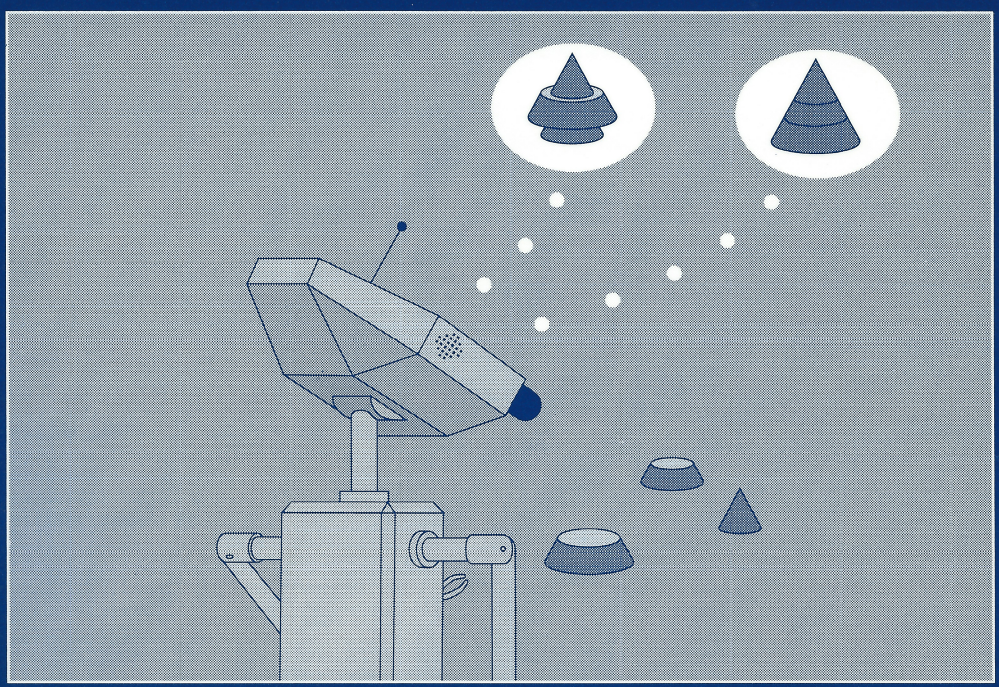
What is on the machine's mind?

Models for reasoning with incomplete and uncertain knowledge



Nico Roos

What is on the machine's mind?

Models for reasoning with incomplete and uncertain knowledge

What is on the machine's mind?

Models for reasoning with incomplete and uncertain knowledge

Proefschrift

Ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft
op gezag van de Rector Magnificus,
Prof. drs. P.A. Schenck,
in het openbaar te verdedigen
ten overstaan van een commissie
aangewezen door het College van Dekanen
op dinsdag 19 februari 1991 te 16.00 uur

door

Nicolaas Roos

informatica ingenieur
geboren te Rotterdam

Dit proefschrift is goedgekeurd door de promotor:

prof. dr. S.C. van Westrhenen

De overige leden van de promotiecommissie zijn:

prof. dr. J. M. Aarts
dr. R. Cooke
prof. S. J. Doorman M.Sc.
prof. dr. ir. J. van Katwijk
prof. dr. J.-J. Ch. Meyer
dr. C. Witteveen

Published and produced by
N.Roos
Rietveld 198
2611 LR Delft
The Netherlands

ISBN 90-90003079-1

Preface

During my study in computer science, I got interested in the the research on formal aspects of knowledge representation. This was the reason why I asked Prof. dr. S. C. van Westrhenen of the group of Theoretical Computer Science, whether it were possible to write my M.Sc. thesis about this subject. I was lucky since he was also interested in this subject and wanted to learn more about it. Because a friend of mine Hans Hellendoorn had asked the same question, a combined project was started in which Hans Hellendoorn investigated theories for handling uncertain and fuzzy knowledge and in which I investigated theories for handling incomplete knowledge. This project as well as our studies were successfully completed in January 1987.

At the end of our study Prof. van Westrhenen asked us to continue our research in a new project that should be completed after four years with the publication of a Ph.D. thesis. The leaders of this project, which was sponsored by the National Aerospace Laboratory (NLR), were dr. C. Witteveen from the group of Theoretical Computer Science and dr. R. J. P. Groothuizen from the NLR. The project was called the *ROOK* (Redeneren met Onzekere en Onvolledige Kennis) project which stands for Reasoning with Uncertain and Incomplete Knowledge. The third aspect of the project was the inexactness of the knowledge which is expressed by the Dutch word 'rook'. In English this word means 'smoke', which carries the meaning of inexactness.

Within this project there have been four topics on which we have been concentrating. Hans Hellendoorn who continued the line of research he started for his M.Sc.thesis, concentrating on reasoning with fuzzy logic. Cees Witteveen started research on Truth Maintenance Systems investigating time complexity and semantical foundations.

My own research concentrated on two topics, non-monotonic logics and certainty measures that express our ignorance about the world. The first one and a half year of the project, I have spent studying the literature on these topics and investigating whether it is possible to define a certainty measure that could be used to realize an efficient diagnostic reasoning process using heuristic knowledge. The results of this research looked promising but there remained some important problems to be solved.

After the discussions I had with Yao Hua Tan at the ECAI-88 in München, I

started to re-examine some ideas I had described in my M.Sc. thesis. These ideas are to view non-monotonic reasoning as a special case of reasoning with inconsistent knowledge and to solve conflicts only when they are being derived using a partial preference relation on the premisses. The first part of this thesis reports on the result of this research.

In July 1989, I completed the research reported on in the first part of my thesis and continued my research on defining a certainty measure that could be used to realize an efficient diagnostic reasoning process. In my first attempts I had used conditional probabilities to relate possible causes to anomalies that can be observed. Since these conditional probabilities depend on the a priori probabilities of the anomalies, e.g. one can have a headache not caused by some disease, incorrect results could be derived. To overcome this problem I wanted to use relations like: '*most* humans have brown eyes' and 'by *many* patients, a brain tumor causes headache'. The problem that arose was how to base correct conclusions on this information. It turned out that the solution was to view a reasoning process as a process of constructing a partial model of the world we are reasoning about. This view on a reasoning process made it possible to give a very natural definition of a *probability* and a *likelihood* measure. The latter can be used to realize an efficient diagnostic reasoning process.

Viewing a reasoning process as a process of constructing a partial model, is not only very useful for handling uncertainty, but also for handling default reasoning. Since in a partial model, the consistency problem is decidable, default conclusions are correct with respect to the partial model. Default conclusions only have to be withdrawn when new information is being added to the partial model that overrules the default conclusion. Hence, default reasoning based on the construction of a partial model does not possess the process non-monotonicity found in other non-monotonic logics.

Unfortunately, the results of my research on default reasoning using a partial model are not reported on in this thesis. They are left out because of two problems that had to be solved first. The first problem was how to order the partial models on the amount of information they possess. The second problem, which was pointed out to me by E. Sandewall at a lecture I gave at the university of Linköping, was caused by an incorrect handling of the uncertainty expressed by nested formulas. Since, I found the solution for both problems only a week before I had to finish this thesis, there was no time left to discuss default reasoning based on the construction of a partial model.

Acknowledgement

I thank everyone who has supported me over the last four year making it possible for me to write this Ph.D. thesis. Without their support, criticism, comments, discussion and friendship it would not have been possible to finish this thesis. I specially want to thank the following persons:

I thank Prof. S. C. van Westrhenen for offering me the opportunity to write this thesis and for arranging the necessary financial support.

I thank Cees Witteveen for leading the *ROOK* project, for the discussions we had about my work and for reading the many draft versions of my thesis.

I thank my colleague Hans Hellendoorn with whom I worked together for many years, for his friendship and for *never* agreeing with me.

I thank Trudie Stoute for her advice about English.

I thank the National Aerospace Laboratory for their financial support. Especially I thank Anneke Donker and Ronny Groothuizen from the NLR for their participation in the *ROOK* project.

Contents

Introduction

Knowledge based systems differ from conventional programs in the way they perform their tasks. A task performed by a conventional program is completely specified by its internal structure designed by the programmer. This implies that the programmer either has to foresee every possible application of the program or (s)he has to place constraints on the applicability of the program.

Knowledge based systems offer a greater flexibility than conventional programs. This flexibility is reached by representing the knowledge needed to perform a task explicitly. The knowledge based system tries to perform the task at hand by manipulating the available domain knowledge.

If the available knowledge is only applicable in specific situations, the only thing we have gained by using a knowledge based system, is an easier and more flexible way of specifying a task to be performed. Although this can be an important advantage, we like to have more. If it is possible to store more generally applicable, and because of this possibly less accurate knowledge, in a knowledge base system, we also can anticipate on unforeseen situations. A disadvantage of the use of this knowledge is that conclusions based on it can be uncertain and can sometimes be wrong. Hence, a knowledge based system may not always perform its task in a correct or optimum way. This, however, is often better than no results at all.

To be able to manipulate generally applicable knowledge correctly, some formal description of this knowledge is needed. This description should specify a method for representing the knowledge and should specify a meaning for this representation. Furthermore, rules for manipulating the knowledge representation are needed. As was argued by J. P. Hayes [25], especially the meaning of the knowledge representation is important. Having a semantics for a knowledge representation enables us to evaluate the situations in which the represented knowledge is *true*. Also the correctness of the rules for manipulating the knowledge can then be verified.

Given these constraints, it seems that logic is a good candidate of knowledge representation. Logics possess a language for representing knowledge, a semantics and a proof theory for manipulating the knowledge. Classical logics, however, are not able to deal with not fully accurate knowledge. Because of this, new logics have appeared. These logics can roughly be divided into two classes, viz. logics for representing uncertain knowledge and non-monotonic logics. The logics for representing

uncertain knowledge are based on extensions of classical logics. In these logics some certainty, belief or probability measure is associated with every proposition.

The non-monotonic logics are a different way of dealing with less accurate knowledge. These logics are used to describe what should hold if possible; i.e. they describe a preferred situation. This allows us to jump to conclusions if the available information is incomplete. If new information gets available, some conclusions may no longer hold. Because of this property, the logics are called non-monotonic.

In the first part of this thesis I will describe a new non-monotonic logic. This logic, which I will call the preference logic, combines the advantages of some existing approaches while avoiding some of their disadvantages. Furthermore, a deduction process in which reason maintenance is integrated, is defined for the logic.

Although non-monotonic logics are a good formalism for representing preferred situations, their reasoning processes, which must be used to determine what holds in a set of preferred situations, are not very intuitive. Like in a classical logic the reasoning processes can only approximate the set of correct conclusions. Here, however, this approximation may contain wrong conclusions. In fact, not only the logics are non-monotonic, but also the approximation of the set of correct conclusions.

To avoid this problem, I will propose a new way of looking at a reasoning process in the second part of this thesis. What I will propose is to view a reasoning process as a process of constructing a partial model of the world we are reasoning about. This way of reasoning has some important advantages over traditional reasoning processes. First of all, it is a more intuitive way of reasoning. Furthermore, the consistency problem is decidable for a partial model. Since all non-monotonic logics depend, directly or indirectly, on consistency checks, this property can be very important for these logics. Finally, it is possible to define two different certainty measures for conclusions derived which express our ignorance about the world with respect to a partial model. One certainty measure expresses the expectation that a conclusion holds, and the other certainty measure expresses whether a conclusion is an explanation for anomalies observed.

One should not expect a description of a complete framework in the second part of this thesis. What I will describe are the foundations of such a framework. Since many problems are still unsolved, much more research is needed. Despite of this, I hope to convince the reader that the approach proposed in this part can be an interesting alternative.

Part I

Non-monotonic logics

1

Non-monotonic reasoning

In literature four different approaches for non-monotonic reasoning can be distinguished, *non-monotonic logics*, *reason maintenance systems*, *inheritance networks* and *belief revision*. In the following section I will briefly describe the properties of these approaches. The reason to describe these properties is twofold. Firstly, it serves as a context in which the research, described in the following chapter, should be placed. Secondly, it summarises those approaches that will be used in the description of the research.

1.1 Non-monotonic logics

The non-monotonic logics all emerged as an extension of some classical logic. Since there is no uniform way in which the logics are extended, I will describe some of the non-monotonic logics, starting with the three most prominent ones that exist today. These are J. McCarthy's *Circumscription* [41], R. Reiter's *Default logic* [49] and R. C. Moore's *Autoepistemic logic* [44]. The first two were published in 1980 in a special issue of the *Artificial Intelligence* journal volume 13, and the *autoepistemic logic*, based on the article of D. McDermott and J. Doyle, *Non-monotonic logic I*

[43], appeared in the same special issue. So it seems that in the last ten years no interesting new approaches have been developed, despite of the fact that these three approaches are not perfect.

1.1.1 Circumscription

McCarthy's circumscription is based on the idea of minimizing the set of n-tuples for which an n-place predicate is *true*. In fact it can be viewed as a sound formalization of the Closed World Assumption. To illustrate the ideas behind circumscription, suppose that we have a set of premisses Σ in which a predicate p with n arguments occurs. Given some standard semantical structure $\langle D, v\rangle$ where D is a set of domain objects and v is a valuation function, we can define truth values for formulas. The truth value of the predicate p is defined in this structure as:

$$\langle D, v\rangle \models p(t_1, ..., t_n) \text{ if and only if } (v(t_1), ..., v(t_n)) \in v(p)$$

So $v(p)$ denotes the set of n-tuples over the domain D for which the predicate p is true. If we would apply circumscription of the predicate p, we should prefer those structures that are models for the premisses Σ, in which p is true for the smallest possible set of n-tuples $v(p) \subseteq D^n$. Now, consider the following example.

Example 1.1

1. $bird(Tweety)$
2. $ostrich(Tweety)$
3. $bird(Woody)$
4. $\forall x\,[bird(x) \wedge \neg abnormal(x) \rightarrow can\_fly(x)]$
5. $\forall x\,[ostrich(x) \rightarrow abnormal(x)]$
6. $\forall x\,[ostrich(x) \rightarrow \neg can\_fly(x)]$

Applying circumscription on the predicate *abnormal* in this example, implies that we prefer those models $\langle D, v\rangle$ in which $v(abnormal) = \{v(Tweety)\}$. Now if $Tweety$ and $Woody$ denote different objects in the domain D, we get the intuitive correct result, since:

$$\langle D, v\rangle \models abnormal(Tweety)$$

$$\langle D, v\rangle \models \neg abnormal(Woody)$$

$$\langle D, v\rangle \models \neg can\_fly(Tweety)$$

$$\langle D, v\rangle \models can\_fly(Woody)$$

After this more or less intuitive definition of circumscription, I will give the formal semantic definition.

Definition 1.2 For each pair of structures $\mathcal{M}, \mathcal{N}$ if $\mathcal{M} = \langle D, v \rangle$, $\mathcal{N} = \langle D, w \rangle$ and $v(p) \subset w(p)$, then $\mathcal{M} <_p \mathcal{N}$.

$\mathcal{M}$ is a minimal model in p for a set of premisses Σ if and only if $\mathcal{M} \in Mod(\Sigma)$ and for no $\mathcal{N} \in Mod(\Sigma)$ there holds:

$$\mathcal{N} <_p \mathcal{M}.$$

The models for circumscription of the predicate p given the premisses Σ, are called 'minimal models'. These are the models for Σ in which the predicate p is true for a minimal set of n-tuples $v(p)$.

We can also give a syntactical characterization of circumscription. For this a second order formula is used. This formula says that the predicate we are circumscribing should be the predicate that is true for the smallest number of n-tuples such that the premisses Σ are satisfied.

Definition 1.3 Let Σ be a set of premisses. Furthermore, let $p, q_1, ..., q_n$ be the predicates which occur in Σ and let $\Sigma = \Sigma(p, q_1, ..., q_n)$.
Then circumscription of p in Σ, denoted as $CIRC(\Sigma; p)$, is equal to:

$$CIRC(\Sigma; p) \equiv \{\forall p', q_1', ..., q_n'[[\bigwedge \Sigma(p', q_1', ..., q_n')] \rightarrow \neg[p' < p]]\} \cup \Sigma(p, q_1, ..., q_n)$$

where $p \leqslant p' \equiv \forall \overline{x}(p(\overline{x}) \rightarrow p'(\overline{x}))$ and $p < p' \equiv p \leqslant p' \wedge \neg[p' \leqslant p]$.

Notice that $p' \geqslant p$ in the structure $\langle D, v \rangle$ if and only if $v(p) \subseteq R$ where R is some relation which represents the predicate variable p'.

