preferred in $SE_\Sigma(Q)$. We show that the procedure returns *true*. Assume that for all minimal $\phi \in \Phi$ both $L\phi \in \Lambda$ and for some $\Lambda' \in Q$ the formula $\neg L\phi$ belongs to Λ'. Now whichever minimal $\phi \in \Phi$ is chosen to be the least element of a strict total order there will always be some $SE_\Sigma(\Lambda'), \Lambda' \in Q$ such that $\phi \in SE_\Sigma(\Lambda)\backslash SE_\Sigma(\Lambda')$. Clearly $SE_\Sigma(\Lambda)$ could not then be P-preferred in $SE_\Sigma(Q)$. Hence there indeed is a minimal $\phi \in \Phi$ such that $L\phi \in \Lambda$ implies $L\phi \in \bigcap Q$ (and hence in the procedure $A \neq \emptyset$), and *extendible* does not immediately return *false*. We analyze by cases the *if* statement.

Let $L\phi \in \Lambda$. The return value is the value returned by extendible(P', Λ, Q) where $P' = \langle \Phi\backslash\{\phi\}, \mathcal{P} \cap (\Phi\backslash\{\phi\} \times \Phi\backslash\{\phi\}) \rangle$, which by induction hypothesis is *true* if and only if $SE_\Sigma(\Lambda)$ is P'-preferred in Q. We show the right-hand side of the equivalence. Because $SE_\Sigma(\Lambda)$ is P-preferred in $SE_\Sigma(Q)$, there is a strict total ordering $d_1, \ldots, d_n$ of Φ that fulfills the condition in Definition 6.1. By Lemma 6.3 there is such an ordering where $d_1 = \phi$. We show that taking $d_2, \ldots, d_n$ to be the strict total ordering $\mathcal{T}$ of $\Phi\backslash\{\phi\}$, $SE_\Sigma(\Lambda)$ fulfills the condition and hence is P'-preferred in Q. Suppose there is $\psi \in (\Phi\backslash\{\phi\} \cap SE_\Sigma(\Lambda))\backslash SE_\Sigma(\Lambda')$ for some $\Lambda' \in Q$. Because $SE_\Sigma(\Lambda)$ is P-preferred in Q, there is some $\psi' \in (\Phi \cap SE_\Sigma(\Lambda'))\backslash SE_\Sigma(\Lambda)$ such that $\psi'\mathcal{T}\psi$. Because $\phi \in SE_\Sigma(\Lambda)$, ψ' has to be in $\Phi\backslash\{\phi\}$. Hence $SE_\Sigma(\Lambda)$ is P'-preferred in Q, and the value returned by the recursive call to *extendible* is *true*.

Let $\neg L\phi \in \Lambda$. The return value is the value returned by extendible$(P', \Lambda, \{\Lambda' \in Q | \neg L\phi \in \Lambda'\})$ where $P' = \langle \Phi\backslash\{\phi\}, \mathcal{P} \cap (\Phi\backslash\{\phi\} \times \Phi\backslash\{\phi\})\rangle$, which by induction hypothesis is *true* if and only if $SE_\Sigma(\Lambda)$ is P'-preferred in $\{\Lambda' \in Q | \neg L\phi \in \Lambda'\}$. We show the right-hand side of the equivalence. Because $SE_\Sigma(\Lambda)$ is preferred in $SE_\Sigma(Q)$, there is a strict total ordering $d_1, \ldots, d_n$ of Φ that fulfills the condition in Definition 6.1. By Lemma 6.3 there is such an ordering where $d_1 = \phi$. We show that taking $d_2, \ldots, d_n$ to be the strict total ordering $\mathcal{T}$ of $\Phi\backslash\{\phi\}$, $SE_\Sigma(\Lambda)$ fulfills the condition and hence is P'-preferred in Q. Suppose there is $\psi \in (\Phi\backslash\{\phi\} \cap SE_\Sigma(\Lambda))\backslash SE_\Sigma(\Lambda')$ for some $\Lambda' \in Q, \neg L\phi \in \Lambda'$. Because $SE_\Sigma(\Lambda)$ is P-preferred in Q, there is some $\psi' \in (\Phi \cap SE_\Sigma(\Lambda'))\backslash SE_\Sigma(\Lambda)$ such that $\psi'\mathcal{T}\psi$. Because $\phi \notin SE_\Sigma(\Lambda')$, ψ' has to be in $\Phi\backslash\{\phi\}$. Hence $SE_\Sigma(\Lambda)$ is P'-preferred in $\{\Lambda' \in Q | \neg L\phi \in \Lambda'\}$, and the value returned by the recursive call to *extendible* is *true*.

($\Rightarrow$) Assume b is *true*. Hence there is minimal $\phi \in \Phi$, such that $L\phi \in \Lambda$ implies $L\phi \in \bigcap Q$. Because $A \neq \emptyset$ the execution proceeds in the *else* branch of the outermost *if*-statement. We analyze by cases the innermost *if*-statement.

Let $L\phi \in \Lambda$. That extendible(P', Λ, Q) returns *true* implies by induction hypothesis that $SE_\Sigma(\Lambda)$ is P'-preferred in Q. This means that there is a strict total ordering $\mathcal{T} : d_1, \ldots, d_n$ of $\Phi\backslash\{\phi\}$ such that the condition in Definition 6.1 is fulfilled for $SE_\Sigma(\Lambda)$. We claim that the ordering $\phi, d_1, \ldots, d_n$ fulfills the condition for $SE_\Sigma(\Lambda)$. Assume there is $\psi \in (\Phi \cap SE_\Sigma(\Lambda))\backslash SE_\Sigma(\Lambda')$ for some $\Lambda' \in Q$. Because $L\phi \in \bigcap Q$, $\psi \in \Phi\backslash\{\phi\}$. Now by induction hypothesis there is $\psi' \in (\Phi\backslash\{\phi\} \cap SE_\Sigma(\Lambda'))\backslash SE_\Sigma(\Lambda)$ such that $\psi'\mathcal{T}\psi$. Therefore $SE_\Sigma(\Lambda)$ is P-preferred in Q.

Let $\neg L\phi \in \Lambda$. That extendible$(P', \Lambda, \{\Lambda' \in Q | \neg L\phi \in \Lambda'\})$ returns *true* implies $SE_\Sigma(\Lambda)$ is P'-preferred in $\{\Lambda' \in Q | \neg L\phi \in \Lambda'\}$. This means that there is a strict total ordering $\mathcal{T} : d_1, \ldots, d_n$ of $\Phi\backslash\{\phi\}$ such that the condition in Definition 6.1 is fulfilled for $SE_\Sigma(\Lambda)$. We claim that the ordering $\phi, d_1, \ldots, d_n$ fulfills the condition for $SE_\Sigma(\Lambda)$. Assume there is $\psi \in (\Phi \cap SE_\Sigma(\Lambda))\backslash SE_\Sigma(\Lambda')$ for some $\Lambda' \in Q$. If $L\phi \in \Lambda'$, take ϕ to be the element $\psi' \in (\Phi \cap SE_\Sigma(\Lambda'))\backslash SE_\Sigma(\Lambda)$. If $\neg L\phi \in \Lambda'$, then because $SE_\Sigma(\Lambda)$ is P'-preferred in $SE_\Sigma(\{\Lambda' \in Q | \neg L\phi \in \Lambda'\})$

there is a $\psi' \in (\Phi\backslash\{\phi\} \cap SE_\Sigma(\Lambda'))\backslash SE_\Sigma(\Lambda)$ and $\psi' \mathcal{T} \psi$. Therefore $SE_\Sigma(\Lambda)$ is P-preferred in Q.

The claim concerning b when Λ is a subset of a full set Λ' of a preferred stable expansion $SE_\Sigma(\Lambda')$ is as follows. The procedure tests only the membership of formulae $L\phi$ in the set Λ, not the presence or absence of formulae $\neg L\phi$. Hence, for a formula $\phi \in \Phi$ for which neither $L\phi \in \Lambda$ nor $\neg L\phi \in \Lambda$, the procedure works like $\neg L\phi \in \Lambda$ would hold. So assume $\Lambda_n = \Lambda \cup \{\neg L\phi | L\phi \in Sf^L(\Sigma)\backslash Sf^{qL}(\Lambda)\}$ is a full set. Whenever $\phi \in \Phi \cap SE_\Sigma(\Lambda_n)$, also $\phi \in \Phi \cap SE_\Sigma(\Lambda')$. Therefore, if $SE_\Sigma(\Lambda')$ is P-preferred in Q, $SE_\Sigma(\Lambda_n)$ is also P-preferred in Q. By the correctness of the algorithm, the procedure call extendible(P, Λ_n, Q) returns *true* if and only if $SE_\Sigma(\Lambda'')$ is P-preferred in Q. □

Lemma 6.5 (Procedure *next*) *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a finite set of formulae, $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization such that $\Phi \subseteq \{\phi | L\phi \in Sf^L(\Sigma)\}$, Λ a consistent subset of $Sf^L(\Sigma) \cup \{\neg L\phi | L\phi \in Sf^L(\Sigma)\}$ such that $Sf^L(\Sigma)\backslash Sf^{qL}(\Lambda) \neq \emptyset$, and Q a finite set of Σ-full sets.*

Then next(P, Σ, Λ, Q) returns a formula χ such that $L\chi \in Sf^L(\Sigma)\backslash Sf^{qL}(\Lambda)$ and the following holds. If S is a set of Σ-full sets $\Lambda' \supseteq \Lambda \cup \{\neg L\chi\}$ such that each $SE_\Sigma(\Lambda')$ is P-preferred in $SE_\Sigma(S \cup Q)$, and S' is a set of Σ-full sets $\Lambda' \supseteq \Lambda \cup \{L\chi\}$ such that each $SE_\Sigma(\Lambda')$ is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$, then each $SE_\Sigma(\Lambda'), \Lambda' \in S$ is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$.

Proof: The proof is by induction on $|\Phi|$. The base case corresponds to the *then*-part of the outermost *if*-statement, and the inductive case to the *else*-part.

Base case $|\Phi| = 0$: For the empty prioritization $P_e = \langle \emptyset, \emptyset \rangle$ every stable expansion is P_e-preferred. By Lemma 6.2 every stable expansion is P_e-preferred in $SE_\Sigma(S \cup S' \cup Q)$. Hence χ can be taken to be any formula such that $L\chi \in Sf^L(\Sigma)\backslash Sf^{qL}(\Lambda)$.

Inductive case $|\Phi| \geq 1$: We analyze by cases whether the sets A and $B \subseteq A$ are empty or non-empty. The case analysis corresponds to the *if* statements of the procedure.

(i) Let the set A be empty. Now no stable expansion $SE_\Sigma(\Lambda')$ such that $\Lambda' \supseteq \Lambda$ is a Σ-full set, can be P-preferred in Q, because for all minimal ϕ, $\phi \in SE_\Sigma(\Lambda')$ and for some $\Lambda'' \in Q$ $\phi \notin SE_\Sigma(\Lambda'')$. Hence the set S is empty, and no matter which element of $Sf^L(\Sigma)\backslash Sf^{qL}(\Lambda)$ is returned, the claim of the lemma holds.

(ii) Let B be non-empty. In this case the value χ returned is a member of B. The claim is that each member E of $SE_\Sigma(S)$ is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$. Let E be a member of $SE_\Sigma(S)$. By assumption E is P-preferred in $SE_\Sigma(S \cup Q)$. Let Λ_1 be a full set corresponding to E, i.e. $E = SE_\Sigma(\Lambda_1)$. Let $\mathcal{T}$ be a strict total order on the basis of which E is P-preferred in $SE_\Sigma(S \cup Q)$. Because $L\phi \in \Lambda$ implies $L\phi \in \bigcap Q$, by Lemma 6.3 the formula ϕ can be taken to be the least formula in the ordering. Because ϕ is in all members of $SE_\Sigma(S')$ and in no member of $SE_\Sigma(S)$, E is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$.

(iii) Let B be empty and A non-empty. In this case either $L\phi \in \Lambda$ or $\neg L\phi \in \Lambda$. The formula ϕ is a minimal element of Φ such that $L\phi \in \Lambda$ implies $L\phi \in \bigcap Q$. The prioritization P' is $\langle \Phi \backslash \{\phi\}, \mathcal{P} \cap (\Phi \backslash \{\phi\} \times \Phi \backslash \{\phi\}) \rangle$. We analyze by cases the last *if*-statement of the procedure.

Assume $\neg L\phi \in \Lambda$. The value χ returned is the one obtained from the recursive call to next$(P', \Sigma, \Lambda, \{\Lambda' \in Q | \neg L\phi \in \Lambda'\})$. Let S and S' be sets that fulfill the conditions stated in the lemma. Take any E in $SE_\Sigma(S)$. Now by induction hypothesis E is P'-preferred in $SE_\Sigma(S \cup S' \cup \{\Lambda' \in Q | \neg L\phi \in \Lambda'\})$, i.e. there is a strict total order $\mathcal{T}$ on $\Phi \backslash \{\phi\}$ that fulfills the condition in Definition 6.1. Let $\mathcal{T}'$ be the strict total order on Φ that extends $\mathcal{T}$ and has ϕ as the least element. We show that E is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$. Let $\Lambda_1 \supseteq \Lambda$ be the full set corresponding to E (i.e. $E = SE_\Sigma(\Lambda_1)$). Let Λ_2 be a either a member of $S \cup S'$ or a member of Q such that $\neg L\phi \in \Lambda_2$. Assume there is $\psi \in (\Phi \cap SE_\Sigma(\Lambda_1)) \backslash SE_\Sigma(\Lambda_2)$. Now ψ must be in $\Phi \backslash \{\phi\}$ because $\neg L\phi \in \Lambda_1$. Because by induction hypothesis $SE_\Sigma(\Lambda_1)$ is P'-preferred in $SE_\Sigma(S \cup S' \cup \{\Lambda' \in Q | \neg L\phi \in \Lambda'\})$, there is θ such that $\theta \mathcal{T}' \psi$ and $\theta \in (\Phi \cap SE_\Sigma(\Lambda_2)) \backslash SE_\Sigma(\Lambda_1)$. Assume $\Lambda_2 \in Q$ and $L\phi \in \Lambda_2$. If there is $\psi \in (\Phi \cap SE_\Sigma(\Lambda_1)) \backslash SE_\Sigma(\Lambda_2)$, then ϕ can be taken to be the element $\theta \in (\Phi \cap SE_\Sigma(\Lambda_2)) \backslash SE_\Sigma(\Lambda_1)$ such that $\theta \mathcal{T}' \psi$. Hence E is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$, if $\neg L\phi \in \Lambda$.