The form of circumscription defined in Definition 1.3 is called *second order formula circumscription*. In Definition 1.3 all predicates in the set of premisses may vary. In the more general form of formula circumscription, we can specify explicitly which predicates may vary. A special case of formula circumscription is when we do not allow any predicate to vary, except for the one we circumscribe [42]. This version of circumscription, called *predicate circumscription*, can be described by the following formula.

$$\forall p'[[\bigwedge \Sigma(p') \wedge \forall \overline{x}[p'(\overline{x}) \rightarrow p(\overline{x})]] \rightarrow \forall \overline{x}[p(\overline{x}) \rightarrow p'(\overline{x})]]$$

If we remove the quantifiers from this formula and represent the resulting formula as an axiom scheme, we get the circumscription introduced by McCarthy [41]. However, this original form of circumscription is not very useful, because predicates not being circumscribed, are not allowed to vary. Therefore, circumscribing the predicate *abnormal* in Example 1.1 will not enable us to conclude that Woody can fly, $can\_fly(Woody)$. Other forms of circumscription that have appeared, are e.g. parallel predicate circumscription, in which we minimize a set of predicates [37]; priorized circumscription [37], in which we minimize a set of predicates according to some preference ordering, and pointwise circumscription [38]. Pointwise circumscription is a

form of circumscription that differs from all other forms of circumscription. It is based on the view that the truth value of an n-place predicate is defined by a function from D^n to $\{true, false\}$. If the truth value is defined by such a function, there is no relation that can be minimized. In fact, we can only minimize a predicate for each point $\xi \in D^n$ by changing, if possible, its truth value from *true* to *false* at this point ξ. This form of minimization can be expressed by the following circumscription formula.

$$\forall x \neg[p(x) \wedge \bigwedge \Sigma(\lambda y\ (p(y) \wedge x \neq y))]$$

To be able to make deductions from $CIRC(\Sigma; p)$, we need to be able to collapse the second order formula into a first order formula. Neither is this always possible nor is there an algorithm for doing this. It is only for a number of special cases that it will be possible to specify how the second order formula can be collapsed into a first order formula. Another objection against circumscription is that although we can represent what should hold in a preferred situation in an intuitively clear way using circumscription, this representation does not seem to be a natural one.

Remark 1.4 When I showed above how circumscription can be applied in Example 1.1, I assumed that *Tweety* and *Woody* denote different objects in the domain. This assumption is called the Unique Name Assumption (UNA) (if ground terms cannot be proved equal, they can be assumed unequal). One might think that the UNA can be modelled using circumscription on the predicate '='. Unfortunately, this is not possible. The reason for this is based on the fact that the truth value of '$x = y$' ($\langle D, v\rangle \models x = y$) is not defined with a relation but as: $v(x) = v(y)$. So '$x = y$' is true if and only if both x and y denote the same object in the domain. Hence, we have to specify explicitly that names denote different objects when we are describing default reasoning using circumscription of the predicate *abnormal* [42].

1.1.2 Default logic

Reiter's default logic is based on extending the predicate logic with a set of special inference rules. These special inference rules are called default rules. A default rule is a rule which adds a new formula to the theory when certain conditions are satisfied. A default rule is a rule of the form:

$$\frac{\alpha(\overline{x})\ :\ \beta_1(\overline{x}), ..., \beta_n(\overline{x})}{\gamma(\overline{x})}$$

where α, $\beta_1, ..., \beta_n$ and γ are formulas containing the free variables $\overline{x}$. Here α is called the prerequisite, $\beta_1, ..., \beta_n$ are called the justifications and γ is called the consequent. A default rule can be viewed as a schemata from which we can generate ***closed default rules*** by substituting the free variables $\overline{x}$ by ground terms.

Roughly speaking, the default rules can be interpreted as: 'for each sequence of ground terms $\bar{t}$: if $\alpha(\bar{t})$ is a known belief and $\neg\beta_1(\bar{t}), ..., \neg\beta_n(\bar{t})$ are consistent with every derivable formula (informally: are not known beliefs) then $\gamma(\bar{t})$ should be believed'. Using this interpretation, we are able to describe the default rule 'birds can fly'.

$$\frac{bird(x) \ : \ can\_fly(x)}{can\_fly(x)}$$

A default theory $E(\Sigma, D)$ consists of a set premisses Σ and a set of default rules D. The following default theory is a reformulation of Example 1.1. The formula in Example 1.1 containing the predicate *abnormal* as a condition in the antecedent of an implication, is replaced by a default rule.

Example 1.5

1. $bird(Tweety)$
2. $ostrich(Tweety)$
3. $bird(Woody)$
4. $\forall x[ostrich(x) \rightarrow \neg can\_fly(x)]$
5. $\dfrac{bird(x) \ : \ can\_fly(x)}{can\_fly(x)}$

When we want to deduce the fact $can\_fly(Woody)$ from this default theory, we need to be able to determine if $\neg can\_fly(Woody)$ is one of our beliefs. This means that $can\_fly(Woody)$ is consistent with what we believe, or informally $\neg can\_fly(Woody)$ cannot be derived from what we believe. Since the problem whether a set of formulas is consistent, is, in general, undecidable, we cannot describe the theory of some default theory $E(\Sigma, D)$ using some executable deduction process. The theory of a default theory $E(\Sigma, D)$, which is called an extension, can only be characterized by a fixed point construction. The following fixed point construction characterizes an extension of a default theory.

Definition 1.6 E is an extension of $E(\Sigma, D)$ if and only if E is a fixed point of Γ, $E = \Gamma(E)$, where for each set of sentences S, $\Gamma(S)$ is defined as the smallest set satisfying:

- $\Sigma \subseteq \Gamma(S)$
- $Th(\Gamma(S)) = \Gamma(S)$
- for each sequence of ground terms $\bar{t}$: if $\dfrac{\alpha(\bar{x}) \ : \ \beta_1(\bar{x}), ..., \beta_n(\bar{x})}{\gamma(\bar{x})} \in D$, $\alpha(\bar{t}) \in \Gamma(S)$ and $\neg\beta_1(\bar{t}), ..., \neg\beta_n(\bar{t}) \notin S$, then $\gamma(\bar{t}) \in \Gamma(S)$.

An extension of a default theory, which satisfies the definition above, does not have to be unique. There may be multiple extensions for the same default theory. An example of such a theory is the following one:

Example 1.7 Let (Σ, D) be a default theory with $\Sigma = \varnothing$ and $D = \{\frac{\top : p}{\neg q}, \frac{\top : q}{\neg p}\}$ This default theory has two extensions, namely $Th(\{\neg q\})$ and $Th(\{\neg p\})$.

It is also possible that a default theory has no extensions at all, as illustrated in the following example.

Example 1.8 Let (Σ, D) be a default theory with $\Sigma = \{p\}$ and $D = \{\frac{\top : q}{\neg p}\}$. This default theory has no extensions.

Reiter introduces two special classes of default rules, which always have an extension, namely normal and semi-normal default rules. Normal default rules are default rules of the form:

$$\frac{\alpha(\overline{x}) : \gamma(\overline{x})}{\gamma(\overline{x})}$$

Since a ground instance of γ can only be an element of an extension if the denial of this instance is not an element of the extension, a default theory only containing normal default rules will have an extension. A semi-normal default looks much like a normal default, but it has more expressive power.

$$\frac{\alpha(\overline{x}) : \gamma(\overline{x}), \beta_1(\overline{x}), ..., \beta_n(\overline{x})}{\gamma(\overline{x})}$$

For the same reason as for a normal default theory, a default theory containing semi-normal defaults will have an extension.

In his original paper Reiter assumed that defaults used by humans could all be modelled by only using normal default rules. Later on he withdraws this assumption [50]. Then he showes that between default rules and between default rules and implications unwanted transitive relations may occur. Consider, for example, the following situation.

- Every Quebecois is a Canadian.
- Canadians are normally native English speakers.
- Quebecois are normally not native English speakers.

Translated into the default logic, this becomes:

1. $\forall x[quebecois(x) \rightarrow canadian(x)]$
2. $\frac{canadian(x) : native\_english\_speaker(x)}{native\_english\_speaker(x)}$

3. $$\frac{quebecois(x) : \neg native\_english\_speaker(x)}{\neg native\_english\_speaker(x)}$$

This default theory has two extensions when we add $quebecois(Pierre)$ to the set of premisses. These extensions are:

- $Th(\{native\_english\_speaker(Pierre)\} \cup \Sigma)$
- $Th(\{\neg native\_english\_speaker(Pierre)\} \cup \Sigma)$

It is clear that only the latter extension is correct. To avoid the former extension in this example, we have to replace the first default rule by:

$$\frac{quebecois(x) : \neg native\_english\_speaker(x) \wedge \neg quebecois(x)}{\neg native\_english\_speaker(x)}$$

What we have done to eliminate the unwanted extensions is formulating an exception on the application of a default rule. So, in describing a default using a default rule, we have to consider every possible exception on its application.

A strange property of default logic is that the default rules are a set of (odd) inference rules that influence the semantics of the premisses while, normally, inference rules belong to the proof theory. As part of the proof theory they should not influence the semantics of the logic.

1.1.3 Autoepistemic logic

Autoepistemic logic is a logic for modelling the knowledge of an ideal introspective agent. Such an agent knows all the logical consequences of its beliefs; for every proposition φ it believes, it believes that it has a belief φ, and for every proposition ψ it does not believe, it believes it has not a belief ψ.

Let a proposition φ about which the agent believes that it believes φ, be described by $B(\varphi)$. Furthermore, let T denote the belief set of an ideal introspective agent. Then the conditions the agent should satisfy can be formulated as follows.

- $T = Th(T)$.
- If $\varphi \in T$, then $B(\varphi) \in T$.
- If $\varphi \notin T$, then $\neg B(\varphi) \in T$.

These conditions on the belief set of an ideal introspective agent were suggested by Stalnaker. He characterized a theory T that satisfies these conditions as *stable*, since no new beliefs can be derived from it.

To show what we can derive by modelling an ideal introspective agent, consider the following line of reasoning. 'I firmly believe that Nixon is still alive. Since if he had died, I would know that.' The belief: 'if Nixon had died, I would have known about that', can be represented by: $\varphi \rightarrow B(\varphi)$. This is equivalent to: $\neg B(\varphi) \rightarrow \neg\varphi$.

Since the belief set does not contain a belief that Nixon has died, $\varphi \notin T$, the belief set will contain $\neg B(\varphi) \in T$. Therefore, $\neg\varphi$ will be believed. Notice that the autoepistemic logic is a non-monotonic logic. Since, if the agent believes φ, $\varphi \in T$ and $\neg\varphi$ cannot be derived any more.

Autoepistimic logic can also be used to describe default rules. For example, the default 'Birds can fly' can be modelled as:

$$\forall x[bird(x) \wedge \neg B\neg can\_fly(x) \rightarrow can\_fly(x)]$$

For a belief set T of an autoepistemic logic, a formal semantics can be defined. In [44], Moore describes a semantics in which he refers to the belief set to define the truth value of $B(\varphi)$. Since, in my opinion, such circular references should not occur in a formal semantics, I will describe the alternative semantics that can be found in [45].

Definition 1.9 Let Ω be the set of possible worlds. $\langle W, R\rangle$ is a complete S5 Kripke structure if and only if $W \subseteq \Omega$ and $R = W \times W$.

$\langle W, R\rangle \models \varphi$ if and only if for every $w \in W$: $w \models \varphi$.

- $w \models p$ if and only if $w(p) = t$.
- $w \models \neg\varphi$ if and only if $w \not\models \varphi$.
- $w \models \varphi \rightarrow \psi$ if and only if $w \not\models \varphi$ or $w \models \psi$.
- $w \models B(\varphi)$ if and only if for every $w' \in W$: $w' \models \varphi$.

Given this definition, the following relation between stable belief sets and complete Kripke structures can be established.

Theorem 1.10 Let T be a set of beliefs.

T is a stable belief set if and only if there exists a complete Kripke structure $\langle W, R\rangle$ such that $\langle W, R\rangle \models T$.

Up till now nothing has been said about what an agent may believe. Here, it is assumed that the belief set is grounded in some base belief set, the premisses; i.e. the base belief set should be a subset of the stable belief set.

Definition 1.11 A belief set T is a stable belief set grounded in a set of premisses Σ, T is an autoepistemic (AE) extension of Σ, if and only if:

$$T = Th(\Sigma \cup \{B(\varphi) \mid \varphi \in T\} \cup \{\neg B(\varphi) \mid \varphi \notin T\})$$

In [32], a belief set that satisfies this definition is called a *weakly grounded* belief set by Konolige. He also gives a different way of characterizing an AE extension of Σ. To be able to do this, he introduces a new entailment relation $\models_{SS}$ by restricting the range of the modal indices on the entailment relation to stable belief sets only. Furthermore, let the set of ordinary formulas in T be denoted by T_0, the ordinary formulas not in T by $\overline{T}_0$, the set $\{B(\varphi) \mid \varphi \in T\}$ by BT and the set $\{\neg B(\varphi) \mid \varphi \notin T\}$ by $\neg B\overline{T}$. Using these new notations, Definition 1.11 can be reformulated as follows.

Definition 1.12 T is a weakly grounded AE extension for Σ if and only if:

$$T = \{\varphi \mid \Sigma \cup BT_0 \cup \neg B\overline{T}_0 \models_{SS} \varphi\}$$

In the weakly grounded AE extensions it is possible to possess beliefs that are not justified by the premisses. For example, let φ be an atomic proposition that does not occur in the premisses. Then, if the premisses possess an AE extension, it will possess an AE extension that contains φ, but it will also possess an AE extension that contains $\neg\varphi$. There is, however, no reason to believe φ in the former extension or to believe $\neg\varphi$ in the latter extension. To get rid of these extensions, Konolige introduces *moderately grounded* extensions.

Definition 1.13 T is a moderately grounded AE extension for Σ if and only if:

$$T = \{\varphi \mid \Sigma \cup B\Sigma \cup \neg B\overline{T}_0 \models_{SS} \varphi\}$$

Moderately grounded AE extensions do not contain more ordinary beliefs than strictly necessary. They are minimal according to the following definition.

Definition 1.14 An AE extension T of Σ is minimal for Σ if and only if: there exists no AE extension S of Σ such that $S_0 \subset T_0$.

Semantically, this definition implies that the model for an AE extension T, which is moderately grounded in Σ, is the most ignorant complete Kripke structure that satisfies Σ; i.e. there is no complete Kripke structure that satisfies Σ and contains more worlds.

Theorem 1.15 T is a moderately grounded AE extension for Σ if and only if there exists a complete Kripke structure $\langle W, R\rangle$ such that:

$$T = Th(\langle W, R\rangle),$$

$$\langle W, R\rangle \models \Sigma$$

and for no $\langle W', R'\rangle$ with $\langle W', R'\rangle \models \Sigma$ there holds:

$$W \subset W'.$$

Finally, Konolige defines an AE extension strongly grounded in Σ. He introduces this version of groundedness to avoid an agent from deriving a proposition φ from $B(\varphi)$ instead of the other way around. So, first an agent must believe a proposition before it may believe that it believes the proposition. Unfortunately, the definition of strongly grounded AE extensions depends on the syntactic properties of the premisses. Before a strongly grounded extension can be determined, the premisses have to be transformed in the following normal form.

$$\neg B(\alpha) \vee B(\beta_1) \vee ... \vee B(\beta_n) \vee \gamma$$

Here, $\alpha, \beta_1, ..., \beta_n$ and γ are ordinary formulas containing no modal operators.

Definition 1.16 Let Σ be a set of premisses in normal form and let T be an AE extension of Σ. Let Σ' be the sentences of Σ whose ordinary part is contained in T.

T is strongly grounded in Σ if and only if:

$$T = \{\varphi \mid \Sigma' \cup B\Sigma' \cup \neg B\overline{T}_0 \models_{SS} \varphi\}$$

It is possible to translate Reiter's default rules to sentences of the autoepistemic logic and vice versa, such that the kernels of the strongly grounded AE extensions are equivalent to the default extensions. The kernel of a stable set T is its set of ordinary formulas T_0.

1.1.4 Reasoning with inconsistent knowledge

The three non-monotonic logics discussed above are only intended as a context for the work described in the next chapter. In this subsection I will discuss three other logics that are closely related with the work described in the next chapter. The first logic I will discuss here is N. Rescher's approach to deal with an inconsistent set of premisses [51].

Hypothetical reasoning

In his book 'Hypothetical Reasoning' Rescher describes how to reason by using an inconsistent set of premisses. He introduces his reasoning method, because he wants to formalize hypothetical reasoning. In particular, he wants to formalize reasoning with belief contravening hypotheses, such as counterfactuals. In case of counterfactual reasoning, we make an assumption of which we know that in fact it is false. For example, let us suppose that Plato had lived during the middle ages. To be able to make such a counter factual assumption, we, temporally, have to give up some of our beliefs to restore consistency. It is, however, not always clear which of our beliefs we have to give up. The following example gives an illustration.

Example 1.17

Beliefs

1. Bizet was of French nationality.
2. Verdi was of Italian nationality.
3. Compatriots are persons who share the same nationality.

Hypothesis Assume that Bizet and Verdi are compatriots.

There are three possibilities to restore consistency. Clearly, we do not like to give 3, but we are indifferent whether we should give up 1 or 2.

To model this behaviour in a logical system, Rescher divides the set of premisses into modal categories. The modalities Rescher proposes are: alethic modalities, epistemic modalities, modalities based on inductive warrant, and modalities based on probability or confirmation. Given a set of modal categories, he selects Preferred Maximal Mutually-Compatible subsets (PMMC subsets) from them. The procedure for selecting these subsets is as follows:

Let $M_0, ..., M_n$ be a family of modal categories.