Assume $L\phi \in \Lambda$. Then $L\phi \in \bigcap Q$. The value χ returned is the one obtained from the recursive call to next(P', Σ, Λ, Q). Let S and S' be sets that fulfill the conditions stated in the lemma. Take any E in $SE_\Sigma(S)$. Now by induction hypothesis E is P'-preferred in $SE_\Sigma(S \cup S' \cup Q)$, i.e. there is a strict total order $\mathcal{T}$ on $\Phi \backslash \{\phi\}$ that fulfills the condition in Definition 6.1. Take $\mathcal{T}'$ be the strict total order on Φ that extends $\mathcal{T}$ and has ϕ as the least element. We show that E is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$. Let $\Lambda_1 \supseteq \Lambda$ be the full set corresponding to E (i.e. $E = SE_\Sigma(\Lambda_1)$). Let Λ_2 be a member of $S \cup S' \cup Q$. Assume there is $\psi \in (\Phi \cap SE_\Sigma(\Lambda_1)) \backslash SE_\Sigma(\Lambda_2)$. Now ψ must be in $\Phi \backslash \{\phi\}$ because $L\phi \in \Lambda_2$. Because by induction hypothesis $SE_\Sigma(\Lambda_1)$ is P'-preferred in $SE_\Sigma(S \cup S' \cup Q)$, there is θ such that $\theta \mathcal{T}' \psi$ and $\theta \in (\Phi \cap SE_\Sigma(\Lambda_2)) \backslash SE_\Sigma(\Lambda_1)$. Hence E is P-preferred in $SE_\Sigma(S \cup S' \cup Q)$, if $L\phi \in \Lambda$. $\square$

Lemma 6.6 (Procedure der_p) *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a finite set of formulae, $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization such that $\Phi \subseteq \{\phi | L\phi \in Sf^L(\Sigma)\}$, $B \cup F$ a consistent set such that $B \subseteq (Sf^L(\Sigma) \cup \{\neg L\phi | L\phi \in Sf^L(\Sigma)\})$ and $F \subseteq Sf^L(\Sigma)$, $\phi \in \mathcal{L}_{ae}$ a formula, and Q a finite set of Σ-full sets.*

The procedure call $der_p(P, \Sigma, B, F, \phi, Q)$ returns (S,b). The set S consists of Σ-full sets $\Lambda \supseteq B \cup F$ such that each $SE_\Sigma(\Lambda)$ is P-preferred in $SE_\Sigma(S \cup Q)$, and each P-preferred stable expansion E of Σ such that $E = SE_\Sigma(\Lambda)$ for some $\Lambda \supseteq B \cup F$, is contained in S. The value b is true *if and only if for some $\Lambda \in S$, test(Σ, Λ, ϕ) returns* true.

Proof: By induction on the size of the set $U = Sf^L(\Sigma) \backslash (Sf^{qL}(B) \cup F)$.

Base case $|U| = 0$: Assume that $F \not\subseteq B$ or for some $\neg L\chi \in B, \Sigma \cup B \cup F \models \chi$. By the correctness of Niemelä's algorithms the set $B \cup F$ is not Σ-full, and ($\emptyset$,*false*) is returned. This is as required, because in this case there are no full sets in S. Assume that $B \cup F = \Lambda$ is Σ-full. By Lemma 6.4, the procedure call extendible(P, Λ, Q) returns *true* if and only if $SE_\Sigma(\Lambda)$ is P-preferred in Q. If the call returns *false*, then ($\emptyset$,*false*) is returned. If the call returns *true*, then ($\{B\}$,test(Σ, B, ϕ)) is returned. In both cases the procedure behaves as claimed. Furthermore, if $SE_\Sigma(\Lambda)$ is P-preferred it is by Lemma 6.2 also P-preferred in Q, and hence contained in S.

Inductive case $|U| \geq 1$: All full sets $\Lambda \supset B \cup F$ contain $\{L\chi \in Sf^L(\Sigma) | \Sigma \cup B \cup F \models \chi\}$. This is by properties of $\models$ and full sets. Hence the call to the procedure is$_m$ does not rule out any full sets from consideration.

The case when the execution of the statement $(B, F) := is_m(\Sigma, B, F)$ reduces U to the empty set is proved exactly like the base case of the induction.

Assume U is non-empty. By induction hypothesis all P-preferred stable expansions $E = SE_\Sigma(\Lambda)$ of Σ such that $\Lambda \supseteq B \cup \{\neg L\chi\} \cup F$, are contained in S. Similarly for $\Lambda \supseteq B \cup F \cup \{L\chi\}$ and S'. Hence the full sets Λ of all P-preferred stable expansions $E = SE_\Sigma(\Lambda)$ of Σ such that $\Lambda \supseteq B \cup F$ are contained in $S \cup S'$.

By induction hypothesis all elements of $SE_\Sigma(S)$ are P-preferred in $SE_\Sigma(S \cup Q)$ and all elements of $SE_\Sigma(S')$ are P-preferred in $SE_\Sigma(S \cup S' \cup Q)$. That all elements of $SE_\Sigma(S)$ are P-preferred in $SE_\Sigma(S \cup S' \cup Q)$ follows from Lemma 6.5. Therefore all elements of $SE_\Sigma(S \cup S')$ are P-preferred in $SE_\Sigma(S \cup S' \cup Q)$.

By induction hypotheses b is *true* if and only if for some $\Lambda \in S$ test(Σ, Λ, ϕ) returns *true* and b' is *true* if and only if for some $\Lambda \in S'$ test(Σ, Λ, ϕ) returns *true*. Therefore (b or b') is *true* if and only if for some $\Lambda \in S \cup S'$ test(Σ, Λ, ϕ) returns *true*. $\square$

Theorem 6.7 (Correctness) *Let $\Sigma \subseteq \mathcal{L}_{ae}$ be a finite set of formulae, $P = \langle \Phi, \mathcal{P} \rangle$ a prioritization such that Φ is finite, and $\phi \in \mathcal{L}_{ae}$ a formula. The procedure call* $der_p(P, \Sigma \cup \{L\phi \rightarrow L\phi | \phi \in \Phi, L\phi \notin Sf^L(\Sigma)\}, \emptyset, \emptyset, \phi, \emptyset)$ *returns (S, b), where b is* true *if and only if there is a P-preferred stable expansion $SE_\Sigma(\Lambda)$ of Σ such that the procedure call test(Σ, Λ, ϕ) returns* true.

Proof: Let $\Sigma' = \Sigma \cup \{L\phi \rightarrow L\phi | \phi \in \Phi, L\phi \notin Sf^L(\Sigma)\}$. The set of stable expansions of Σ is exactly the same set as the set of stable expansions of Σ', because all formulae in $\Sigma' \backslash \Sigma$ are tautologies. The idea in extending Σ to Σ' is to fulfill the requirement that for each formula $\phi \in \Phi$ in the prioritization, $L\phi \in \Sigma$. The procedures *next*, *extendible* and der_p depend on this requirement.

By Lemma 6.6 $SE_{\Sigma'}(S)$ contains all P-preferred stable expansions. Because each member of $SE_{\Sigma'}(S)$ is P-preferred in $SE_{\Sigma'}(S)$, each of its members is P-preferred by Lemma 6.2. The value of b is *true* if and only if some of the stable expansions $SE_{\Sigma'}(\Lambda), \Lambda \in S$ is such that test(Σ', Λ, ϕ) returns *true*. $\square$

The decision procedures for prioritized cautious and brave reasoning are obtained by supplying different procedures *test*. For prioritized brave reasoning the procedure *test* simply tests $\Sigma \cup \Lambda \models_L \phi$, and for prioritized cautious reasoning it tests $\Sigma \cup \Lambda \not\models_L \phi$ and the value finally returned by der_p is negated. Procedures for testing the existence of preferred stable expansions or for computing all preferred stable expansions can be constructed in a similar way.

The algorithm in Figures 8,9 is given in a form that makes its correctness proof more concise. An improvement that is useful to make in the implementations of the algorithm is to terminate the computation immediately after the first preferred stable expansion with test(Σ, Λ, ϕ) returning *true* is found. This can be achieved by replacing the innermost begin-end block of the main procedure by the piece of code in Figure 10.

Some small improvements can be made to the book-keeping mechanism of already-found preferred stable expansions and the calls to the procedures *extendible* and *next*. There currently seem to be much redundant computation in this part of the algorithm, though the effect on the efficiency is small. Developments in the theorem-proving techniques for ordinary autoepistemic logic are likely to benefit also the decision procedures for prioritized autoepistemic logic.

6.2 An automatic theorem-prover

An automatic theorem-proving system for nonmonotonic logics has been implemented by us in Standard ML [Milner and Tofte, 1991]. The system con-

```
begin
  χ := next(P, Σ, B ∪ F, Q);
  (S, b) := der_p(P, Σ, B ∪ {¬Lχ}, F, φ, Q);
  if b = true
  then return (S, true)
  else
    begin
      (S', b') := der_p(P, Σ, B, F ∪ {Lχ}, φ, S ∪ Q);
      return (S ∪ S', b')
    end
end
```

Figure 10: An improvement to the decision procedure

sists of about 3000 lines or 95kB of source code (excluding the parser and the lexical analyzer). The system implements theorem-provers for Moore's original autoepistemic logic, L-hierarchic autoepistemic logic of Niemelä [1991], default logic of Reiter [1980], prioritized autoepistemic logic defined by the author (the algorithm is described in Section 6.1), and the stratified subset of autoepistemic logic [Marek and Truszczyński, 1991; Niemelä and Rintanen, 1992].

The theorem-proving algorithms for the Moore and the L-hierarchic versions of autoepistemic logic were given by [Niemelä, 1993]. The theorem-provers for default logic are implemented using a simple translation to L-hierarchic autoepistemic logic. The correctness of the translation was proved by [Niemelä, 1991]. The implementations of all these algorithms are parameterized with respect to the theorem-prover for a classical logic. This means that addition of new classical theorem-provers is easy. Currently there are theorem-provers for classical propositional logic (one based on the analytic tableau of [Smullyan, 1968] and another on a variant of the Davis–Putnam procedure as given in [Lee and Plaisted, 1992]) and its Horn clause subset. The Horn subset is useful because of its low time complexity: there is a linear time algorithm for testing satisfiability of sets of propositional Horn clauses [Dowling and Gallier, 1984]. Hence each of the afore-mentioned logics can be used as two versions, one based on classical propositional logic and another based on the more efficient (though less expressive) propositional Horn clause logic.

The undecidability of the first-order predicate logic implies the undecidability of nonmonotonic logics based on it. For example, autoepistemic logic is not even semi-decidable when its underlying classical logic is semidecidable. This makes the use of classical predicate logic for nonmonotonic theorem-proving inconvenient. To support the expression of knowledge-bases of some practical value in the theorem-prover, a facility that supports the expression of formula

```
Decision procedures for AEL and DL v0.0 February 1993 (C) Jussi Rintanen
set nixon [ cons $a : nixon, clyde ;
republican(nixon), quaker(nixon),
republican($a) & -L-pacifist($a) => pacifist($a),
quaker($a) & -Lpacifist($a) => -pacifist($a) ];
Theory nixon defined: 6 formulae.
show nixon;
{republican(nixon),
quaker(nixon),
republican(nixon)&-L-pacifist(nixon)=>pacifist(nixon),
republican(clyde)&-L-pacifist(clyde)=>pacifist(clyde),
quaker(nixon)&-Lpacifist(nixon)=>-pacifist(nixon),
quaker(clyde)&-Lpacifist(clyde)=>-pacifist(clyde)}
Execution time 0.01 seconds
expansions ael nixon;
2 expansions:
0: {-Lpacifist(clyde), -Lpacifist(nixon), -L-pacifist(clyde), L-pacifist(nixon)}
1: {-Lpacifist(clyde), Lpacifist(nixon), -L-pacifist(clyde), -L-pacifist(nixon)}
Execution time 0.14 seconds
cautious ael nixon pacifist(nixon);
Formula pacifist(nixon) is not in all expansions
Execution time 0.09 seconds
extensions dl { p, p : q / q, p : -q / -q };
2 extensions:
0: {L-q, Lp, -L--q}
1: {-L-q, Lp, L--q}
Execution time 0.06 seconds
expansions pael [ cons $a : tweety, clyde;
   { flies(tweety) < -flies(tweety) }
bird(tweety), penguin(tweety), bird(clyde),
bird($a) & -L-flies($a) => flies($a),
penguin($a) & -Lflies($a) => -flies($a)];
1 preferred stable expansion:
0: {Lflies(clyde), -Lflies(tweety), -L-flies(clyde), L-flies(tweety)}
Execution time 0.07 seconds
```

Figure 11: An example run of the automatic theorem-prover

schemata and the instantiation of them has been implemented. Using this facility, a set of formula schemata is replaced by a larger set of ordinary propositional formulae before giving the set to a theorem-proving algorithm. A sample run of the theorem-proving system is given in Figure 11.

7 Priorities in diagnosis and inheritance reasoning

7.1 Diagnosis

In this section we show how priorities can be used for restricting the number of diagnoses computed for a diagnosis problem. The section builds on the work of Reiter [1987] the new ingredient being the use of priorities. Reiter defines a framework of *consistency-based diagnosis* for doing model-based diagnosis. He indicates the connection of consistency-based diagnosis to nonmonotonic reasoning by showing how it can be formalized in default logic.

A *system with observations* is a triple (SD,COMPONENTS,OBS) where SD is a set of formulae (the system description), COMPONENTS is the set of components of the system that can possibly fail, and OBS is a set of atomic formulae (the observations). A *diagnosis* for (SD,COMPONENTS,OBS) is a minimal set $\Delta \subseteq$ COMPONENTS such that

$$SD \cup OBS \cup \{ab(c)|c \in \Delta\} \cup \{\neg ab(c)|c \in COMPONENTS\backslash\Delta\}$$

is consistent. Informally, a diagnosis is a minimal set of components that have to be assumed to behave incorrectly in order to avoid contradicting the observations and the system description. We demonstrate this definition and the use of priorities with it using an example system that is a circuit consisting of *and*, *or*, *xor* and *not* gates. The circuit computes the arithmetical sum of two sequences of binary digits (Reiter gives a similar example). Part of our circuit is given in Figure 12. The circuit has three inputs, the carry (C_{i-1}) and two binary digits (A_i and B_i), and two outputs, the carry (C_i) and the sum (S_i). An adder for n digits can be obtained by connecting n units in sequence, each carry output to the next carry input.

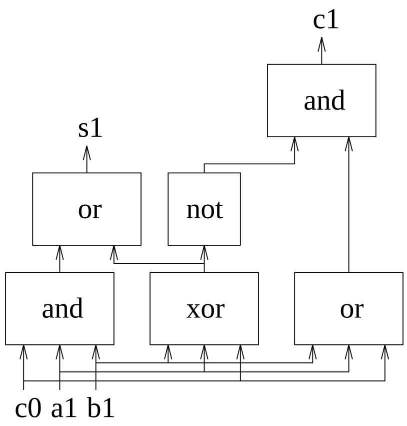


Figure 12: A circuit for adding two binary numbers

The representation of this circuit in autoepistemic logic is straightforward. For example, the and-gate in unit i with three inputs corresponds to the following formulae of the system description SD.

$$\begin{array}{l} on(a_i) \wedge on(b_i) \wedge on(c_i) \rightarrow (\neg ab(and_i) \leftrightarrow on(and_i)) \\ \mathit{off}(a_i) \rightarrow (\neg ab(and_i) \leftrightarrow \mathit{off}(and_i)) \\ \mathit{off}(b_i) \rightarrow (\neg ab(and_i) \leftrightarrow \mathit{off}(and_i)) \\ \mathit{off}(c_i) \rightarrow (\neg ab(and_i) \leftrightarrow \mathit{off}(and_i)) \end{array}$$

For instance, the first formula says that if the inputs a_i, b_i and c_i are *on*, then the output of the and-gate is *on* if and only if the gate works properly (i.e. if it is not abnormal). Similar sets of formulae describe other gates of the circuit. By default, all gates work correctly but when the correctness assumption contradicts the observed behaviour of the circuit, it has to be given up. This correctness assumption is formalized using formulae of the following form.

$$\neg Lab(and_i) \rightarrow \neg ab(and_i)$$

The observations OBS concerning the inputs and outputs of the circuit are directly given as atomic formulae, for example

$$on(a_1), on(b_1), \mathit{off}(c_1), \mathit{off}(s_1), on(c_2).$$

The set COMPONENTS consists of the names of the gates and_1, or_1 and so on. The connection between the atomic formulae involving *on* and *off* are established simply using equivalences $on(P) \leftrightarrow \neg \mathit{off}(P)$ for inputs and gates P.