1. Select a maximal consistent subset of M_0 and let this be the set S_0.
2. Form S_i by adding as many premisses of M_i to S_{i-1} as possible without disturbing the consistency of S_i.

S_n is a PMMC-subset.

Given these PMMC-subsets, Rescher defines two entailment relations.

- Compatible-Subset (CS) entailment. A formula is CS entailed if it follows from every PMMC-subset.
- Compatible-Restricted (CR) entailment. A formula is CR entailed if it follows from some PMMC-subset.

Unfortunately, Rescher does not define a formal semantics for his logic.

Two approaches based on the ideas of Rescher are D. Poole's *framework for default reasoning* [48] and G. Brewka's *preferred subtheories* [6]. Poole, however, does not recognise this fact.

A logical framework for default reasoning

The central idea behind Poole's approach is that default reasoning should be viewed as *scientific theory formation*. Given a set of facts about the world and a set of hypotheses, a subset of the hypotheses which together with the facts can explain an *observation*, have to be selected. Of course, this selected set of hypotheses has to be consistent with the facts. A default rule is represented in Poole's framework by a hypothesis containing free variables. Such a hypothesis represents a set of ground instances of the hypothesis. Each of these ground instances can be used independent of the other instances in an explanation.

Definition 1.18 Let $\mathcal{F}$ be a set of facts and let Δ be a set of possible hypotheses. A *scenario* of $(\mathcal{F}, \Delta)$ is a set $D \cup \mathcal{F}$ where D is in the set of ground instances of Δ such that $D \cup \mathcal{F}$ is consistent.

Definition 1.19 If g is a closed formula, then an *explanation* of g from $\mathcal{F}, \Delta$ is a scenario of $(\mathcal{F}, \Delta)$ that implies g.

Definition 1.20 An *extension* of $(\mathcal{F}, \Delta)$ is a set of logical consequences of a maximal (with respect to the set inclusion) scenario of $(\mathcal{F}, \Delta)$.

Notice that this framework corresponds with Rescher's approach in which we have two modal categories, $M_0 = \mathcal{F}$ and $M_1 = \Delta$.

Given a set of facts and a set of hypotheses, it may occur that we do not want to accept some scenario as an explanation for a formula g. To block these unwanted scenarios, Poole extends his framework with a set of constraints. A scenario must be consistent with these constraints. In this way constraints can be used as a filter on the set of possible scenarios.

Definition 1.21 Let $\mathcal{F}$ be a set of facts, let Δ be a set of possible hypotheses and let C be a set of constraints. A *scenario* of $(\mathcal{F}, C, \Delta)$ is a set $D \cup \mathcal{F}$ where D is in the set of ground instances of Δ such that $D \cup \mathcal{F} \cup C$ is consistent.

Poole does not specify a semantics for his logic. He does not need a semantics, because he views default reasoning as scientific theory formation; i.e. he is only searching for (instances of) hypotheses that can be added to the facts, in order to explain the observations. If, however, we also want to use the framework for making predictions, then, clearly, a semantics is needed.

Preferred subtheories

G. Brewka generalizes Poole's framework for default reasoning. He introduces a partial preference relation on the hypotheses. Furthermore, he defines facts as the most preferred hypotheses. Following Rescher, Brewka selects preferred maximal consistent subsets of the set of hypotheses and calls them preferred subtheories.

With this generalization Brewka can solve the following two drawbacks of Poole's model.

- It is not possible to represent exceptions of exceptions in an elegant way. Due to this, the number of defaults needed to represent a situation, may become quite large.
- It is not possible to represent priorities between defaults directly.

In Brewka's model constraints do not occur any more. Instead, the preferred subtheories are generated, using the preference relation on the hypotheses. Although default reasoning can be described more elegantly with the preferred subtheories of Brewka, the subtheories contain less expressive power than the original logical framework for default reasoning of Poole. To illustrate this, Example 3.1 can be used.

Since Brewka presents his model as a generalization of Poole's model, there is, actually, no need for a new formal semantics. Poole argues that default reasoning should be viewed like scientific theory formation. The hypotheses, used to describe

defaults, only serve as a tentative theory used to explain observations. With this view on default reasoning, one can stay in the semantic domain of first order logic. This approach seems to be perfect for modelling defaults like: *birds lay eggs.* Observing that a bird laying eggs, this default can be used as an explanation. Poole's approach is not intended to deduce that some bird Tweety lays eggs. Since Tweety can be a male bird, the derivation of such a conclusion is undesirable. There are, however, defaults from which we do wish to derive default conclusions. Examples of such defaults are the following normative rules.

- Someone who has a driving licence, is permitted to drive a car.
- Someone who has a driving licence, but has drunk too much, is forbidden to drive a car.

Brewka's preferred subtheories can be defined using the following definitions. Let Σ be a set of hypotheses and let $(\Sigma, \prec)$ be a strict partial order on these hypotheses. Since hypotheses containing free variables (the defaults) denote a set of ground instances of these hypotheses, an expanded set of hypotheses $\overline{\Sigma}$ is introduced.

Definition 1.22 Let Σ be a set of hypotheses. The expanded set of hypotheses $\overline{\Sigma}$ is the smallest set in which every formula of Σ (containing free variables) is replaced by the set of ground instances of this formula.

Here it is assumed that a ground instance of a hypothesis has the same preferences as the original hypothesis of which it is an instance.

Definition 1.23 Let $(\Sigma, \prec)$ be the preference relation on a set of hypotheses Σ. The expanded preference relation $(\overline{\Sigma}, \prec)$ on the expanded set of hypotheses $\overline{\Sigma}$, is the smallest strict partial order containing $(\Sigma, \prec)$, which is invariant under the expansion of Σ to $\overline{\Sigma}$.

Now that the set of hypotheses and the preference relations are defined, the definition of a preferred subtheory can be given.

Definition 1.24 Let Σ be a set of hypotheses and $(\Sigma, \prec)$ be a preference relation defined on these hypotheses. Furthermore, let $\sigma_1, \sigma_2, \ldots$ be some enumeration[1] of $\overline{\Sigma}$ such that for every $\sigma_j \prec \sigma_k \in (\overline{\Sigma}, \prec)$: $k < j$.

S is a preferred subtheory of Σ if and only if $S = S_m$ where:

$$S_0 = \varnothing$$

and for $0 \leqslant i < m$

$$S_{i+1} = \begin{cases} S_i \cup \{\sigma_i\} & \text{if } S_i \cup \{\sigma_i\} \text{ is consistent} \\ S_i & \text{otherwise} \end{cases}$$

[1] Assuming a finite language a transfinite enumeration is needed when functions are allowed in the language. In that case the index i in the set $\sigma_1, \ldots, \sigma_i$ is understood to be some ordinal number.

Note that the most preferred hypothesis is always added. Because $(\overline{\Sigma}, \prec)$ is a partial order on $\overline{\Sigma}$, different enumerations of $\overline{\Sigma}$ may exist. Therefore, there may exist more than one preferred subtheory.

Given the preferred subtheories, Brewka defines two notions of provability: weak provability and strong provability.

- A formula φ is *weakly provable* from Σ if and only if there exists a preferred subtheory S of Σ such that $S \vdash \varphi$.
- A formula φ is *strongly provable* from Σ if and only if for every preferred subtheory S of Σ there holds $S \vdash \varphi$.

Notice that weakly provable corresponds with Rescher's CR entailment and that strongly provable corresponds with Rescher's CS entailment.

In Chapter 4 I will define a semantics for the preferred subtheories. The set of formulas entailed by the set of models is equal to the set of strongly provable formulas. This set will be denoted by Δ.

Definition 1.25 Let $S^1, ..., S^n$ be preferred subtheories that satisfy Definition 1.24. Then:

$$\Delta = \bigcap_{i \geq 1} Th(S^i)$$

1.1.5 Deriving new defaults

The last approach I want to mention here is Degrande's *conditional logic for prototypical properties* [15]. This logic, which is, in its original version, not a non-monotonic logic, possesses a property that cannot be found in any non-monotonic logic. In this logic one can reason about default rules. New default rules can be derived from old default rules and other propositions. Delgrande defines a semantic and a proof theory for his logic. For the proof theory soundness and completeness is proven. In [16], Delgrande extends his logic to enable the deduction of default conclusions. Unfortunately, the proof theory uses consistency checks. Therefore, no executable deduction process can be constructed.

1.1.6 Semantics

Till 1987 the semantics of non-monotonic logics was not well developed. The only logic having a clear and well defined semantic from the start, was circumscription. This is probably the reason why researchers are still interested in circumscription, despite of the fact that it is not useful for practical applications. Default logic and Rescher's approach to deal with inconsistencies did, originally, not have a semantics at all. As far as I know, the semantic described in Chapter 4 is the first semantics that is defined for Rescher's approach.

More clarity about the semantics of non-monotonic logics arose with Y. Shoham's paper *Non-monotonic logic: meaning and utility* [59]. In this paper Shoham argues that the difference between monotonic logic and non-monotonic logic is based on an alternative definition of logical entailment. In the classical monotonic logics a sentence is entailed by a set of premisses if and only if this sentence is true in every model for the set of premisses. In non-monotonic logic we change this definition. In these logics we define a strict partial preference relation on the set of semantical structures for a logic. Using this preference relation, we define a set of preferred models for a set of premisses. This means that we select a subset of the set of models for a set of premisses. Now, a sentence is entailed in a non-monotonic logic, it is preferentially entailed, if and only if this sentence is true in every preferred model of a set of premisses. The different forms of non-monotonic logics can now be realized, using different preference relations.

Definition 1.26 Let Str be a set of semantical structures and let $\langle Str, \sqsubset \rangle$ be a preference relation on the set of semantical structures. This relation must be transitive, but not reflexive or symmetric.

Using the preference relation $\sqsubset$, we can define a set of preferred models for a set of premisses.

Definition 1.27 Given a set of premisses Σ. Let $Mod(\Sigma)$ be the set of models for Σ. $Mod_{\sqsubset}(\Sigma)$ is the set of preferred models for Σ if and only if:

1. $Mod_{\sqsubset}(\Sigma) \subseteq Mod(\Sigma)$
2. for each $\mathcal{M}$:

   if $\mathcal{M} \in Mod(\Sigma)$ and for each $\mathcal{N} \in Mod(\Sigma)$: $\neg[\mathcal{M} \sqsubset \mathcal{N}]$, then $\mathcal{M} \in Mod_{\sqsubset}(\Sigma)$.

Using the set of preferred models, we can define the notion of preferred entailment.

Definition 1.28 A set of premisses preferentially entails a sentence, $\Sigma \models_{\sqsubset} p$, if and only if every preferred model for a set of premisses satisfies this sentence, i.e. for each $\mathcal{M} \in Mod_{\sqsubset}(\Sigma)$:

$$\mathcal{M} \models p.$$

It is not difficult to see that the standard monotonic logics are a special case in the framework described above. We get a standard monotonic logic if for all sets of premisses the set of preferred models is equal to the set of models. Furthermore, notice that it is possible to have a consistent set of premisses which preferentially entails a sentence and the denial of that sentence. For this it is only necessary that the set of preferred models of a set of premisses is empty.

An important theorem which does not hold in non-monotonic logics is the deduction theorem. A weaker version of this theorem still holds.

Theorem 1.29 Let Δ be a set of premisses and let p and q be two sentences.

If $\Delta \cup \{p\} \models_{\sqsubset} q$, then $\Delta \models_{\sqsubset} p \rightarrow q$.

Not all forms of non-monotonic reasoning can be modelled in the framework of Shoham. An important form of non-monotonic reasoning that can be modelled in this framework is circumscription. The following preference relation ensures that the preferred models are the same as the minimal models of Definition 1.2.

Definition 1.30 Let p be some atomic predicate.

For each pair of structures $\mathcal{M}, \mathcal{N}$ if $\mathcal{M} = \langle D, v \rangle$ and $\mathcal{N} = \langle D, w \rangle$, then:

$\mathcal{N} \sqsubset_p \mathcal{M}$ if and only if $v(p) \subset w(p)$.

Another example for which this frame work can be used, is the modelling of the Unique Name Assumption.

Definition 1.31 The result of applying the Unique Name Assumption (UNA) on a theory can be characterized by the following preference relation on the semantical structures.

For every structure $\mathcal{M}, \mathcal{N}$:

$\mathcal{N} \sqsubset_{UNA} \mathcal{M}$ if and only if $NEQ(\mathcal{N}) \subset NEQ(\mathcal{M})$,

where $NEQ(\mathcal{M}) = \{t_1 \neq t_2 \mid t_1, t_2 \in Terms \text{ en } \mathcal{M} \models t_1 \neq t_2\}$.

S. Kraus, D. Lehmann and M. Magidor generalized Shoham's view on the semantics for non-monotonic logics [33]. In their paper 'Non- monotonic reason, preferential models and cumulative logics', they study non-monotonic logics along two different lines. They study non-monotonic logics proof-theoretically by investigating the non-monotonic consequence relation $\mid\!\sim$, and they study it semantically by developing semantics based on Shoham's ideas. They also establish connections between the two lines.

The interpretation they give for the non-monotonic consequence relation $\alpha \mid\!\sim \beta$ is: 'if α, then normally β' or 'β is a plausible consequence of α'. Given this consequence relation, they identify a number of properties a non-monotonic logic should satisfy. According to Kraus et al. any logical system should at least satisfy the following properties, since they cannot think of any interesting weaker system.

- $\alpha \mid\!\sim \alpha$ (Reflexivity).
- $\dfrac{\models \alpha \leftrightarrow \beta, \alpha \mid\!\sim \gamma}{\beta \mid\!\sim \gamma}$ (Left Logical Equivalence).
- $\dfrac{\models \alpha \rightarrow \beta, \gamma \mid\!\sim \alpha}{\gamma \mid\!\sim \beta}$ (Right Weakening).

- $\dfrac{\alpha \wedge \beta \mathrel{|\!\sim} \gamma, \alpha \mathrel{|\!\sim} \beta}{\alpha \mathrel{|\!\sim} \gamma}$ (Cut).
- $\dfrac{\alpha \mathrel{|\!\sim} \beta, \alpha \mathrel{|\!\sim} \gamma}{\alpha \wedge \beta \mathrel{|\!\sim} \gamma}$ (Cautious Monotonicity).

A system that satisfies these properties is called system **C**, for *cumulative*. Some properties that can be derived in system **C** are:

- $\dfrac{\alpha \mathrel{|\!\sim} \beta, \beta \mathrel{|\!\sim} \alpha, \alpha \mathrel{|\!\sim} \gamma}{\beta \mathrel{|\!\sim} \gamma}$ (Equivalence).
- $\dfrac{\alpha \mathrel{|\!\sim} \beta, \alpha \mathrel{|\!\sim} \gamma}{\alpha \mathrel{|\!\sim} \beta \wedge \gamma}$ (And).
- $\dfrac{\alpha \mathrel{|\!\sim} \beta \rightarrow \gamma, \alpha \mathrel{|\!\sim} \beta}{\alpha \mathrel{|\!\sim} \gamma}$ (Modus Ponens in the Consequent).

System **C** can be made stronger by adding new properties on the consequence relation. The first property Kraus et al. add to system **C**, is the following one:

- $\dfrac{\alpha_0 \mathrel{|\!\sim} \alpha_1, \alpha_1 \mathrel{|\!\sim} \alpha_2, \ldots, \alpha_{k-1} \mathrel{|\!\sim} \alpha_k, \alpha_k \mathrel{|\!\sim} \alpha_0}{\alpha_0 \mathrel{|\!\sim} \alpha_k}$ (Loop).

A system that also satisfies this property, is called system **CL**, for *cumulative with loop*. According to Kraus et al. there is another property a non-monotonic logic should possess.

- $\dfrac{\alpha \mathrel{|\!\sim} \gamma, \beta \mathrel{|\!\sim} \gamma}{\alpha \vee \beta \mathrel{|\!\sim} \gamma}$ (Or).

A system that also satisfies this property, is called system **P**, for *preferential*. The addition of the rule Or to **CL** (or to **C** since **C** + Or implies Loop) is not beyond criticism. Kraus et al. argue that in counter examples for the rule Or, like the one described below, there is a hidden epistemic operator. If we make this operator explicit, the problem will disappear.

> For example, if I knew α, it would be an abnormal situation and if I knew $\neg\alpha$, it would be an abnormal situation as well.

There are, however, situations in which no hidden epistemic operator is involved, but in which the rule Or still leads to unwanted conclusions. This can be shown by extending the example Kraus et al. gave in defence of the rule Or.

Example 1.32 If John attends the party, normally the evening will be great, and if Cathy attends the party, normally, the evening will be great. But if both attend the party, the evening will not be great.

From this example it is clear that we may not conclude that if at least one of them attends the party, the evening will be great. So we can conclude that, although there are situations in which we would like to have the rule Or, we can also imagine situations in which it leads to undesirable conclusions

Kraus et al. distinguish two other systems **CM** and **M**, which stand for *cumulative monotonic* and *monotonic*. Since both systems are monotonic, they will not be considered here any further.