Reiter shows how consistency-based diagnosis can be naturally formalized within default logic. The diagnoses of a system are directly obtained as the extensions of a default theory. For each component there is a default rule that states the correctness assumption: if it is consistent to believe that the component is working correctly, believe so.

Theorem 7.1 ([Reiter, 1987]) *Consider a system (SD,COMPONENTS) under observation OBS where SD and OBS are sets of formulae. Then E is an extension for the default theory*

$$DT = \left(\left\{ \frac{: \neg ab(c)}{\neg ab(c)} | c \in COMPONENTS \right\}, SD \cup OBS \right)$$

if and only if for some diagnosis Δ *for (SD,COMPONENTS,OBS),* $E = \{\Pi | \Delta \text{ predicts } \Pi\}$ *(where* Δ *predicts* Π *means* $SD \cup OBS \cup \{ab(c) | c \in \Delta\} \cup \{\neg ab(c) | c \in COMPONENTS \backslash \Delta\} \models \Pi$*).*

Because of the close connection between default logic and autoepistemic logic Reiter's theory of diagnosis can be easily expressed also in autoepistemic logic.

Corollary 7.2 *Consider a system (SD,COMPONENTS) under observation OBS where SD and OBS are sets of formulae. Then E is a stable expansion of the set of formulae*

$$DT = \{\neg Lab(c) \rightarrow \neg ab(c) | c \in COMPONENTS\} \cup SD \cup OBS$$

if and only if for some diagnosis Δ *for (SD,COMPONENTS,OBS),* $E \cap \mathcal{L} = \{\Pi | \Delta \textit{ predicts } \Pi\}$.

Proof: The result is immediate by the correspondence between default theories with rules of the form $\frac{:\phi}{\phi}$ and autoepistemic theories with formulae of the form $\neg L \neg \phi \rightarrow \phi$. □

If the inputs and outputs conform to the model of the circuit, there will be only one diagnosis (stable expansion) that corresponds to the situation where all gates work correctly. In case of incorrect behaviour, there may be several stable expansions corresponding to the different diagnoses that explain the behaviour. In this multiple expansions situation it may be useful to compute only a subset of all the diagnoses if some of the gates are more likely to misbehave than others. This is where the priorities come in. For example that the NOT-gate is less likely to fail than the XOR-gate, can be expressed as follows.

$$ab(not_1) \mathcal{P} ab(xor_1)$$

Consider the circuit in Figure 12 and let the inputs C_0, A_1, B_1 be *off, on, on* respectively, and let the outputs S_1, C_1 be *off* and *off* respectively. Clearly there has to be something wrong, because the sum $1 + 1$ is 10 in binary, not 00. The diagnoses are the following.

1: $or2_1, xor3_1$
2: $or3_1$
3: not_1
4: $and2_1$

Take for example the first diagnosis. It indicates that the observed behaviour is possible if the XOR-gate and the OR-gate with two inputs behave incorrectly. That is, the output of the XOR-gate is incorrectly *on*, and the output of the NOT-gate is correctly *off* and consequently the output carry is *off*.

Because the XOR-gate outputs *on* the OR-gate with two inputs must be broken as well to explain its output *off*. Let us assume that for some reason the OR-gates are much less likely to behave incorrectly than the others. This assumption can be given as a prioritization P where $ab(or3_1)\mathcal{P}ab(\gamma)$ and $ab(or2_1)\mathcal{P}ab(\gamma)$ for all other gates γ. For this prioritization the preferred stable expansions correspond to the diagnoses 3 and 4, just as expected.

The computation time of all four stable expansions on the theorem-proving system described in the previous section is 44.8 seconds. With the priorities, the computation of the two preferred stable expansions is only 6.87 seconds.

7.2 Inheritance reasoning

In this section we discuss the use of priorities in inheritance reasoning. There are two kinds of conflicts that arise in inheritance. First, there are conflicts between attribute values in superclasses and subclasses. These conflicts are always solved in favour of the values in subclasses. We discuss this subject in connection with a variant of a theory of single inheritance developed by Gelfond and Przymusinska [1990]. The second kind of conflicts in inheritance arise when multiple inheritance, i.e. the possibility of more than one parent classes for a class or an instance is allowed. As an example of such an inheritance theory we give a variant of a theory developed by Stein [1992].

7.2.1 Expressing specificity as priorities

Gelfond and Przymusinska [1990] present several theories of inheritance that are based directly on autoepistemic logic. We choose the simplest theory that supports single-inheritance with both strict and defeasible positive links and defeasible negative links. There are three disjoint classes of entities, objects, classes and properties. We denote these classes respectively by Ω, Γ, Π. Each object can be connected to a number of classes by strict positive links (PS), each class to a number of classes by strict positive links and to a number of properties by positive or negative defeasible links (PD,ND). That an object X has a certain property P is expressed as $\mathrm{H}(X, C)$.

The first four axiom schemata describe the propagation of properties.

1.1 $\mathrm{H}(X, C)$ if $X \in \Omega, C \in \Gamma$ and there is a strict link from X to C
1.2 $L\mathrm{PS}(c_1, c_2) \wedge L\mathrm{H}(x, c_1) \rightarrow \mathrm{H}(x, c_2)$
1.3 $L\mathrm{PD}(c, p) \wedge L\mathrm{H}(x, c) \wedge \neg L\mathrm{ab}(x, c, p) \rightarrow \mathrm{H}(x, p)$
1.4 $L\mathrm{ND}(c, p) \wedge L\mathrm{H}(x, c) \wedge \neg L\mathrm{ab}(x, c, p) \rightarrow \neg\mathrm{H}(x, p)$

The abnormality predicates above prevent the propagation of a property

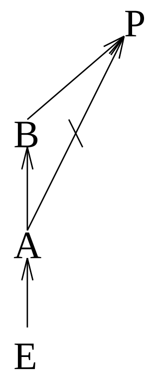


Figure 13: A simple inheritance hierarchy

through a link that is defeated by a more specific link. First there are two axiom schemata for expressing that two links are incompatible.

3.1 $\mathrm{PD}(c_1,p) \wedge \mathrm{ND}(c_2,p) \rightarrow \mathrm{I}(c_1,c_2,p)$
3.2 $\mathrm{I}(c_1,c_2,p) \rightarrow \mathrm{I}(c_2,c_1,p)$

Gelfond and Przymusinska code specificity in the following axiom schemata.

2.1 $\mathrm{PS}(c_1,c_2) \rightarrow \mathrm{MS}(c_1,c_2)$
2.2 $\mathrm{MS}(c_1,c) \wedge \mathrm{MS}(c,c_2) \rightarrow \mathrm{MS}(c_1,c_2)$
5.1 $L\mathrm{H}(x,c_1) \wedge L\mathrm{H}(x,c_2) \wedge L\mathrm{I}(c_1,c_2,p) \wedge L\mathrm{MS}(c_1,c_2) \wedge \neg L\mathrm{ab}(x,c_1,p)$
$\rightarrow \mathrm{ab}(x,c_2,p)$

Instead of using the above three schemata we can use priorities. The axiom schema 5.1b below replaces 5.1 and the prioritization P defined as $P = \langle \Phi, \mathcal{P} \rangle$ replaces the axiom schemata 2.1 and 2.2.

$$\begin{aligned} \Phi &= \{\mathrm{ab}(x,c_1,c_2) | x \in \Omega, \{c_1,c_2\} \subseteq \Gamma\} \\ \mathcal{P} &= \{(\mathrm{ab}(x,c_1,p),\mathrm{ab}(x,c_2,p)) | x \in \Omega, p \in \Pi, \{c_1,c_2\} \subseteq \Gamma, \\ &\qquad \text{there is a strict link from } c_1 \text{ to } c_2\}. \end{aligned}$$

5.1b $L\mathrm{H}(x,c_1) \wedge L\mathrm{H}(x,c_2) \wedge L\mathrm{I}(c_1,c_2,p) \wedge \neg L\mathrm{ab}(x,c_1,p) \rightarrow \mathrm{ab}(x,c_2,p)$

The improvement over the original representation of Gelfond and Przymusinska is minor. However, there are likely to be computational benefits due to the smaller number of formulae and the advantage of having priorities in computing the stable expansions. Results pointing to this direction are given in Section 6.

Consider the simple inheritance hierarchy in Figure 13. There is an object E that belongs to the class A, which is a subclass of the class B. Member of class A typically do *not* have the property P (as indicated by the negative edge), and members of class B typically do have the property P. The instances of the axiom schemata above together with the atomic formulae $\mathrm{PS}(A, B), \mathrm{ND}(A, P), \mathrm{PD}(B, P), \mathrm{H}(E, A)$ have two stable expansions, one containing the formulae $\mathrm{ab}(E, A, P)$ and $\mathrm{H}(E, P)$ that indicate that it is abnormal for E to inherit property P through the class A and that E has the property P, and the other containing the formulae $\mathrm{ab}(E, B, P)$ and $\neg\mathrm{H}(E, P)$ that indicate that it is abnormal for E to inherit property P through the class B and that E does not have the property P. Not both of these stable expansions are P-preferred. Because $\mathrm{ab}(E, A, P)\mathcal{P}\mathrm{ab}(E, B, P)$, only the latter stable expansion is P-preferred, i.e. the one that indicates that E does not have the property P.

7.2.2 Priorities in multiple inheritance

In [Stein, 1992] combinations of skeptical/credulous, on-path/off-path theories of inheritance are developed. Stein also gives polynomial time algorithms for several decision problems. As a basis of our example we take the inheritance theory with off-path pre-emption developed in [Stein, 1992, Section 8]. We show how a similar theory can be embedded in prioritized autoepistemic logic, and how resolution of conflicts caused by multiple inheritance can be performed using priorities.

The set of edges in an inheritance hierarchy Γ is denoted by E_Γ. Sequences of edges $(a_1, a_2), (a_2, a_3), \ldots, (a_{n-1}, a_n)$ are written as $a_1 \cdot a_2 \cdot \cdots \cdot a_n$. If the last edge is negative (defeating) the path is written as $a_1 \cdot a_2 \cdot \cdots a_{n-1} \cdot \neg a_n$. For her off-path inheritance, Stein defines *admissibility* as

> An edge $v \cdot x$ (respectively, $v \cdot \neg x$) is off-path admissible in Γ w.r.t. a if no truly unambiguous off-path admissible prefix $a \cdot s_1 \cdots s_n \cdot v$ contains a preemptor of $v \cdot x$ (respectively, $v \cdot \neg x$).

The definition of a preemptor is as follows.

> A node s is a preemptor of $v \cdot x$ (respectively, $v \cdot \neg x$) w.r.t a if $s \cdot \neg x \in E_\Gamma$ (respectively, $s \cdot x \in E_\Gamma$).

Admissibility is used in the definition of supported inference paths:

> We say that Γ off-path supports $a \rightarrow x$ (respectively $a \not\rightarrow x$) if there is some sequence of off-path admissible edges $a \cdot s_1 \cdots s_n \cdot x$ (respectively, $a \cdot s_1 \cdots s_n \cdot \neg x$) in Γ w.r.t a.

These definitions can be expressed in autoepistemic logic as follows. The positive defeasible edges and negative defeasible edges are written respectively as $\mathrm{PD}(c_1, c_2)$ and $\mathrm{ND}(c_1, c_2)$. That $v \cdot \ldots \cdot w$ is a part of an admissible path with respect to a is expressed as $\mathrm{AP}(a, v, w)$. That $v \cdot \ldots \cdot \neg w$ is a suffix of an admissible path with respect to a is expressed as $\mathrm{NAP}(a, v, w)$. Similarly, $\mathrm{AE}(a, v, w)$ and $\mathrm{NAE}(a, v, w)$ refer to edges $v \cdot w$ and $v \cdot \neg w$ in paths that are admissible w.r.t. a. That there is a preemptor of the edge $v \cdot x$ on an admissible path from a to v with respect to a is expressed as $\mathrm{PE}(a, v, x)$. Or if the edge is negative, $\mathrm{NPE}(a, v, x)$.

$$\begin{aligned}
\mathrm{PD}(a, c) &\rightarrow \mathrm{AE}(a, a, c) && (2)\\
\mathrm{AP}(a, s, v) \wedge \mathrm{PD}(v, x) \wedge \neg L\mathrm{PE}(a, v, x) &\rightarrow \mathrm{AE}(a, v, x) && (3)\\
\mathrm{AP}(a, s, v) \wedge \mathrm{ND}(v, x) \wedge \neg L\mathrm{NPE}(a, v, x) &\rightarrow \mathrm{NAE}(a, v, x) && (4)\\
\mathrm{AP}(a, s, v) \wedge \mathrm{NAE}(a, v, w) &\rightarrow \mathrm{NAP}(a, s, w) && (5)\\
\mathrm{AE}(a, s, v) &\rightarrow \mathrm{AP}(a, s, v) && (6)\\
\mathrm{AP}(a, s, v) \wedge \mathrm{AE}(a, v, w) &\rightarrow \mathrm{AP}(a, s, w) && (7)
\end{aligned}$$

Axioms for preemption are the following.

$$\begin{aligned}
\mathrm{AP}(a, s, v) \wedge \mathrm{ND}(s, x) &\rightarrow \mathrm{PE}(a, v, x) && (8)\\
\mathrm{AP}(a, s, v) \wedge \mathrm{PD}(s, x) &\rightarrow \mathrm{NPE}(a, v, x) && (9)
\end{aligned}$$

The definitions above (and their translations to autoepistemic logic) determine the set of admissible edges and paths. A path is admissible if it is not defeated by specificity considerations. There can still be paths from a node to another node some of which end at a negative edge and others at a positive edge. Next we show how these kinds of conflicts are resolved. A number of axioms are given for class membership. Priority information is used for solving the conflicts that arise in the presence of pairs of conflicting edges.

$$\begin{aligned}
& && \mathrm{HO}(a, v, v) && (10)\\
\mathrm{AE}(a, v, w) \wedge \mathrm{HO}(a, w, p) \wedge \neg ab(a, v, w, p) &\rightarrow && \mathrm{HO}(a, v, p) && (11)\\
\mathrm{NAE}(a, v, w) \wedge \mathrm{HO}(a, w, p) \wedge \neg ab(a, v, w, p) &\rightarrow && \mathrm{NHO}(a, v, p) && (12)\\
\mathrm{AE}(a, v, w) \wedge \mathrm{NHO}(a, w, p) \wedge \neg ab(a, v, w, p) &\rightarrow && \mathrm{NHO}(a, v, p) && (13)\\
\mathrm{HO}(a, v, w) &\rightarrow && \neg\mathrm{NHO}(a, v, w) && (14)\\
\neg L ab(a, v, w, p) &\rightarrow && \neg ab(a, v, w, p) && (15)
\end{aligned}$$

Axiom 10 just states the reflexivity of subset relation. Axioms 11-13 propagate properties downward through positive and negative edges. Axiom 14 states the incompatibility of $\mathrm{HO}(a, v, w)$ and $\mathrm{NHO}(a, v, w)$. Axiom 15 states that as many properties as possible should be allowed to propagate from w to v with respect to a.