Now I will turn to the semantic side of the systems **C**, **CL** and **P**. As I mentioned before, Kraus et al. developed their semantics on the ideas of Shoham [58, 59]. They define a model as a set of state, representing possible states of affairs, and a binary preference relation on the states. This preference relation is used, for example, to prefer states in which Tweety can fly to states in which it cannot. The non-monotonic consequence relation $\alpha \mid\!\sim \beta$ is satisfied by such a model if and only if β is satisfied in the most preferred states that satisfy α.

The states in a model are labelled with a non empty set of worlds. Because a label of a state is not limited to a single world, these models possess more expressive power than Shoham's account. Even when states are labelled with single worlds, they still possess some extra expressive power since different states can be labelled with the same world. The following definitions describe their models.

Definition 1.33 A model is a triple $\langle S, l, < \rangle$ where S is a set of states, $l : S \rightarrow (2^{\mathcal{U}} - \{\varnothing\})$ is a labelling function that assigns a non empty set of worlds from a *universe* $\mathcal{U}$ of worlds to every state, and $<$ is a binary preference relation on the state S. Furthermore, the smoothness condition of Definition 1.35 must be satisfied.

Definition 1.34 Let $\langle S, l, < \rangle$ be a model. A formula α is satisfied by a state $s \in S$, $s \models\!\!\models \alpha$, if and only if for every $w \in l(s)$: $w \models \alpha$. Furthermore, let $\widehat{\alpha} = \{s \in S \mid s \models\!\!\models \alpha\}$ be the set of all states that satisfy α.

Definition 1.35 A set of states $\widehat{\alpha}$ is smooth if and only if for every $s \in \widehat{\alpha}$ there exists a minimal $t \in \widehat{\alpha}$ such that $t \leqslant s$.

A model $\langle S, l, < \rangle$ satisfies the smoothness condition if and only if for every formula α their holds: $\widehat{\alpha}$ is smooth.

Given the definitions of a model, we can formally define the meaning of the non-monotonic consequence relation.

Definition 1.36 Let $W = \langle S, l, < \rangle$ be a model. $\alpha \mid\!\sim_W \beta$ if and only if for any minimal state $s \in \widehat{\alpha}$, $s \models\!\!\models \beta$.

A model $\langle S, l, < \rangle$ without any additional conditions is called a *cumulative model*. For the cumulative models and system **C**, we can prove a representation theorem, relating the two characterizations of non-monotonic logics.

Theorem 1.37 A consequence relation is a consequence relation for system **C** if and only if it is defined by some cumulative model.

We can limit the set of cumulative models to those models whose preference relation is a strict partial order. These models are called *cumulative ordered models.* For the cumulative ordered models and system **CL**, we can also prove a representation theorem.

Theorem 1.38 A consequence relation is a consequence relation for system **CL** if and only if it is defined by some cumulative ordered model.

We can limit the set of cumulative ordered models to models whose labelling function assigns only single worlds to a state. The models are called *preferential models.* They correspond with Shoham's semantics for non-monotonic logics. Also for the preferential models and system **P**, we can prove a representation theorem.

Theorem 1.39 A consequence relation is a consequence relation of system **P**, if and only if it is defined by a preferential model.

1.2 Reason maintenance systems

Reason maintenance systems have emerged as an implementation technique for practical reasoning systems. Central in these systems is a dependency network. This network represents the dependencies between propositions, which are represented as nodes. These dependencies describe on which propositions belief in some proposition is based.

A dependency network can be used in two different ways. It can either be used to determine the propositions that can be believed, given the dependencies between them, or it can be used to determine minimal consistent sets of assumptions on which belief in a proposition is based.

The former approach originates from J. Doyle's JTMS [17]. In a JTMS a dependency is described by an n-tuple

$$\langle m_1, ..., m_j; n_1, ..., n_k \rightarrow c \rangle ,$$

which is called a justification. In such a justification, $m_1, ..., m_j$ are called the monotonic antecedents, $n_1, ..., n_k$ are called the non-monotonic antecedents and c is called the consequent. Consequent and antecedents represent nodes in the dependency network. The justifications justify belief in a proposition c if and only if there exists a justification whose monotonic antecedents $m_1, ..., m_j$ are believed, and if none of its non-monotonic antecedents $n_1, ..., n_k$ are believed. The problem of determining the set of believed propositions is for general dependency networks an NP-Hard problem.

The latter approach originates from J. de Kleer's ATMS [31]. In an ATMS the justifications are used to determine for a proposition each minimal set of assumptions

of which the elements must be believed to justify belief in this proposition. To determine these sets of assumptions, in the original ATMS only monotonic justifications are used. Monotonic justifications are justifications that have no non-monotonic antecedents.

In [24], J. W. Goodwin argues that reason maintenance systems should be viewed as a process oriented logic that focuses on the process of inference in a logic. In this, it differs from the logic itself, which only focuses on derivability; i.e. logics only characterize the set of theorems that follow from the set of premisses. If a non-monotonic logic possesses a proof theory, this proof theory is only intended to verify, at least in principle, whether a set of formulas is a set of theorems given the premisses. A proof theory is not a theory about how we can determine or approximate the set of theorem in practical situations. Furthermore, since some non-monotonic logic do not satisfy the compactness theorem, they do not have a proof theory. For example, Reiter's default logic does not have a proof theory.

Goodwin's *process oriented logic* concentrates on the *process* of determining the set of theorems. He argues that reasoning should be viewed as a process of adopting new constraints on what is *currently proven*. These constraints are added as justifications in a dependency network. For example, using the modus ponens, constraints of the form: 'if φ and $\varphi \rightarrow \psi$ are currently proven, then ψ must be currently proven' can be derived. The reason maintenance process uses these constraints to determine what is currently proven. The set of propositions currently proven, may change non-monotonically if new constraints are added. In the limit, however, the set of currently proven propositions will become equal to the set of theorems. According to Goodwin, the process non-monotonicity in this approximation process is just another side of non-monotonic logics.

1.3 Inheritance networks

Inheritance networks emerged as a branch of the semantic networks. Like the semantic networks, the inheritance networks use a graph for representing knowledge. In an inheritance network the nodes represent specific objects and classes of objects. The edges between the nodes represent the *is-a* or *is-not-a* relations. To derive conclusions from an inheritance network, procedures taking the network as input are used. Although all today's networks agree on the fact that the procedure should use pre-emption [63, 28, 64, 61], i.e. the properties of a class can be overruled by a subclass, they do not agree if multiple inheritance occurs. In [64], Touretzky et al. discuss some of the intuitions behind multiple inheritance. In my opinion, the problems with multiple inheritance arise from the lack of a model based semantics for their inheritance networks. Some authors have defined the semantics of the network in terms of propositions in a non-monotonic logic [19, 62, 5, 20]. This implies that the inheritance network is only used as an implementation technique for a subset of the logic. F. Bacchus [2] and L. Shastri [57] both describe a semantics for inheri-

tance networks in which the *is-a* relation is modelled as denoting that a percentage of the objects of a class also belongs to some other class. As Bacchus points out, this interpretation of the *is-a* link limits the applicability of an inheritance networks.

In [34], T. Krishnaprasad ad W. Kifer claim to describe a model based semantics for inheritance networks. What they actually describe, however, is a semantics for a very limited logic language. In [35], the same authors together with D. S. Warren do describe a model based semantics for inheritance networks. To be able to handle the problem of multiple inheritance, they define a preference relation between the *is-a* and *is-not-a* relations, using a special link between their consequent nodes.

A number of problems of inheritance networks are their limited expressive power, the ad hoc nature of their proof procedures and the lack of a model based semantics for most of the networks. Therefore, in my opinion, if there is any practical use for inheritance networks, it is for implementing a limited subset of some non-monotonic logic. In this thesis I will not consider inheritance networks any further.

1.4 Belief revision

The reason to call a non-monotonic logic non-monotonic is that the set of theorems of a set of premisses does not have to grow monotonically in such a logic if the set of premisses is growing monotonically; i.e. it does not satisfy the property:

$$A \subset B \text{ if and only if } TH(A) \subseteq TH(B),$$

where A and B are sets of premisses and TH is a function that maps a set of premisses on its set of theorems. If we identify the set of theorems as the belief set of an ideal omniscient reasoning agent, we can say that the agent may have to revise its belief set when it receives new information.

P. Gärdenfors has studied the revision of an agent's belief set independent of some underlying (non-monotonic) logic. In his book 'Knowledge in Flux' [23], Gärdenfors studies the *dynamics* of belief. He identifies a belief set with an epistemic state containing a deductively closed set of propositions an agent holds to be true. Given such a belief set of an ideal reasoning agent, Gärdenfors studies the changes of the belief set when new information is received. He describes three possible ways of changes, *expansion, revision* and *contraction.* These changes are the result of new information coming available. They are described by three functions, $B^+[\alpha]$, $B^*[\alpha]$ and $B^-[\alpha]$, denoting respectively the expansion, revision and contraction of a belief set B with respect to a formula α. For each of these three functions, he describes a set of *rationality postulates* that should be satisfied by the appropriated belief changes.

Expansion

Expansion is the change being the result of learning something. Gädenfors identifies the following postulates for expansion of a belief set B with respect to a formula α.

The key idea behind these postulates, and also behind the postulates for revision and contraction, is that we want to retain the old beliefs as much as possible when changing the belief set.

1. $B^+[\alpha]$ is a belief set.
2. $\alpha \in B^+[\alpha]$.
3. $B \subseteq B^+[\alpha]$.
4. If $\alpha \in B$, then $B^+[\alpha] = B$.
5. If $B \subseteq H$, then $B^+[\alpha] \subseteq H^+[\alpha]$.
6. For all belief sets B and all formulas α, $B^+[\alpha]$ is the smallest belief set that satisfies the postulates 1 to 5.

Given these postulates, the following important theorem can be proven.

Theorem 1.40 The expansion function satisfies the postulates 1 to 6 if and only if $B^+[\alpha] = Th(B \cup \{\alpha\})$.

revision

Revision is more or less the same as expansion, except for demanding the resulting belief set to be consistent. This implies that if an agent learns α and $\neg\alpha$ is in its belief set, it must give up some of its beliefs to be able to accept α. Hence, the belief set changes non-monotonically. So, belief revision is closely related to non-monotonic reasoning. In fact, according to Gärdenfors, one of the reasons why an agent can learn α while believing $\neg\alpha$ is because $\neg\alpha$ is some default assumption. Learning α implies that $\neg\alpha$ has to be withdrawn from the belief set. The changes of the belief set must, of course, be minimal. Gärdenfors identifies the following basic set of postulates for belief revision.

1. $B^*[\alpha]$ is a belief set.
2. $\alpha \in B^*[\alpha]$.
3. $B^*[\alpha] \subseteq B^+[\alpha]$.
4. If $\neg\alpha \notin B$, then $B^+[\alpha] \subseteq B^*[\alpha]$.
5. $B^*[\alpha] = B_\perp$ if and only if $\vdash \neg\alpha$. Here $B_\perp$ denotes the inconsistent belief set.
6. If $\vdash \alpha \leftrightarrow \beta$, then $B^*[\alpha] = B^*[\beta]$.

Contraction

Contraction is the change that results from stopping to believe a formula while no new formulas are added. This kind of change may occur when an agent learns that what he had learned before, comes from an irreliable source. For contraction Gärdenfors identifies the following basic set of postulates.

1. $B^-[\alpha]$ is a belief set.
2. $B^-[\alpha] \subseteq B$.
3. If $\alpha \notin B$, then $B^-[\alpha] = B$.
4. If $\not\vdash \alpha$, then $\alpha \notin B^-[\alpha]$.
5. If $\alpha \in B$, then $B \subseteq (B^-[\alpha])^+[\alpha]$.
6. If $\vdash \alpha \leftrightarrow \beta$, then $B^-[\alpha] = B^-[\beta]$.

Between expansion, revision and contraction the following two relations can be established.

Theorem 1.41 Let the contraction function satisfy the contraction postulates 1 to 4 and 6 and let the expansion function satisfy the expansion postulates 1 to 6. Then the revision function defined by:

$$B^*[\alpha] = (B^-[\neg\alpha])^+[\alpha]$$

satisfies the revision postulates 1 to 6.

Theorem 1.42 Let the revision function satisfy the revision postulates 1 to 6. Then the contraction function defined by:

$$B^-[\alpha] = B \cap B^*[\alpha]$$

satisfies the contraction postulates 1 to 6.

1.5 Research goals

In Default logic and in autoepistemic logic belief in a proposition can be based on not believing some other proposition. In Default logic this is realized by the justifications in the default rules, and in autoepistemic logic by formulas containing sub-formulas of the form $\neg B(\beta)$. Believing a formula α because we do not believe a formula β, actually implies that we implicitly assume the formula β to be false. Since β must be either true or false (no intuitionistic logics are considered), and since we may

not believe α if β is true, β must be false. In Default logic and autoepistemic logic, however, the truth value of the implicit assumptions is left undefined.

In my opinion, these implicit assumptions should be stated explicitly in a non-monotonic logic. In [55], E. Sandewall uses the same view when he defines his functional semantics for non-monotonic logics. One way to avoid these implicit assumptions is to view non-monotonic reasoning as a special case of reasoning with inconsistent knowledge. This motivated my research, reported in the following chapters. Considering non-monotonic reasoning to be a special case of reasoning with inconsistent knowledge, I wanted to investigate whether it is possible to create a non-monotonic logic based on a reasoning process that solves conflicts after they are being derived. To solve the conflicts derived, generalizing Rescher's approach, I will use a strict partial preference relation on the premisses.

If such a non-monotonic logic is possible, its relation with other non-monotonic reasoning systems should be investigated. Here, we are especially interested in the relation between its semantics and the semantics for non-monotonic logics described by Kraus et al. Furthermore, we are interested in the behaviour of the logic when new information is added. I will compare this behaviour with Gärdenfors's postulates for belief revision.

2

A preference logic

In this chapter I will describe a new non-monotonic logic based on the ideas of N. Rescher [51]. This logic is intended for reason with inconsistent knowledge. To handle inconsistencies, a partial preference relation on the set of premisses is used to choose a culprit when an inconsistency is determined. Using J. W. Goodwin's view on non-monotonic reasoning [24], a deduction process based on the generation of justifications will be developed. Furthermore, a semantics based on the ideas of Y. Shoham [59] will be defined for the logic. It is shown that this semantics fits in the family of preferred models, defined by Kraus et al. [33].

The logic can also be used for default reasoning. To be able to do this, default reasoning has to be viewed as a special case of reasoning with with inconsistent knowledge. Default rules must, of course, have a lower preference than facts. In that case they can be overruled by the facts.

2.1 Basic concepts

To be able to reason with inconsistent knowledge, I will consider premisses to be assumptions. These premisses are assumed to be true as long as we do not derive a

contradiction from them. If, however, a contradiction is derived, we have to determine the premisses on which the contradiction is based. The premisses on which a contradiction is based are the premisses used in the derivation of the contradiction. When we know these premisses, we have to remove one of them to block the derivation of the contradiction. To select a premiss to be removed, I will use a preference relation. This preference relation must define a strict partial ordering on the set of premisses. Using the preference relation, we have to remove a least preferred premiss of the inconsistent set, thereby blocking the derivation of the contradiction.

Example 2.1 Let Σ denote a set of premisses,

$$\Sigma = \{1.\ \varphi, 2.\ \varphi \rightarrow \psi, 3.\ \neg\psi, 4.\ \alpha\}$$

and $(\Sigma, \prec)$ a preference relation on Σ:

$$(\Sigma, \prec) = \{3 \prec 1, 3 \prec 2\}$$

From Σ, ψ can be derived using the premisses 1 and 2. Furthermore, a contradiction can be derived from ψ and premiss 3. Hence, the contradiction is based on the premisses 1, 2 and 3. Since premiss 3 is the least preferred premiss on which the contradiction is based, it has to be removed.

Three problems may arise when trying to remove a contradiction.

- Firstly, we have to be able to determine the premisses on which a contradiction is based. These are the premisses that are used in the derivation of the contradiction. To solve this problem, justifications are introduced. Such a justification, called *in_justification*, describes the premisses from which a formula is derived. Using Goodwin's view on justifications [24], in_justification also functions as a constraint on the set of formulas that we can believe. This set will be called the belief set.

- Secondly, a premiss that has been removed, may have to be placed back because the contradiction causing its removal cannot occur any more. This may take place because of some other contradiction being derived. To solve this problem, another kind of justifications is introduced. This type of justification is called an *out_justification*. An out_justification describes which premiss must be removed when other premisses are still assumed to be true. It is a constraint on the set of premisses we assume to be true.

- Thirdly, there need not exist a single least preferred premiss in the set of premisses on which a contradiction is based. In such a situation there are two possible choices.

  - Do nothing. The contradiction is not solved but this does not imply that the contradiction will not be solved at all [52].

- Consider the results of the removal of every alternative apart. As a result of this policy, we have to consider different subsets of the set of premisses. It is possible that these subsets will converge to one consistent subset of the set of premisses. If this happens, the result of the two approaches will be the same.

As already mentioned in the introduction, default reasoning will be treated as a special case of reasoning with inconsistent knowledge. Default rules are general rules which may contradict specific information. When this occurs, we have to prefer specific information to general information. For example, the specific information 'Tweety cannot fly because it is a penguin' should be preferred to the general information 'Birds can fly'. The question is how to represent the general information. It is not possible to describe the sentence 'Birds can fly' by:

$$\forall x[Bird(x) \rightarrow Can\_fly(x)]$$

If there is one bird that cannot fly, this premiss will be removed making it impossible to derive for any bird that it can fly. Since this is undesirable, I will introduce an alternative approach to represent defaults. In the predicate logic a premiss φ containing free variables $\overline{x}$ is equivalent to $\forall \overline{x}\varphi$. Here, however, like Reiter's open default rules, a formula φ containing free variables, is interpreted as denoting a set of instances of this formula. When a member of this set is the least preferred premiss of a set on which a contradiction is based, only this instance is removed.

Example 2.2 Suppose that the following premisses are given.

1. $Bird(x) \rightarrow Can\_fly(x)$
2. $Bird(Tweety)$
3. $\neg Can\_fly(Tweety)$

If the second and the third premiss are preferred to the instance of the first premiss:

$$Bird(Tweety) \rightarrow Can\_fly(Tweety)$$

then only this instance will be removed, but not the first premiss.

In Reiter's default logic ground terms must be substituted for the free variables that occur in a default rule. Here I will not limit the instances of a formula to ground instances only. Allowing every possible instance of a formula has two advantages. Firstly, we can avoid proliferation of ground instances. If we derive a formula containing free variables, every instance of the formula will also be a logical consequence of the premisses unless it is overruled by some other formula. Secondly, as a result of this, it becomes possible to derive new default rules in the logic.

A question that has to be answered yet is: 'how is the preference relation defined on the premisses related to a preference relation on *instances* of these premisses?'. To motivate the answer of this question, consider the following situation. Suppose that a problem can be described by two premisses of which one contains a free variable.

1. $\varphi(x)$
2. $\forall x \neg \varphi(x)$

Clearly this set of premisses is inconsistent. Now suppose that the second premiss is preferred to the first, then we have to remove the whole set of premisses denoted by the first premiss. Because we can also derive a contradiction with each instance of the first premiss, the second premiss has to be preferred to each instance of the first premiss. Therefore, each instance of the set generated by a premiss containing free variables should have the same preferences as this premiss.

Condition 2.3 Every instance of a premiss $\dot{\varphi}$ containing free variables should have the same preference as φ.

A possible extension of the logic would be to permit that a preference relation is specified on the instances of a formula.

2.2 Formal definitions

In the formal description of the preference logic, unification will be used [39]. To unify two formulas, a substitution of terms for free variables may be required. Such a substitution θ for the free variables is denoted by placing $[\theta]$ behind a formula. A substitution that has to be carried out on every formula of a set of formulas, or on every formula occurring in a justification, is denoted in the same way.

The preference logic is based on an ordinary first order logic L. A set of premisses Σ of this logic is some subset of this language L. On this set of premisses a preference relation can be defined. This preference relation $\prec$ for a set of premisses Σ generates a strict partial order $(\Sigma, \prec)$.

Because premisses containing free variables are viewed as representing a set of instances of those premisses, an extended set of premisses $\overline{\Sigma}$, also containing all instances, is introduced.

Definition 2.4 Let S be a set of formulas. Then $\overline{S}$ denotes the *extended set of formulas*, which also contains all instances of the formulas of S.

$$\overline{S} = \{\varphi \mid \psi \in S \text{ and for some substitution } \theta : \varphi = \psi[\theta]\}$$

In case a contradiction is derived, a formula from the extended set of premisses $\overline{\Sigma}$ has to be withdrawn. To be able to do this, it is necessary to extend the preference

relation. This extended preference relation should satisfy Condition 2.3 and should again be a strict partial order. The preference relation for the extended set of premisses is defined as follows:

Definition 2.5 Let Σ be a set of premisses and let $(\Sigma, \prec)$ be a strict partial preference relation on Σ. Furthermore, let $(\overline{\Sigma}, \prec)$ be the preference relation on the extended set of premisses $\overline{\Sigma}$. $(\overline{\Sigma}, \prec)$ is the smallest strict partial order containing $(\Sigma, \prec)$, being invariant under term substitution in the premisses of Σ.

One should notice that there need not be an extended preference relation $(\overline{\Sigma}, \prec)$, as can be seen in the following example.

Example 2.6 Let Σ be a set of premisses and let $(\Sigma, \prec)$ be a preference relation on these premisses.

$$\begin{aligned} \Sigma &= \{\varphi(x), \varphi(a), \psi\} \\ (\Sigma, \prec) &= \{\varphi(x) \prec \psi, \psi \prec \varphi(a)\} \end{aligned}$$

Clearly, there does not exist an asymetric preference relation on the extended set of premisses.

Now the set of extended premisses and their preference relation has been defined, the justifications can be defined. Two kinds of justifications, *in-justifications* and *out-justifications*, are distinguished. The in-justifications are used to denote that a formula is believed if the premisses in the antecedent are believed, while the out-justifications are used to denote that a premiss can no longer be believed (must be withdrawn) when the premisses in the antecedent are believed.

Definition 2.7 Let Σ be a set of premisses. Then the set of possible in_ and out_justifications is defined as follows:

$$\begin{aligned} In\_Just(\Sigma) &= \{P \Rightarrow \varphi \mid P \subset \overline{\Sigma} \text{ and } \varphi \in L\} \\ Out\_Just(\Sigma) &= \{P \not\Rightarrow \varphi \mid P \subset \overline{\Sigma} \text{ and } \varphi \in \overline{\Sigma}\} \end{aligned}$$

2.3 The deduction process

Instead of deriving new formulas, in the preference logic only new justifications are derived. These justifications are generated by the inference rules. Because the inference rules are defined on justifications and not on formulas, and because justifications function as constraints on the belief set, *Reason (Truth) Maintenance* can be viewed as part of the deduction process. So, a deduction process in the preference logic can be viewed as a process of belief revision in the same way as Goodwin's *logical process theory* [24]. The deduction process will finally terminate

with the belief set, being the set of theorems for the models of the set of premisses and the preference relation. How these models are defined, can be found in section 2.5.

A deduction process for the preference logic starts with an initial set of justifications J_0. This initial set J_0 contains an in-justification for every premiss. These justifications indicate that a formula is believed if the corresponding premisses are believed.

Definition 2.8 Let Σ be a set of premisses. Then the set of initial justifications J_0 is defined as follows:

$$J_0 = \{\{\varphi\} \Rightarrow \varphi \mid \varphi \in \Sigma\}$$

Each set of justifications J_i with $i > 0$ is generated from the set J_{i-1} by adding new justifications. How these justifications are determined, depends on the deduction system used. In the following description of the preference logic I will use an axiomatic deduction system for a language L, only containing the logical operators $\rightarrow$ and $\neg$ and the quantifier $\forall$. The logical axioms used originates from [18]. There is no specific reason why I have chosen these axioms. In fact, any set of logical axioms for a first order logic with the modus ponens as the only inference rule can be used here.

Axioms Let φ be a generalization of ψ if and only if for some $n \geqslant 0$ and some variables $x_1, ..., x_n$:

$$\forall x_1, ..., \forall x_n \, \psi.$$

Since this definition includes the case $n = 0$, any formula is a generalization of itself.

The logical axioms are all the generalizations of the formulas described by the following schemata.

1. Tautologies.
2. $\forall x \varphi(x) \rightarrow \varphi(y)$ where y is a variable that does not occur in φ; i.e. y is a free variable in $\varphi(y)$.
3. $\forall x(\varphi \rightarrow \psi) \rightarrow (\forall x \varphi \rightarrow \forall x \psi)$.
4. $\varphi \rightarrow \forall x \varphi$ where x does not occur in φ.

The second axiom scheme differs from the axiom scheme described in [18]. In [18] this axiom scheme is stated as follows:

$\forall x \varphi(x) \rightarrow \varphi(t)$ where t is a term containing no variables that occur in φ.

Since here a formula containing free variables denotes a set of instances, clearly both formulations are equivalent. An advantage of the formulation chosen here is that no unnecessary instances of formulas will be generated by the deduction process described below.

Because an axiomatic approach is used, justifications for the axioms have to be introduced. Since axioms cannot be withdrawn, an axiom will always have an in-justification with an antecedent equal to the empty set. An axiom is introduced by the following axiom rule.

Rule 2.9 An axiom φ gets an in_justification $\varnothing \Rightarrow \varphi$.

In the deduction system two inference rules will be used, namely the modus ponens and the contradiction rule. The modus ponens introduces a new in_justification for some formula. This justification is constructed from the justifications for the antecedents of the modus ponens.

Rule 2.10 Let φ and $\psi \rightarrow \mu$ be two formulas with justifications, respectively $P \Rightarrow \varphi$ and $Q \Rightarrow (\psi \rightarrow \mu)$.

> If φ and ψ can be unified with a most general unifier θ, then the formula $\mu[\theta]$ gets an in_justification $((P \cup Q) \Rightarrow \mu)[\theta]$.

While the modus ponens introduces a new in_justification, the contradiction rule introduces a new out_justification to eliminate a contradiction.

Rule 2.11 Let φ and $\neg\psi$ be formulas with justifications $P \Rightarrow \varphi$ and $Q \Rightarrow \neg\psi$. Let φ and ψ be unifiable and θ be a most general unifier.

> If $R = \mathit{min}(P \cup Q)$, then each premiss $\eta \in R$ gets an out_justification $(((P \cup Q)/\eta) \not\Rightarrow \eta)[\theta]$.

> Here, the function $\mathit{min}(X)$ selects the set of least preferred premisses from a set of premisses X.

In order to guarantee that the current belief set will approximate the set of theorems of the premisses with respect to the preference relation, we have to guarantee that the process creating new justifications is fair. By this I mean that this process does not forever defer the addition of some possible justification to the set of justifications.

Assumption 2.12 The reasoning process will not defer the addition of any possible justification to the set of justifications forever.

If a fair process is used, the following theorems hold. The first theorem guarantees the soundness of the in_justifications; i.e. the antecedent of an in_justification logically implies the consequent of the in_justification. The second theorem guarantees the completeness of the in_justifications; i.e. if a formula is logically implied

by a subset of the premisses, then there exists a corresponding in_justification. Finally, the third and fourth theorem guarantee respectively the soundness and the completeness of the out_justifications.

Theorem 2.13 *Soundness*
For each $i \geqslant 0$:

if $P \Rightarrow \varphi \in J_i$, then for each substitution θ:

$P[\theta] \subseteq \overline{\Sigma}$ and $P[\theta] \models \varphi[\theta]$.

Proof By the soundness of first order logic,

if $P[\theta] \vdash \varphi[\theta]$, then $P[\theta] \models \varphi[\theta]$.

Therefore, we only have to prove that for each $i \geqslant 0$:

if $P \Rightarrow \varphi \in J_i$, then for each substitution θ:

$P[\theta] \subseteq \overline{\Sigma}$ and $P[\theta] \vdash \varphi[\theta]$.

I will prove this by induction on the index i of J_i.

- For $i = 0$:

  $\{\varphi\} \Rightarrow \varphi \in J_0$ if and only if $\varphi \in \Sigma$.

  Since $\overline{\Sigma}$ is closed under term substitution, for each substitution θ:

  $\varphi[\theta] \in \overline{\Sigma}$.

  Therefore, for each substitution θ:

  $\{\varphi[\theta]\} \vdash \varphi[\theta]$.

- Proceeding inductively, suppose that $P \Rightarrow \varphi \in J_{k+1}$.
  Then:

  $P \Rightarrow \varphi \in J_{k+1}$ if and only if $P \Rightarrow \varphi \in J_k$ or $P \Rightarrow \varphi$ has been added by Rule 2.9 or 2.10.

  – If $P \Rightarrow \varphi \in J_k$, then, by the induction hypothesis, for each substitution θ:

  $P[\theta] \subseteq \overline{\Sigma}$ and $P[\theta] \vdash \varphi[\theta]$.

– If $P \Rightarrow \varphi$ is introduced by Rule 2.9, then it is an axiom.
Therefore, $P = \varnothing$ and for each substitution θ:

$$\vdash \varphi[\theta].$$

– If $P \Rightarrow \varphi$ is introduced by Rule 2.10, then there is a $Q \Rightarrow \alpha \in J_k$, $R \Rightarrow (\beta \rightarrow \psi) \in J_k$, such that α and β are unifiable with a most general unifier θ.
So we have:

$$P = (Q \cup R)[\theta] \text{ and } \varphi = \psi[\theta].$$

According to the induction hypothesis for each substitution ζ:

$$Q[\theta \circ \zeta], R[\theta \circ \zeta] \subseteq \overline{\Sigma},$$

$$Q[\theta \circ \zeta] \vdash \alpha[\theta \circ \zeta]$$

and

$$R[\theta \circ \zeta] \vdash (\beta \rightarrow \psi)[\theta \circ \zeta].$$

Therefore, for each substitution ζ:

$$P[\zeta] \subseteq \overline{\Sigma} \text{ and } P[\zeta] \vdash \varphi[\zeta].$$

□

Theorem 2.14 *Completeness*
For each $P \subseteq \overline{\Sigma}$:

if $P \models \varphi$, then for some $i \geqslant 0$:

$$Q \Rightarrow \psi \in J_i$$

and for some substitution θ:

$$Q[\theta] \subseteq P \text{ and } \psi[\theta] = \varphi.$$

Proof Let $P \subseteq \overline{\Sigma}$ and $P \models \varphi$.
By the completeness of first order logic,

$$\text{if } P[\theta] \models \varphi[\theta], \text{ then } P[\theta] \vdash \varphi[\theta].$$

Since $P \vdash \varphi$, there exists a deduction sequence $\langle \varphi_0, \varphi_1, ..., \varphi_n \rangle$ such that $\varphi_n = \varphi$ and for each $j \leqslant n$: either

- $\varphi_j \in P$, or
- φ_j is an axiom, or

- there exists a φ_k and a φ_l with $k, l < j$ and $\varphi_l = \varphi_k \rightarrow \varphi_j$.

The theorem will be proven, using induction on the length n of the deduction sequence.

- For $n = 1$, $\langle \varphi_1 \rangle$ is the deduction sequence for $P \vdash \varphi$.
  - If $\varphi_1 \in P$, then $\varphi_1 \in \overline{\Sigma}$ and there exists a $\psi \in \Sigma$ such that for some substitution θ:
    $$\psi[\theta] = \varphi.$$
  - If φ_1 is an axiom, then there exists some $i_0 \geqslant 0$ such that:
    $$J_{i_0} = J_{i_0 - 1} \cup \{\varnothing \Rightarrow \varphi_0\} \text{ and } \varnothing \Rightarrow \varphi_0 \text{ is added by Rule 2.9.}$$

  Hence the theorem holds for deduction sequences of length 1.

- Proceeding inductively, let $\langle \varphi_0, \varphi_1, ..., \varphi_{m+1} \rangle$ be a deduction sequence for $P \vdash \varphi_{m+1}$.
  - If $\varphi_{m+1} \in P$, then $\{\psi\} \Rightarrow \psi \in J_0$ and for some substitution θ:
    $$\varphi_{m+1} = \psi[\theta].$$
  - If φ_{m+1} is an axiom, then there is some i_{m+1} such that:
    $$J_{i_{m+1}} = J_{i_{m+1}-1} \cup \{\varnothing \Rightarrow \varphi_{m+1}\} \text{ and } \varnothing \Rightarrow \varphi_{m+1} \text{ is added by Rule 2.9.}$$
  - If there exists a φ_k and a φ_l with $k, l \leqslant m + 1$ and $\varphi_l = \varphi_k \rightarrow \varphi_{m+1}$, then, by the induction hypothesis, there exists some i_k and some i_l such that:
    $$Q \Rightarrow \alpha \in J_{i_k},$$
    $$R \Rightarrow (\beta \rightarrow \psi) \in J_{i_l}$$
    and for some substitution θ:
    $$Q[\theta] \subseteq P \text{ and } \varphi_k = \alpha[\theta],$$
    and for some substitution ζ:
    $$R[\zeta] \subseteq P \text{ and } \varphi_l = (\beta \rightarrow \psi)[\zeta].$$
    Since $\alpha[\theta] = \beta[\zeta] = \varphi_k$, α and β are unifiable.
    Let ξ be a most general unifier.
    Because of the fairness Assumption 2.12, there exists some i_{m+1} with $i_k, i_l < i_{m+1}$ such that:
    $$(Q \cup R \Rightarrow \psi)[\xi] \in J_{i_{m+1}}$$

and for some substitution σ:

$$\varphi_{m+1} = \psi[\xi \circ \sigma].$$

Hence there exists some i_{m+1} such that $S \Rightarrow \psi \in J_{i_{m+1}}$ and for some substitution θ:

$$\varphi_{m+1} = \psi[\theta].$$

□

Theorem 2.15 *Soundness*
For each $i \geqslant 0$:

if $P \not\Rightarrow \varphi \in J_i$, then for each substitution θ:

$(P \cup \{\varphi\})[\theta] \subseteq \overline{\Sigma}$,

$(P \cup \{\varphi\})[\theta]$ is not satisfiable

and for each $\psi \in P[\theta]$:

$\psi \not\leftarrow \varphi[\theta]$.