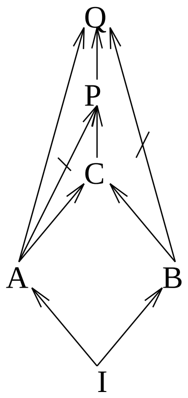


Figure 14: An inheritance hierarchy with multiple inheritance

The idea here is that the immediate parent classes of an object and the immediate superclasses of a class are put into a total order. This is an often used practise in object-oriented systems that support multiple inheritance. When in conflict, the least class in the ordering takes precedence over other parent (or super) classes. The ordering is expressed as a prioritization of the atomic formulae $ab(a, v, w, p)$. If inheriting properties from w_1 to v is preferred to inheriting them from w_2 to v, then $ab(a, v, w_1, p)\mathcal{P}ab(a, v, w_2, p)$ (for all focus nodes a and nodes or properties p).

We demonstrate the inheritance theory using the hierarchy in Figure 14. This example is expressed in autoepistemic logic as the set of formula Σ which consists of instances of the axiom schemata above and the following atomic formulae.

$$\mathrm{PD}(I, A), \mathrm{PD}(I, B), \mathrm{PD}(A, C), \mathrm{PD}(B, C),$$

$$\mathrm{PD}(C, P), \mathrm{ND}(A, P), \mathrm{PD}(A, Q), \mathrm{ND}(B, Q)$$

There are two stable expansions E_1 and E_2 of Σ. In this case, whether the element I has the property P or not is determined by the preemption strategy of the theory. The off-path strategy says that the positive edge from C to P is preempted because there is at least one subclass of C from which there is a negative link to P. The class represented by the node A is such. Hence the path $I \cdot A \cdot \neg P$ is admissible, but the path $I \cdot B \cdot C \cdot P$ is not. The ambiguity about the property Q however remains. The path $I \cdot A \cdot Q$ as well as the path $I \cdot B \cdot \neg Q$ are admissible. Both E_1 and E_2 contain $\mathrm{AE}(a, b, c)$ for all admissible edges $b \cdot c$ in $\{I \cdot A, I \cdot B, A \cdot C, B \cdot C, A \cdot Q\}$ with respect to a, and $\mathrm{NAE}(a, b, c)$ for all admissible edges $b \cdot \neg c$ in $\{A \cdot \neg P, B \cdot \neg Q\}$ with respect to a.

In concluding whether an object a has a property p or not (whether $\mathrm{HO}(a, p, p)$ or $\mathrm{NHO}(a, p, p)$), not all admissible edges can be used. Each stable expansion may contain only one of atomic formulae $\mathrm{HO}(a, v, p)$ or $\mathrm{NHO}(a, v, p)$ because of Axiom 14: The conclusion of these formulae using Axioms 11–13 has as a prerequisite formulae $\neg ab(a, v, w, p)$, which are included in the stable expansion only as long as consistency is preserved. Let the prioritization P be such that $ab(\theta, I, A, \pi)\mathcal{P}ab(\theta, I, A, \pi)$ for all focus nodes θ and for all nodes (properties) π. In our example, each stable expansion contains only one of $ab(I, I, A, Q)$ and $ab(I, I, B, Q)$. In this case $ab(I, I, B, Q) \in E_1$ and $ab(I, I, A, Q) \in E_2$. The prioritization P resolves the inheritance conflict in favour of A: only E_1 is P-preferred and E_1 contains the formula $\mathrm{HO}(I, I, Q)$.

8 Conclusions

In this work, we have presented a formalism for nonmonotonic reasoning called prioritized autoepistemic logic. We have analyzed its properties and given decision procedures for it. This logic is a proper extension of autoepistemic logic [Moore, 1985]. Autoepistemic logic – like other nonmonotonic logics – differ from classical logics in several ways. Most importantly, it is nonmonotonic and was designed for expressing reasoning using defaults. Secondly, unlike classical (monotonic) logics, in general, there is no unique set of consequences for a set of formulae. This property of multiple extensions is caused by the presence of defaults: different extensions are obtained depending on which way conflicts between defaults are resolved. The priority mechanism of prioritized autoepistemic logic can be used for preventing conflict resolution in unintended ways. Hence in several cases the number of stable extensions (or stable expansions in the case of autoepistemic logic) can be made smaller than what it would be in Moore's autoepistemic logic. A suitable prioritization can restrict the number of stable expansions to one. As the standard way of dealing with multiple extensions is to take the intersection of them, supplying priorities thus increases the conclusions that can be derived from a set of formulae.

Our prioritization mechanism is general in the sense that it is not tightly tied to autoepistemic logic. The mechanism can be easily transferred to any other formalism with multiple extensions, e.g. other modal nonmonotonic logics [Marek *et al.*, 1991], logic programs (stable model semantics [Gelfond and Lifschitz, 1988]), and default logic. Our prioritization mechanism can be seen from an abstract point of view. There is a space of objects (extensions), a number of points (formulae in the prioritization), a partial order on the points, and a method for comparing the objects given the points. By supplying an alternative space and points, the mechanism can be applied in different contexts. The preferred subtheories of Brewka [1991] is an instance of our notion of priorities. Here the space consists of maximal consistent subsets of a set of formulae, and the set of points is the set of formulae. Our alternative notion of priorities in Section 5 can be seen similarly. Ryan's structured theories are an instance of the framework as follows. The objects are models, the points are formulae, and the comparison method uses the natural consequences of the points for comparing.

Despite the loose coupling between autoepistemic logic and our priority mechanism, we have been able to demonstrate that the decision procedure for prioritized autoepistemic logic can benefit from the priorities. Priorities give a justification for pruning the search trees traversed in the decision procedures. Unfortunately, there is typically no asymptotic decrease in the resources needed. The existence problem of stable expansions coincides with the existence problem of preferred stable expansions, and these problems are polynomial time Turing reducible to respective problems of cautious reasoning. The problem of existence of stable expansions is solvable in polynomial time only in syntactically very restricted cases. A class that becomes tractable with sufficient priorities is that of propositional Horn clauses and autoepistemic formulae of the form $\neg L \neg \phi \rightarrow \phi$ where ϕ is a Horn clause, which corresponds to a superclass of a class of default theories shown NP-complete in [Stillman, 1990].

There are some important questions from the point of view of practical applicability of prioritized autoepistemic logic. First, we have demonstrated the computational benefits of priorities in the case when priorities are on the L-subformulae of premises. What are the computational benefits when the priorities are *not* on these formulae? This question could be answered by directly trying to devise efficient theorem-proving algorithms for arbitrary prioritizations. However, because the prioritizations on the L-subformulae seem to be primary from the computational point of view, it seems that a more useful approach would be to investigate how an arbitrary prioritization can be replaced by a prioritization on L-subformulae. This investigation would also lead to a deeper understanding of our priorities. The second more difficult question is, whether our priorities can be connected to probabilities in a useful way. The relevance of this question stems from application areas in which autoepistemic logic can be used and where probability information is available. Such an application area is model-based diagnosis. Interesting questions concern the existence of prioritizations that correspond to (in some suitable sense) to given probability information. Similarly, given a prioritization, what kind of probabilities correspond to it? These questions are interesting because of their potential implications on efficient reasoning procedures for diagnostic reasoning.

Acknowledgements

I am very grateful to Prof. Leo Ojala for financial and other arrangements that made this work possible, and to Dr Ilkka Niemelä for helpful comments, discussions and advice, and especially for suggesting the subject of this work. The research was financed from funds from Suomen Akatemia. Additional financial support has been provided by the following foundations: Alfred Kordelinin yleinen edistys- ja sivistysrahasto, Heikki ja Hilma Honkasen säätiö.

Otaniemi, December 1993

References

Anderson, I.: *Combinatorics of Finite Sets.* Oxford Science Publications, Oxford, 1987.

Arrow, K. J.: *Social Choice and Individual Values.* Yale University Press, New Haven, CT, 2nd edition, 1963.

Arrow, K. J.: Values and collective decision making. In P. Laslett and W. G. Runciman, editors, *Philosophy, Politics and Society (third series)*, pages 215–232. Basil Blackwell, Oxford, 1967.