Proof The theorem will be proven using induction to the index i of the set of justifications J_i.

- For $i = 0$: the theorem holds vacuously because there is no $P \not\Rightarrow \varphi \in J_0$.
- Proceeding inductively, suppose that $P \not\Rightarrow \varphi \in J_{k+1}$.
  $P \not\Rightarrow \varphi \in J_{k+1}$ if and only if $P \not\Rightarrow \varphi \in J_k$ or $P \not\Rightarrow \varphi$ has been added by Rule 2.11.
  - If $P \not\Rightarrow \varphi \in J_k$, then, by the induction hypothesis, for a substitution θ:

    $(P \cup \{\varphi\})[\theta] \subseteq \overline{\Sigma}$,

    $(P \cup \{\varphi\})[\theta]$ is not satisfiable

    and for each $\psi \in P[\theta]$:

    $\psi \not\leftarrow \varphi[\theta]$.
  - If $P \not\Rightarrow \varphi$ is introduced by Rule 2.11, then there is an $R \Rightarrow \alpha \in J_k$, $Q \Rightarrow \neg\beta \in J_k$ and α and β are unifiable.
    If ζ is a most general unifier, then $\varphi \in min((Q \cup R)[\zeta])$ and $P = ((R \cup Q)[\zeta])/\varphi$.
    By Theorem 2.13 for each substitution ξ:

$$R[\xi], Q[\xi] \subseteq \overline{\Sigma},$$

$$R[\xi] \vdash \alpha[\xi]$$

and

$$Q \vdash \neg\beta[\xi].$$

Hence for each substitution θ:

$$(P \cup \{\varphi\})[\theta] \subseteq \overline{\Sigma},$$

and

$(P \cup \{\varphi\})[\theta]$ is inconsistent.

Since inconsistency implies unsatisfiability, for each substitution θ:

$$(P \cup \{\varphi\})[\theta] \subseteq \overline{\Sigma},$$

$(P \cup \{\varphi\})[\theta]$ is not satisfiable

and for each $\psi \in P[\theta]$:

$$\psi \not< \varphi[\theta].$$

□

Theorem 2.16 *Completeness*
For each $P \subseteq \overline{\Sigma}$:

if P is a minimal unsatisfiable set of premisses and $Q = min(P)$, then for some $i \geqslant 0$ there holds for each $\varphi \in Q$:

$$R \not\Rightarrow \psi \in J_i,$$

and for some substitution θ:

$$Q = R[\theta] \text{ and } \varphi = \psi[\theta].$$

Proof Let P be a minimal unsatisfiable subset of $\overline{\Sigma}$ with $Q = min(P)$.
Since P is a minimal unsatisfiable set, P is a minimal inconsistent set.
Therefore, there exists a formula α such that:

$$P \vdash \alpha \text{ and } P \vdash \neg\alpha.$$

By Theorem 2.14 there exists a $j \geqslant 0$:

$$S \Rightarrow \beta \in J_j$$

and for some substitution ζ:

$$S[\zeta] \subseteq P \text{ and } \alpha = \beta[\zeta].$$

Also by Theorem 2.14 there exists a $k \geqslant 0$:

$$T \Rightarrow \neg\gamma \in J_k$$

and for some substitution ξ:

$$T[\xi] \subseteq P \text{ and } \alpha = \gamma[\xi].$$

Since β and γ are unifiable, there exists a most general unifier σ.
Hence, for some substitutions θ_1, θ_2:

$$(S[\sigma \circ \theta_1] \cup T[\sigma \circ \theta_2]) \subseteq P.$$

Since P is minimal inconsistent:

$$(S[\sigma \circ \theta_1] \cup T[\sigma \circ \theta_2]) = P.$$

Hence, for some substitution θ:

$$(S[\sigma \circ \theta_1] \cup T[\sigma \circ \theta_2]) = (S \cup T)[\sigma \circ \theta]$$

and

$$Q = min(P) = min((S \cup T)[\sigma \circ \theta]).$$

Therefore, there exists an $l > j, k$ such that for each $\varphi \in Q$, there is a $\psi \in min(S \cup T)$:

$$\varphi = \psi[\theta],$$

$$((R \cup S)/\psi)[\sigma] \not\Rightarrow \psi[\sigma] \in J_l$$

and

$$R = P/\varphi = (((S \cup T)/\psi)[\sigma])[\theta].$$

Hence for some $i \geqslant 0$:

$$P/\varphi \not\Rightarrow \psi \in J_i$$

and for some substitution θ:

$$P = Q[\theta] \text{ and } \psi[\theta] = \varphi.$$

□

Given a set of justifications, there may exist one or more subsets of the set of premisses which can be believed. Such a subset contains the premisses that do not have to be withdrawn because of an out-justification. Suppose that J_i is a set of justifications derived by a reasoning agent and that $\Delta \subseteq \overline{\Sigma}$ is a subset of the premisses that are believed by the reasoning agent. Then for each premiss ψ such that for some out_justification $P \not\Rightarrow \varphi \in J_i$ and some substitution θ, there holds that $P[\theta] \subseteq \Delta$ and $\psi = \varphi[\theta]$, one may not believe ψ. The set of premisses that may not be believed given a set of justification J_i, is denoted by $Out_i(\Delta)$.

Definition 2.17

$$Out_i(S) = \{\varphi[\theta] \mid P \not\Rightarrow \varphi \in J_i, \text{ and for some substitution } \theta : P[\theta] \subseteq S\}$$

The set of premisses Δ that we may believe, must, of course, be equal to the set of premisses we will get after having removed all the premisses that we may not believe; i.e. $\Delta = \overline{\Sigma} - Out_i(\Delta)$. The sets of premisses that satisfy these requirements, are defined by the following fixed point definition.

Definition 2.18 Let $\overline{\Sigma}$ be an extended set of premisses and let J_i be a set of justifications. Furthermore, let $\mathcal{A}_i$ be the set containing all the subsets of the premisses that can be believed given the out_justifications in J_i.
Then:

$$\mathcal{A}_i = \{\Delta \mid \Delta = \overline{\Sigma} - Out_i(\Delta)\}$$

After having determined all the sets of premisses that can be believed, the set of derived formulas that can be believed, can be determined given the in_justifications. This set is defined as:

Definition 2.19 Let J_i be a set of justifications and $\mathcal{A}_i$ be the corresponding the of sets of believed premisses.
The set of formulas B_i that can be believed (*the belief set*) is defined as:

$$B_i = \{\psi[\theta] \mid \text{ for each } \Delta \in \mathcal{A}_i \text{ there is a } P \Rightarrow \psi \in J_i \text{ such that for some substitution } \theta\text{: } P[\theta] \subseteq \Delta\}$$

Property 2.20 For each $\varphi \in B_i$: $[\Delta \vdash \varphi$ for each $\Delta \in \mathcal{A}_i]$

Proof Suppose $\varphi \in B_i$.
Then for each $\Delta \in \mathcal{A}_i$ there exists a

$$P \Rightarrow \psi \in J_i$$

and for some substitution θ:

$$\varphi[\theta] = \psi \text{ and } P[\theta] \subseteq \Delta.$$

Therefore, by Theorem 2.13:

$$P \vdash \varphi \text{ and } P[\theta] \subseteq \Delta$$

Hence, for each $\Delta \in \mathcal{A}_i$:

$$\Delta \vdash \varphi.$$

□

J_∞ is defined as the set of all justifications which can be derived.

Definition 2.21 $J_\infty = \bigcup_{i \geqslant 0} J_i$

The corresponding sets of premisses and formulas that can be believed, will be denoted by $\mathcal{A}_\infty$ and by B_∞. For J_∞, $\mathcal{A}_\infty$ and B_∞ the following properties can be proven:

Property 2.22 For each $\Delta \in \mathcal{A}_\infty$: $\Delta \subseteq \overline{\Sigma}$.

Proof Since $\Delta = \overline{\Sigma} - Out(\Delta)$, $\Delta \subseteq \overline{\Sigma}$. □

Property 2.23 For each $\Delta \in \mathcal{A}_\infty$: Δ is maximal consistent.

Proof Suppose that some $\Delta \in \mathcal{A}_\infty$ is inconsistent.
Then there exists a minimal inconsistent subset M of Δ.
Let $\varphi \in min(M)$.
Then by Theorem 2.16 there exists an i with

$$P \not\Rightarrow \psi \in J_i$$

and for some substitution θ:

$(P \cup \{\psi\})[\theta] = M$ and $\varphi = \psi[\theta]$.

Hence $P \not\Rightarrow \psi \in J_\infty$.
Because $P[\theta] \subseteq \Delta$, $\varphi \notin \Delta$.
Contradiction.

Suppose that some $\Delta \in \mathcal{A}_\infty$ is not maximal consistent.
Then there exists a $\varphi \in (\overline{\Sigma} - \Delta)$ and $\{\varphi\} \cup \Delta$ is consistent.
Since $\varphi \in (\overline{\Sigma} - \Delta)$, $\varphi \in Out_\infty(\Delta)$.
Therefore, there exists a $P \not\Rightarrow \psi \in J_\infty$ and for some substitution θ:

$P[\theta] \subseteq \Delta$ and $\varphi = \psi[\theta]$.

Since $P \not\Rightarrow \psi \in J_\infty$, $(P \cup \{\psi\})[\theta]$ is inconsistent.
Hence $\Delta \cup \{\varphi\}$ is inconsistent.
Contradiction. □

Property 2.24 If each minimal inconsistent subset of $\overline{\Sigma}$ has only one least preferred element and there exists no infinite sequence of minimal inconsistent subsets such that a minimal element of one subset is an element of another subset in which it is not a minimal element, then $|\mathcal{A}_\infty| = 1$.

Proof Suppose that the condition of the property holds and that $|\mathcal{A}_\infty| > 1$.
Then there exist at least two subsets Δ, Δ' of $\overline{\Sigma}$.
Let φ be any formula such that $\varphi \notin \Delta$ and $\varphi \in \Delta'$.
By Theorem 2.16 there exists a $P \not\Rightarrow \psi \in J_\infty$ and for some substitution θ such that:

$\varphi = \psi[\theta]$

and

$(P \cup \{\psi\})[\theta]$ is a minimal inconsistent set.

Because each minimal inconsistent set has only one least preferred element, for every $\eta \in P$ there holds:

$\psi[\theta] \prec \eta[\theta]$.

Since $\varphi \notin \Delta$ and $\varphi \in \Delta'$, there exists an $\eta \in P$:

$\eta[\theta] \in \Delta$ and $\eta[\theta] \notin \Delta'$.

Hence there exists an infinite sequence of minimal inconsistent subsets such that a minimal element of one subset is a non minimal element of another subset. Contradiction.

Hence Δ is unique. □

Property 2.25

$$B_\infty = \bigcap_{\Delta \in \mathcal{A}_\infty} Th(\Delta)$$

where $Th(S) = \{\varphi \mid S \vdash \varphi\}$

Proof For each $\Delta \in \mathcal{A}_\infty$:

$$B_\infty \subseteq Th(\Delta)$$

because according to Property 2.20:

if $\varphi \in B_\infty$, then for each $\Delta \in \mathcal{A}_\infty$:

$$\Delta \vdash \varphi.$$

Suppose there exists a φ such that:

$$\varphi \notin B_\infty \text{ and } \varphi \in \bigcap_{\Delta \in \mathcal{A}_\infty} Th(\Delta).$$

Since $\varphi \in \bigcap_{\Delta \in \mathcal{A}_\infty} Th(\Delta)$, for each $\Delta \in \mathcal{A}_\infty$:

$$\Delta \vdash \varphi.$$

By Theorem 2.14 for every $\Delta \in \mathcal{A}_\infty$, there exists some i and some $Q \Rightarrow \psi \in J_i$ such that for some substitution θ:

$$Q[\theta] \subseteq \Delta \text{ and } \varphi = \psi[\theta].$$

Therefore, for every $\Delta \in \mathcal{A}_\infty$ there exists some i and some $Q \Rightarrow \psi \in J_\infty$ such that for some substitution θ:

$$Q[\theta] \subseteq \Delta \text{ and } \varphi = \psi[\theta].$$

Hence, by Definition 2.20:

$$\varphi \in B_\infty.$$

Contradiction.

Hence $B_\infty = Th(\Delta_\infty)$. □

2.4 Determination of the belief set

In this section I will describe an algorithm that can be used to determine a set of premisses that can be believed, given a set of out_justifications. The algorithm determines a single set Δ from $\mathcal{A}_i$ given the justifications J_i. The idea behind this algorithm, which was suggested by C .Witteveen, is to order the current set of out_justifications. Suppose that we prefer an out_justification j_1 to an out_justification j_2 if two conditions are satisfied. Firstly, the consequent of j_1 must occur in the antecedent of j_2 and secondly, the consequent of j_2 may not occur in the antecedents of j_1. Since the consequent of an out_justification is never preferred to any premiss in the antecedent, the ordering defined is a strict partial ordering. Clearly, if a justification j_1 is preferred to j_2, the application of j_1 will influence the application of j_2, but not the other way around. If there is no ordering between two justifications, then either they cannot influence each other or they mutually influence each other. If j_1 and j_2 mutually influence each other, then the consequent of j_1 occurs in the antecedent of j_2 and the other way around. Therefore, the application of j_1 cannot undo the application of j_2 and vice versa. Finally, if j_1 and j_2 do not influence each other, it makes no difference which one is applied first. Hence, applying the most preferred out_justifications first, will guarantee that the application of an out_justification will never have to be undone.

To make the algorithm more efficient, first all non minimal inconsistent sets of premisses described by the out_justifications are removed by determining the set *min_out_just*. This can be executed in $\mathcal{O}(n^2)$ steps, where n is the number of out_justifications. The order of the out_justifications can also be determined in $\mathcal{O}(n^2)$ steps. Finally, the repeat loop can be executed in $\mathcal{O}(n \cdot k)$ steps, where k is the number of premisses in *prem*. Hence, the algorithm can be executed in $\mathcal{O}(n \cdot max(k, n))$ steps.

begin
prem := $\{\varphi \mid \varphi$ occurs in some justification of $J_i\}$;
min_out_just := $\{P \not\Rightarrow \varphi \mid P \not\Rightarrow \varphi \in J_i$ and there is no $Q \not\Rightarrow \psi \in J_i$ such that $Q \cup \{\psi\} \subset P \cup \{\varphi\}\}$;
for each $P \not\Rightarrow \varphi \in$ *min_out_just* and for each $Q \not\Rightarrow \psi \in$ *min_out_just*:
$P \not\Rightarrow \varphi > Q \not\Rightarrow \psi$ if and only if $\varphi \in Q$ and $\psi \notin P$;
delta := *prem*;

repeat
  $P \not\Rightarrow \varphi \in max(min\_out\_just)$;
  $min\_out\_just := min\_out\_just - \{P \not\Rightarrow \varphi\}$;
  for every substitution θ such that $P[\theta] \subseteq delta$
  do $delta := delta/\varphi[\theta]$;
until $min\_out\_just = \emptyset$;
$out := prem - delta$;
return $delta, out$;
end.

The algorithm determines a set *delta* and a set *out* such that for some $\Delta \in \mathcal{A}_i$:

$$\Delta = \{\varphi \in \overline{\Sigma} \mid \varphi \text{ is an instance of some } \psi \in delta \text{ and if } \varphi \text{ is an instance of some } \mu \in out, \text{ then } \psi \text{ is an instance of } \mu\}.$$

It is not difficult to modify the algorithm in such a way that it determines every element of $\mathcal{A}_i$. Since the cardinality of $\mathcal{A}_i$ can grow exponential by the number of out_justifications, as shown in the following example, the determination of all the sets in $\mathcal{A}_i$ will be of exponential time complexity.

Example 2.26 Let $\{\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_n\}$ be a set containing $2n$ premisses. Furthermore, let each pair of premisses α_i and β_i be in conflict with each other. Then we can derive the following set of out_justifications.

$$\{\{\alpha_1\} \not\Rightarrow \beta_1, \{\beta_1\} \not\Rightarrow \alpha_1, ..., \{\alpha_n\} \not\Rightarrow \beta_n, \{\beta_n\} \not\Rightarrow \alpha_n\}$$

Clearly we can create 2^n different consistent subsets of the set of premisses by choosing either α_i or β_i.

Instead of determining every $\Delta \in \mathcal{A}_i$, we can try to determine $\bigcap \mathcal{A}_i$. This set will contain all premisses about which we do not have any doubt that we can believe them. If we can determine this set efficiently, we can approximate B_i with the set:

$$\{\varphi[\theta] \mid \text{there is a } P \Rightarrow \varphi \in J_i \text{ such that for some substitution } \theta\text{: } P[\theta] \in \bigcap \mathcal{A}_i\} \subseteq B_i$$

Unfortunately, it was proven by K. O. ten Bosch that the determination of $\bigcap \mathcal{A}_i$ is NP-Hard in the propositional case [4]. Ten Bosch proved this by showing that the decision problem 'is the premiss φ an element of $\bigcap \mathcal{A}_i$' is co-NP-Complete. I will give a reformulation of his proof below. In the proof I will limit the preference logic to a propositional logic. Furthermore, a labelling function $l : prem \rightarrow \{in, out\}$ is introduced. This function describes which premisses of the set of premisses *prem* are in a set $\Delta \in \mathcal{A}_i$.