Baker, A. B. and M. L. Ginsberg: A theorem prover for prioritized circumscription. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 463–467, Detroit, 1989. Morgan Kaufmann Publishers.

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

Brewka, G.: Preferred subtheories: an extended logical framework for default reasoning. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1043–1048, Detroit, 1989. Morgan Kaufmann Publishers.

Brewka, G.: *Nonmonotonic Reasoning: Logical Foundations of Commonsense.* Cambridge University Press, Cambridge, UK, 1991.

de Condorcet: *Essai sur l'application de l'analyse à la probabilité des décisions redues à la pluralité des voix.* Paris, 1785.

Dowling, W. F. and J. H. Gallier: Linear-time algorithms for testing the satisfiability of propositional Horn formulae. *Journal of Logic Programming*, 1(3):267–284, 1984.

Doyle, J. and M. P. Wellman: Impediments to universal preference-based default theories. *Artificial Intelligence*, 49:97–128, 1991.

Fagin, R., J. D. Ullman, and M. Y. Vardi: On the semantics of updates in databases. In *Proceedings of the Second ACM SIGACT–SIGMOD Symposium on Principles of Database Systems*, pages 352–365, Atlanta, March 1983. The Association for Computing Machinery.

Gärdenfors, P.: *Knowledge in Flux.* The MIT Press, Cambridge, MA, 1988.

Garey, M. R. and D. S. Johnson: *Computers and Intractability.* W.H. Freeman and Company, San Francisco, 1979.

Geffner, H.: Conditional entailment: closing the gap between defaults and conditionals. Preprints of Third International Workshop on Nonmonotonic Reasoning, 1990.

Geffner, H.: *Default Reasoning: Causal and Conditional Theories.* MIT Press, Cambridge, MA, 1992.

Gelfond, M. and V. Lifschitz: The stable model semantics for logic programming. In *Proceedings of the 5th International Conference on Logic Programming*, pages 1070–1080, Seattle, August 1988. The MIT Press.

Gelfond, M. and H. Przymusinska: Formalization of inheritance reasoning in autoepistemic logic. *Fundamenta Informaticae*, 13(4):430–443, 1990.

Gelfond, M.: On stratified autoepistemic theories. In *Proceedings of the 6th National Conference on Artificial Intelligence*, pages 207–211, Seattle, July 1987. American Association for Artificial Intelligence.

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

Grosof, B. N.: Generalizing prioritization. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 289–300, Cambridge, MA, 1991. Morgan Kaufmann Publishers.

Junker, U. and G. Brewka: Handling partially ordered defaults in TMS. In *Symbolic and quantitative approaches to uncertainty. European Conference*, number 548 in Lecture Notes in Computer Science, pages 211–218, Marseille, 1991. Springer-Verlag.

Junker, U.: Prioritized defaults: Implementation by TMS and application to diagnosis. In J. Mylopoulos, editor, *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 310–315, Sydney, 1991. Morgan Kaufmann Publishers.

Kanellakis, P. C.: Elements of relational database theory. In J. Van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 1073–1156. Elsevier, 1990.

Konolige, K.: Hierarchic autoepistemic theories for nonmonotonic reasoning: Preliminary report. In *Proceedings of the 2nd Workshop on Non-Monotonic Reasoning*, number 346 in Lecture Notes in Artificial Intelligence, pages 42–59, Grassau, Germany, June 1989. Springer-Verlag.

Kramer, B. M. and J. Mylopoulos: Representation of knowledge. In S. C. Shapiro, editor, *Encyclopedia of Artificial Intelligence*, volume 2, pages 882–890. John Wiley & Sons, 1987.

Lee, S.-J. and D. A. Plaisted: Eliminating duplication with the hyper-linking strategy. *Journal of Automated Reasoning*, 9:25–42, 1992.

Lifschitz, V.: Computing circumscription. In A. Joshi, editor, *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 121–127, Los Angeles, August 1985. Morgan Kaufmann Publishers.

MacNish, C.: Hierarchical default logic. In *Symbolic and quantitative approaches to uncertainty. European Conference*, number 548 in Lecture Notes in Computer Science, pages 246–253, Marseille, 1991. Springer-Verlag.

Marek, W. and M. Truszczyński: Autoepistemic logic. *Journal of the ACM*, 38:588–619, 1991.

Marek, W., G. F. Shvarts, and M. Truszczyński: Modal nonmonotonic logics: Ranges, characterization, computation. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 395–404, Cambridge, MA, April 1991. Morgan Kaufmann Publishers.

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

McCarthy, J.: Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

Milner, R. and M. Tofte: *The Definition of Standard ML*. The MIT Press, Cambridge, MA, 1991.

Moore, R. C.: Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.

Nebel, B.: Belief revision and default reasoning: Syntax-based approaches. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 417–428, Cambridge, MA, 1991. Morgan Kaufmann Publishers.

Niemelä, I. and J. Rintanen: On the impact of stratification on the complexity of nonmonotonic reasoning. In B. Nebel, C. Rich, and W. Swartout, editors, *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, pages 627–638, Cambridge, MA, October 1992. Morgan Kaufmann Publishers.

Niemelä, I.: Towards automatic autoepistemic reasoning. In J. Van Eijck, editor, *Proceedings of the European Workshop on Logics in Artificial Intelligence—JELIA'90*, number 478 in Lecture Notes in Artificial Intelligence, pages 428–443, Amsterdam, September 1990. Springer-Verlag.

Niemelä, I.: Constructive tightly grounded autoepistemic reasoning. In J. Mylopoulos, editor, *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 399–404, Sydney, August 1991. Morgan Kaufmann Publishers.

Niemelä, I. N. F.: On the decidability and complexity of autoepistemic reasoning. *Fundamenta Informaticae*, 17:117–155, 1992.

Niemelä, I. N. F.: A unifying framework for nonmonotonic reasoning. In B. Neumann, editor, *Proceedings of the 10th European Conference on Artificial Intelligence*, pages 334–338, Vienna, 1992. John Wiley & Sons.

Niemelä, I.: *Autoepistemic Logic as a Unified Basis for Nonmonotonic Reasoning.* PhD thesis, Helsinki University of Technology, August 1993.

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

Przymusinska, H.: The embeddability of hierarchic autoepistemic logic in autoepistemic logic. In Z. W. Ras, editor, *Methodologies for Intelligent Systems, 4: Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*, pages 485–493, New York, 1989. North-Holland.

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

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

Ryan, M.: Defaults and revision in structured theories. In *Proceedings of the Sixth Annual IEEE Symposium on Logic in Computer Science*, pages 362–373, Amsterdam, 1991. IEEE Computer Society Press.

Ryan, M.: Representing defaults as sentences with reduced priority. In B. Nebel, C. Rich, and W. Swartout, editors, *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning,* pages 649–660, Cambridge, MA, October 1992. Morgan Kaufmann Publishers.

Smullyan, R.: *First-Order Logic*. Springer-Verlag, Berlin, 1968.

Stein, L. A.: Resolving ambiguity in nonmonotonic inheritance networks. *Artificial Intelligence*, 55(2–3):259–310, 1992.

Stillman, J.: It is not my default: the complexity of reasoning in default logic. In *Proceedings of the 9th National Conference on Artificial Intelligence*, pages 571–578, Boston, August 1990. Morgan Kaufmann Publishers.

Tan, Y. H. and J. Treur: Constructive default logic and the control of defeasible reasoning. In B. Neumann, editor, *Proceedings of the 10th European Conference on Artificial Intelligence*, pages 299–303, Vienna, 1992. John Wiley & Sons.

Toyama, K., Y. Inagaki, and T. Fukumura: Knowledge representation based on autoepistemic logic for multiple agents. Technical report 90-1-06, Chukyo University, School of Computer and Cognitive Sciences, March 1991.

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, DC, July 1991. The MIT Press.