Definition 2.27 Let *prem* be a set of premisses and let J be a set of out justifications. Furthermore, let $l : prem \rightarrow \{in, out\}$ be a labelling function.

l is a valid labelling of *prem* with respect to J if and only if for every $\varphi \in prem$:

$l(\varphi) = out$ if and only if there exists a $P \not\vdash \varphi$ such that for every $\psi \in P$:

$l(\psi) = in$.

Observation 2.28

$$\mathcal{A}_i = \{\{\varphi \mid l(\varphi) = in\} \mid l \text{ is a valid labelling with respect to } J_i\}.$$

Now we can formulate the decision problem as follows.

NAME: Belief Intersection (BI)
INSTANCE: A set of propositional premisses *prem*, a set of out_justifications J and a premiss $\varphi \in prem$.
QUESTION: Is $l(\varphi) = in$ for every labelling l that is valid given the out_justifications J.

I will prove that this decision problem co-NP-Complete by proving that the complementary problem is NP-Complete.

NAME: co-BI
INSTANCE: A set of propositional premisses *prem*, a set of out_justifications J and a premiss $\varphi \in prem$.
QUESTION: Is $l(\varphi) = out$ for some labelling l that is valid given the out_justifications J.

The proof that this problem is NP-Complete, is based on a reduction of the problem SAT to co-BI. For this reduction the following transformation is used.

Let $P = \{p_1, ..., p_m\}$ be a set of propositional variables.
Furthermore, let $C = \{c_1, ..., c_n\}$ be a set of clauses where $c_i = \{e_{i,1}, ..., e_{i,\ell_i}\}$ and for every $j \in \{1, ..., \ell_i\}$ there holds either that $e_{i,j} = p_k$ or that $e_{i,j} = \overline{p}_k$ for some $p_k \in P$.

The sets *prem*, J and $\langle prem, \prec \rangle$ are constructed as follows.

- $prem = U \cap W \cap \{\varphi, \psi\}$ where
  - $U = \{u_1, ..., u_m\}$, and
  - $W = \{w_1, ..., w_m\}$.
- $J = J_1 \cap J_2 \cap \{\{\psi\} \not\vdash \varphi\}$ where
  - $J_1 = \{\{u_1\} \not\vdash w_1, \{w_1\} \not\vdash u_1, ...\{u_m\} \not\vdash w_m, \{w_m\} \not\vdash u_m\}$,
  - $J_2 = \{A_1 \not\vdash \psi, ..., A_n \not\vdash \psi\}$ where $A_i = \{a_{i,1}, ..., a_{i,\ell_i}\}$ and
    - $a_{i,j} = w_k$ if and only if $e_{i,j} = p_k$,

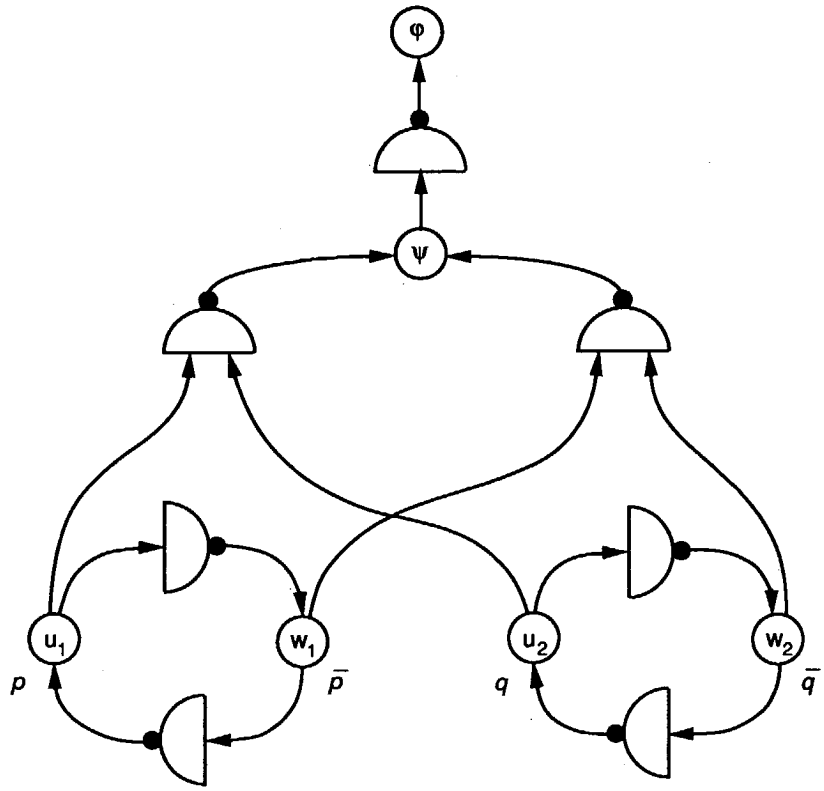


Figure 2.1: The reduction of $\{\{p, q\}, \{\overline{p}, \overline{q}\}\}$

  * $a_{i,j} = u_k$ if and only if $e_{i,j} = \overline{p}_k$.
- $\langle prem, \prec \rangle = \{(u_i, \psi), (w_i, \psi) \mid 1 \leqslant i \leqslant m\} \cup \{(\psi, \varphi)\}$

Clearly, since for every $Q \not\vdash \alpha \in J$ and for every $\beta \in Q$ there holds: $\beta \not\prec \alpha$, J is a set of out_justifications that can be generated by a set of premisses.
Now, let $t : P \rightarrow \{true, false\}$ be a truth assignment.
Then a corresponding labelling $l : prem \rightarrow \{in, out\}$ is defined as:

- $l(u_i) = in$ and $l(w_i) = out$ if and only if $t(p_i) = true$,
- $l(u_i) = out$ and $l(w_i) = in$ if and only if $t(p_i) = false$.

An illustration of this reduction can be found in figure 2.1. In this figure the result of reducing $\{\{p, q\}, \{\overline{p}, \overline{q}\}\}$ is shown.

Given the construction, the following lemma can be proven.

Lemma 2.29 There exists a valid labelling with respect to J such that:

$l(\varphi) = out$ if and only if the corresponding truth assignment t satisfies C.

Proof

$\Leftarrow$ Let t be a truth assignment such that C is satisfied and let l be the corresponding labelling.
We have to prove that l is a valid labelling and that φ is labelled *out*.
Clearly, l is valid with respect to J_1.
Since C is satisfied, in every c_i there must be an $e_{i,j}$ whose truth value is *true*.

- Suppose that $e_{i,j} = p_k$.
  Then $t(p_k) = true$, $l(u_k) = in$ and $l(w_k) = out$.
  Furthermore, since $e_{i,j} = p_k$, $w_k \in A_i$ and $l(w_k) = out$.
- Suppose that $e_{i,j} = \overline{p}_k$. Then $t(p_k) = false$, $l(u_k) = out$ and $l(w_k) = in$.
  Furthermore, since $e_{i,j} = \overline{p}_k$, $u_k \in A_i$ and $l(u_k) = out$.

Hence, for every c_i there is an $a_{i,j} \in A_i$ such that: $l(a_{i,j}) = out$.
So, there is no out-justification that forces ψ to be labelled with *out*.
Therefore, $l(\psi) = in$ and $l(\varphi) = out$ and l is valid with respect to J.

$\Rightarrow$ Let l be a valid labelling such that $l(\varphi) = out$.
We have to prove that the corresponding truth assignment t satisfies C.
Since $l(\varphi) = out$, $l(\psi) = in$.
Since $l(\psi) = in$, for every $A_i \not\Rightarrow \varphi$ there exists an $a_{i,j} \in A_i$ and $l(a_{i,j}) = out$.

- Suppose that $e_{i,j} = p_k$.
  Then $a_{i,j} = w_k$, $l(u_k) = in$ and $l(w_k) = out$.
  Furthermore, since $e_{i,j} = p_k$, $t(p_k) = true$ and c_i is satisfied.
- Suppose that $e_{i,j} = \overline{p}_k$.
  Then $a_{i,j} = u_k$, $l(u_k) = out$ and $l(w_k) = in$.
  Furthermore, since $e_{i,j} = \overline{p}_k$, $t(p_k) = false$ and c_i is satisfied.

Hence, every c_i is satisfied.
Therefore, C is satisfied.

□

Theorem 2.30 co-BI is an element of the class of NP-Complete problems.

Proof Clearly, given some labelling l and a proposition φ, it can be verified in polynomial time whether or not l is a valid labelling such that $l(\varphi) = out$.
Therefore, co-BI in NP.

By Lemma 2.29 there exists a polynomial transformation from SAT to co-BI.
Therefore, co-BI in NPC. □

Since, co-BI in NPC, BI in co-NPC. Hence, the problem of determining $\bigcap A_i$ must be NP-Hard.

2.5 The semantics for the logic

The semantics of the preference logic is based on the ideas of Y. Shoham [58, 59]. In [58, 59] Shoham argues that the difference between monotonic logic and non-monotonic logic is a difference in the definition of the entailment relation. In a monotonic logic a formula is entailed by the premisses if it is true in every model for the premisses. In a non-monotonic logic, however, a formula is entailed by the premisses if it is preferentially entailed by a set of premisses; i.e. if it is true in every preferred model for the premisses. These preferred models are determined by defining an acyclic partial preference order on the models.

The semantics for the preference logic differs slightly from Shoham's approach. Since the set of premisses may be inconsistent, the set of models for these premisses can be empty. Therefore, instead of defining a preference relation on the models of the premisses, a partial preference relation on the set of semantical structures for the language is defined. Given such a preference relation on the structures, the models for the premisses are the most preferred semantical structures. Hence, an appropriate preference relation on the structures has to be defined. This preference relation is based on the following ideas.

- The premisses are assumptions about the world we are reasoning about.
- We are more willing to give up believing a premiss with a low preference than a premiss with a high preference.

Therefore, a structure satisfying more premisses with a higher preference ($\prec$) than some other structure, is preferred ($\sqsubset$) to this structure.

In the preference logic we have to choose between premisses in case a minimal inconsistent subset of $\overline{\Sigma}$ does not contain a least preferred element. Choosing some premiss can be viewed as preferring the alternative choices to this premiss. So the original preference relation is extended by making choices. In case a premiss containing free variables is chosen, this choice is made for every instance of this premiss. Hence, the extension of the preference relation belonging to this choice, should also satisfy Condition 2.3. Now a structure satisfies more premisses than some other structure if this is the case for every linear extension of $(\overline{\Sigma}, \prec)$ which satisfies Condition 2.3. The following definitions describe this formally.

Definition 2.31 A semantical structure (interpretation) is a tuple $\langle D, v \rangle$ where:

- D is a domain of objects, and
- v is a valuation function that assigns objects to constants, functions to function symbols and relations to predicate symbols.

The set of all semantical structures is denoted by Str.

Definition 2.32 Let $\mathcal{M}$ be a semantical structure and let $\overline{\Sigma}$ be an extended set of premisses.
Then the premisses $Prem(\mathcal{M}) \subseteq \overline{\Sigma}$ that are satisfied by $\mathcal{M}$, are defined as:

$$Prem(\mathcal{M}) = \{\varphi \mid \varphi \in \overline{\Sigma} \text{ and } \mathcal{M} \models \varphi\}$$

Definition 2.33 Let $\overline{\Sigma}$ be an extended set of premisses and let $(\overline{\Sigma}, \prec)$ be a preference relation on $\overline{\Sigma}$. Furthermore, let Str be the set of structures for the language L and $(Str, \sqsubset)$ be a preference relation on these structures.
For every structure $\mathcal{M}, \mathcal{N}$ there holds:

$\mathcal{N} \sqsubset \mathcal{M}$ if and only if $Prem(\mathcal{M}) \neq Prem(\mathcal{N})$ and for every linear extension of $(\overline{\Sigma}, \prec)$ satisfying Condition 2.3 and for every $\varphi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$, there is a $\psi \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$ such that:

$$\varphi \prec \psi$$

and for no $\eta \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$:

$$\psi \prec \eta.$$

Given the preference relation between the structures, the set of models for the premisses can be defined.

Definition 2.34 Let $\overline{\Sigma}$ be an extended set of premisses and let $Mod_{\sqsubset}(\overline{\Sigma})$ denote the models for the premisses $\overline{\Sigma}$.

$\mathcal{M} \in Mod_{\sqsubset}(\overline{\Sigma})$ if and only if there exists no structure $\mathcal{N}$ such that:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

Now the following important theorem, guaranteeing the soundness and the completeness of the preference logic, holds:

Theorem 2.35 Let $\mathcal{M}$ be a partial model, let $\mathcal{A}_\infty$ be the corresponding set of consistent sets of believed premisses and let B_∞ be the corresponding belief set.
Then:

$$Mod_{\sqsubset}(\overline{\Sigma}) = \bigcup_{\Delta \in \mathcal{A}_\infty} Mod(\Delta) = Mod(B_\infty)$$

where $Mod(S)$ denotes the set of classical models for a set of formulas S.

Proof From Property 2.25 follows immediately:

$$\bigcup_{\Delta \in \mathcal{A}_\infty} Mod(\Delta) = Mod(B_\infty)$$

The proof of

$$Mod_{\sqsubset}(\overline{\Sigma}) = \bigcup_{\Delta \in \mathcal{A}_\infty} Mod(\Delta)$$

can be divided into the proof of the soundness

$$\bigcup_{\Delta \in \mathcal{A}_\infty} Mod(\Delta) \subseteq Mod_{\sqsubset}(\overline{\Sigma})$$

and the proof of the completeness

$$Mod_{\sqsubset}(\overline{\Sigma}) \subseteq \bigcup_{\Delta \in \mathcal{A}_\infty} Mod(\Delta)$$

of the preference logic.

Completeness Suppose that for some $\Delta \in \mathcal{A}_\infty$ and some $\mathcal{M} \in Mod(\Delta)$:

$$\mathcal{M} \notin Mod_{\sqsubset}(\overline{\Sigma}).$$

Then there exists a structure $\mathcal{N}$:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

According to Proposition 2.23, since $Prem(\mathcal{M}) = \Delta$:

$$\Delta \not\subset Prem(\mathcal{N}).$$

Hence, there exists a $\varphi \in (\Delta - Prem(\mathcal{N}))$.
Now by Definition 2.33 for each linear extension of $(\overline{\Sigma}, \prec)$ there exists a $\psi \in (Prem(\mathcal{N}) - \Delta)$ and $\varphi \prec \psi$.
Since $\psi \notin \Delta$, there exists a $P \not\Rightarrow \eta \in J_\infty$ and for some substitution θ:

$$P[\theta] \subseteq \Delta \text{ and } \psi = \eta[\theta].$$

Now, $P[\theta] \not\subseteq Prem(\mathcal{N})$, otherwise $Prem(\mathcal{N})$ would be inconsistent.
Hence, there exists a $\mu \in P[\theta]$:

$$\mu \in (\Delta - Prem(\mathcal{N})) \text{ and } \mu \not\prec \psi.$$

Therefore, there exists a linear extension of $(\overline{\Sigma}, \prec)$ such that:

$$\varphi \prec \psi \prec \mu.$$

Contradiction.

Hence,

$$\bigcup_{\Delta \in \mathcal{A}_\infty} Mod(\Delta) \subseteq Mod_{\sqsubset}(\overline{\Sigma}).$$

Soundness Suppose there exists a structure $\mathcal{M} \in Mod_{\sqsubset}(\overline{\Sigma})$ such that:

$$Prem(\mathcal{M}) \neq \overline{\Sigma} - Out_\infty(Prem(\mathcal{M})).$$

Then there exists a φ such that either:

- $\varphi \in Prem(\mathcal{M})$ and $\varphi \notin \overline{\Sigma} - Out_\infty(Prem(\mathcal{M}))$, or:
- $\varphi \notin Prem(\mathcal{M})$ and $\varphi \in \overline{\Sigma} - Out_\infty(Prem(\mathcal{M}))$.

Suppose that $\varphi \in Prem(\mathcal{M})$ and $\varphi \notin \overline{\Sigma} - Out_\infty(Prem(\mathcal{M}))$.
Hence, there exists a $P \not\Rightarrow \psi \in J_\infty$ and for some substitution θ:

$$P[\theta] \subseteq Prem(\mathcal{M}) \text{ and } \varphi = \psi[\theta].$$

Because $P[\theta] \subseteq Prem(\mathcal{M})$, $Prem(\mathcal{M})$ is inconsistent.
Contradiction.

Hence,

$$Prem(\mathcal{M}) \subseteq \overline{\Sigma} - Out_\infty(Prem(\mathcal{M})).$$

Suppose $\varphi \notin Prem(\mathcal{M})$ and $\varphi \in \overline{\Sigma} - Out_\infty(Prem(\mathcal{M}))$.
Then $Prem(\mathcal{M}) \cup \{\varphi\}$ is either consistent or inconsistent.
If it is consistent, then for each structure $\mathcal{N} \in Mod(Prem(\mathcal{M}) \cup \{\varphi\})$:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

Contradiction.

Hence $Prem(\mathcal{M}) \cup \{\varphi\}$ is inconsistent.
Therefore, there exists at least one minimal inconsistent subset of $Prem(\mathcal{M}) \cup \{\varphi\}$.
Let P be such a minimal inconsistent subset.
Now suppose that $\varphi \in min(P)$.
Then by Theorem 2.16 there exists an $R \not\Rightarrow \psi$ and for some substitution θ:

$$P/\varphi = R[\theta] \text{ and } \varphi = \psi[\theta].$$

Since $R[\theta] \subseteq Prem(\mathcal{M})$, $\varphi \notin \overline{\Sigma} - Out_\infty(Prem(\mathcal{M}))$.

Hence for each minimal inconsistent subset P:

$$\varphi \notin min(P).$$

Let MIN be the union of all the sets $min(P)$ for each minimal inconsistent subset P of $Prem(\mathcal{M}) \cup \{\varphi\}$.
For each $\eta \in MIN$ there holds:

$$\eta \prec \varphi.$$

Clearly, the set $(Prem(\mathcal{M}) \cup \{\varphi\}) - MIN$ is consistent.
Let $\mathcal{N} \in Mod((Prem(\mathcal{M}) \cup \{\varphi\}) - MIN)$.
Because for each $\eta \in (Prem(\mathcal{M}) - Prem(\mathcal{N})$:

$$\eta \prec \varphi,$$

and because $\varphi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$ there holds:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

Contradiction.

Hence,

$$\overline{\Sigma} - Out_\infty(Prem(\mathcal{M})) \subseteq Prem(\mathcal{M}).$$

□

2.6 Some properties of the logic

In this section I will discuss some properties of the logic. Firstly, I will relate the logic to the general framework for non-monotonic logics described by S. Kraus, D. Lehmann and M. Magidor [33]. Secondly, I will compare the behaviour of the logic when new information is added with Gärdenfors's theory for belief revision [23].

2.6.1 Preferential models and cumulative logics

In [33] Kraus et al. describe a general framework for the study of non-monotonic logics. They distinguish five general logical systems and show how each of them can be characterized by the properties of the consequence relation. Furthermore, for each consequence relation a different class of models is defined. The consequence relations and the classes of models are related to each other by representation theorems.

The consequence relation relevant for the preference logic is the preferential consequence relation of system **P**. I will show that the preference relation on the semantic structures, described in the previous section, corresponds with a preferential model described by Kraus et al.

Lemma 2.36 Let Σ be a set of premisses and let $(\Sigma, \prec)$ be a preference relation on the premisses. Furthermore, let $\widehat{\alpha} = \{\mathcal{M} \mid \mathcal{M} \models \alpha\}$, let $\Sigma' = \Sigma \cup \{\alpha\}$ and let $(\Sigma', \prec') = (\Sigma/\alpha, \prec) \cup \{\langle \varphi, \alpha \rangle \mid \varphi \in \Sigma/\alpha\}$.

Then $\mathcal{M} \in Mod_{\sqsubset'}(\overline{\Sigma'})$ if and only if $\mathcal{M} \in \widehat{\alpha}$ and for no $\mathcal{N} \in \widehat{\alpha}$:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

Proof

- Suppose that $\mathcal{M} \in \widehat{\alpha}$ and $\mathcal{N} \notin \widehat{\alpha}$, i.e. $\mathcal{M} \models \alpha$ and $\mathcal{N} \not\models \alpha$.
  Then by Definition 2.32:

  $$Prem(\mathcal{M}) \neq Prem(\mathcal{N}).$$

  Therefore,

  $$\alpha \in (Prem(\mathcal{M}) - Prem(\mathcal{N})),$$

  for each $\varphi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$ there holds:

  $$\varphi \prec' \alpha,$$

  and for no $\eta \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$ there holds:

  $$\alpha \prec' \eta.$$

  Hence by Definition 2.33 for each $\mathcal{M} \in \widehat{\alpha}$ and $\mathcal{N} \notin \widehat{\alpha}$:

  $$\mathcal{N} \sqsubset' \mathcal{M}.$$

- Suppose that $\mathcal{M}, \mathcal{N} \in \widehat{\alpha}$.
  Since $\mathcal{M}, \mathcal{N} \models \alpha$, for each $\varphi \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$ and for each $\psi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$:

- $\varphi \prec \psi$ if and only if $\varphi \prec' \psi$, and
- $\varphi \prec \psi$ if and only if $\varphi \prec' \psi$.

Hence, for each $\mathcal{M}, \mathcal{N} \in \widehat{\alpha}$:

$$\mathcal{N} \sqsubset' \mathcal{M} \text{ if and only if } \mathcal{N} \sqsubset \mathcal{M}.$$

Hence, $\mathcal{M} \in Mod_{\sqsubset'}(\overline{\Sigma}')$ if and only if $\mathcal{M} \in \widehat{\alpha}$ and for no $\mathcal{N} \in \widehat{\alpha}$:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

□

Theorem 2.37 Let Σ be a set of premisses and let $(\Sigma, \prec)$ be a preference relation on the premisses.

$\langle S, l, < \rangle$ is a preferential model for $\Sigma, (\Sigma, \prec)$ if and only if $S = Str$, $l : S \to S$ is the identity function and for each $\mathcal{M}, \mathcal{N} \in S$:

$$\mathcal{M} < \mathcal{N} \text{ if and only if } \mathcal{N} \sqsubset \mathcal{M}.$$

Proof Since the relation $\sqsubset$ defines a strict partial order on Str, so does $<$ on S. Since l is a function from Str to Str, l assigns a single 'world' to each state.

Suppose that $<$ is not smooth.
Then by Lemma 2.36 for some formula α and some $\mathcal{M} \in \widehat{\alpha}$ there exists no $\mathcal{N} \in Mod_{\sqsubset'}(\overline{\Sigma}')$ such that:

$$\mathcal{M} \sqsubseteq \mathcal{N}.$$

So, by Definition 2.33 for each $\mathcal{N} \in Mod_{\sqsubset'}(\overline{\Sigma}')$ there exists a linear extension of $(\Sigma, \prec)$ and there exists a most preferred $\varphi \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$ such that for no $\psi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$:

$$\varphi \prec \psi.$$

Since $\mathcal{N} \in Mod_{\sqsubset'}(\overline{\Sigma}')$,

$$\{\varphi\} \cup Prem(\mathcal{N}) \text{ is inconsistent.}$$

Let Γ be a minimal inconsistent subset of $\{\varphi\} \cup Prem(\mathcal{N})$.
Clearly, $\varphi \in \Gamma$.
Furthermore, since $\mathcal{N} \in Mod_{\sqsubset'}(\overline{\Sigma}')$,

$$\varphi \in min(\{\varphi\} \cup Prem(\mathcal{N})).$$

Because we consider a linear extension of $(\overline{\Sigma, \prec})$, there holds:

for each $\psi \in \Gamma$:

$$\varphi \prec \psi.$$

Contradiction.
Hence, $<$ is smooth.

Hence, $\langle S, l, < \rangle$ is a preferential model according to the definition of Kraus et al. □

Now I will relate the consequence relation of system **P** to the preference logic. To motivate the relation I will describe below, recall that $\alpha \mathrel{|\!\sim} \beta$ should be interpreted as: 'if α, normally β'. Hence, if we assume α, we must assume that α is true beyond any doubt. To realize this, we must add α as a premiss and prefer it to every other premiss, otherwise we cannot guarantee that α is an element of the belief set B_∞. If α is indeed an element of B_∞, we have to prove that β will also be an element of B_∞.

Theorem 2.38 Let $W = \langle S, l, < \rangle$ be a preferential model for $\Sigma, (\Sigma, \prec)$. Then the following equivalence holds:

$\alpha \mathrel{|\!\sim_W} \beta$ if and only if

$$\Sigma' = \Sigma \cup \{\alpha\},$$

$$(\Sigma', \prec') = (\Sigma/\alpha, \prec) \cup \{\langle \varphi, \alpha \rangle \mid \varphi \in \Sigma/\alpha\}$$

and $\beta \in B'_\infty$.

Proof According to Theorem 2.35:

$\beta \in B'_\infty$ if and only if for each $\mathcal{M} \in Mod_{\sqsubset'}(\overline{\Sigma'})$:

$$\mathcal{M} \models \beta.$$

Therefore, by Lemma 2.36:

$\beta \in B'_\infty$ if and only if for each $\mathcal{M} \in min(\widehat{\alpha})$:

$$\mathcal{M} \models \beta.$$

Hence, by Definition 1.36 we have:

$\beta \in B'_\infty$ if and only if $\alpha \mid\!\sim_W \beta$.

□

Corollary 2.39 Let $W = \langle S, l, < \rangle$ be a preferential model for $\Sigma, (\Sigma, \prec)$. Then:

$$B_\infty = \{\alpha \mid \mid\!\sim_W \alpha\}$$

2.6.2 Belief revision

In [23], Gärdenfors describes three different ways in which a belief set can be revised, viz. *expansion, revision* and *contraction*. Expansion is a simple change that follows from the addition of a new formula. Revision is a more complex form of adding a new formula. Here the belief set must be changed in such a way that the resulting belief set is consistent. Contraction is the change necessary to stop believing some formula. For each of these forms of belief revision, Gärdenfors has formulated a set of *rationality postulates*.

In this subsection I will investigate which of the postulates are satisfied by the preference logic. To be able to do this, the set B_∞ is identified as a belief set. Here expansion, revision and contraction of this belief set with respect to the formula α will be denoted by respectively: $B^+_\infty[\alpha]$, $B^*_\infty[\alpha]$ and $B^-_\infty[\alpha]$.

Expansion

To expand a belief set with respect to a formula α, α should be added to the set of premisses that generate the belief set. Since the preference logic does not allow an inconsistent belief set, only if $\neg\alpha$ does not belong to the belief set, α can be added. Otherwise, the logic would start revising the belief set. Adding α to the set of premisses, however, is not sufficient to guarantee that α will belong to the new belief set. Take for example the following set of premisses and preference relation.

$$\Sigma = \{1 : \alpha \wedge \beta, 2 : \neg\alpha \wedge \beta, 3 : \alpha \wedge \neg\beta, 4 : \neg\alpha \wedge \beta\}$$

$$(\Sigma, \prec) = \{3 \prec 2, 4 \prec 1\}$$

Clearly, adding α to Σ does not result in believing α. Hence, the second postulate of expansion is not satisfied. To guarantee that α belongs to the new belief set, we have to prefer α to any other premiss. If, however, we prefer α to every other premiss in the example above, the third postulate for expansion will not be satisfied. Hence, expansion of a belief set is not possible in the preference logic. The reason for this is that the reasons for believing a formula in a belief set are not taken into account by the postulates for expansion. Because of this internal structure, revision instead of expansion takes place.

Revision

For revision of a belief set B_∞ with respect to a formula α, we have to add α as a premiss and prefer it to any other premiss. With this implementation of the revision process, some of the postulates for revision of the belief set with respect to α are satisfied. The postulates not being satisfied, relate revision to expansion. Expansion, however, is not defined for the preference logic.

Theorem 2.40 Let B_∞ be the belief set for the premisses Σ with preference relation $(\Sigma, \prec)$.
Suppose that $B^*_\infty[\alpha]$ is the belief set of $\Sigma \cup \{\alpha\}$ with preference relation:

$$(\Sigma \cup \{\alpha\}, \prec) = ((\Sigma, \prec) \restriction (\Sigma/\alpha \times \Sigma/\alpha)) \cup \{\langle \varphi, \alpha \rangle \mid \varphi \in \Sigma/\alpha\};$$

i.e. $B^*_\infty[\alpha] = \{\beta \mid \alpha \mid\sim_W \beta\}$ where W is a preferential model for $\Sigma \cup \{\alpha\}, (\Sigma \cup \{\alpha\}, \prec)$.
Then the following postulates are satisfied.

1. $B^*_\infty[\alpha]$ is a belief set.
2. $\alpha \in B^*_\infty[\alpha]$.
6. If $\vdash \alpha \leftrightarrow \beta$, then $B^*_\infty[\alpha] = B^*_\infty[\beta]$.

Proof

1. This follows from Property 2.25
2. Since $\alpha \mid\sim_W \alpha$ (reflexivity), $\alpha \in B_\infty[\alpha]$.
6. Since $\dfrac{\models \alpha \leftrightarrow \beta, \alpha \mid\sim_W \gamma}{\beta \mid\sim_W \gamma}$ (left logical equivalence), if $\vdash \alpha \leftrightarrow \beta$, then $B^*_\infty[\alpha] = B^*_\infty[\beta]$.

□

Contraction

It is not possible to realise contraction of a belief set in the preference logic in a straight forward way. To be able to contract a formula α from a belief set B_∞, we have to determine the premisses on which belief in this formula is based. This information can be found in the applicable in_justification that supports the formula α. When we have determined these premisses, we have to remove some of them. This can be done in two different ways. We can either add the following out_justifications to J_∞

$$\{P/\varphi \not\Rightarrow \varphi \mid P \Rightarrow \alpha \in J_\infty, \varphi \in \mathit{min}(P)\}$$

or we have to remove some premisses from Σ. Choosing the latter solution, there need not exist one unique new belief set not containing α. Because Gärdenfors assumes a unique new belief set, this solution cannot be compared with his rationality postulates.

The former solution, which requires a modification of the preference logic, does give us a unique new belief set. However, it can only be applied if J_∞ has been determined. Using a first order logic, this will never be possible. Given this solution, we can easily verify that only the most trivial postulates 1, 3, 4 and 6 are satisfied.

3

Related work

In this chapter I will discuss some related approaches.

3.1 Hypothetical reasoning

The preference logic is closely related to N. Rescher's approach to deal with inconsistent knowledge [51]. This comes, of course, not as a surprise, since the preference logic is based on the ideas described by Rescher. Therefore, it is possible to translate Rescher's approach into the preference logic. Rescher divides a set of premisses into a finite number of modal categories, $M_0, ..., M_m$. Here M_0 contains the premisses we never want to give up and M_m contains the premisses we prefer to give up, if we have to give up some premiss to restore consistency. This can be modelled in the preference logic by preferring every premiss of a modal category M_i to every premiss of M_j with $i < j$. Since Rescher only considers propositional logic, the set of PMMC subsets is equal to the set $\mathcal{A}_\infty$. Furthermore, the set of CS entailed formulas is equal to the set B_∞, and the set of CR entailed formulas is equal to the set $\bigcup_{\Delta \in \mathcal{A}_\infty} Th(\Delta)$.

In Chapter 4 I will describe a semantics that is based on the semantics of the

preference logic, for Brewka's preferred subtheories. Since the preferred subtheories are a direct generalization of Rescher's work, the semantics described can also be used as a semantics for Rescher's work.

3.2 A framework for default reasoning

The preference logic is also related to D. Poole's framework for default reasoning [48]. Poole introduces two sets of premisses, facts and hypotheses. The set of facts is always consistent and cannot be removed. The set of hypotheses, however, may be inconsistent. Furthermore, a hypothesis may contain free variables. Each hypothesis containing free variables denotes a set of *ground instances of the hypothesis.* From the hypothesis a maximal consistent subset has to be selected, which can explain, together with the facts, some closed formula.

This framework can be represented in the preference logic by preferring each fact to each hypothesis. Because in Poole's framework each hypothesis containing free variables represents a set of *ground instances* instead of a set of instances, we have to restrict each set $\Delta \in \mathcal{A}_\infty$ to formulas containing no free variables. The result will be equal to the set of maximal scenarios of Poole's framework.

Although Poole's framework can be expressed in the preference logic, the philosophies behind the two approaches are quite different. Poole's work is based on the idea that default reasoning is a process of selecting consistent sets of hypotheses, which can explain a set of observations. In the preference logic, however, a consistent set of preferred assumptions is determined from which conclusions are drawn. This set of preferred assumptions may change due to new information.

In Poole's framework constraints can be added to denote that some set of hypotheses may not be used as an explanation. These constraints express that some explanations are preferred to others. This is realized by making the latter explanations inconsistent through the addition of constraints.

Poole's framework without constraints can be modelled in the preference logic. Since in the preference logic a preference relation on the premisses generates a preference relation on consistent subsets of the premisses, we may wonder if the preference relation described by the constraints can be modelled in the preference logic. Unfortunately, the answer is 'no'. This is illustrated by the following example.

Example 3.1

Facts: φ, ψ.

Defaults: $\varphi \rightarrow \alpha, \varphi \rightarrow \neg\beta, \psi \rightarrow \neg\alpha, \psi \rightarrow \beta$.

Constraints: $\neg(\alpha \wedge \beta), \neg(\neg\alpha \wedge \neg\beta)$.

Without the constraints this theory has four different extensions. These extensions are the logical consequences of the following scenarios.

$$S_1 = \{\varphi, \psi, \varphi \rightarrow \alpha, \varphi \rightarrow \neg\beta\}$$

$$S_2 = \{\varphi, \psi, \psi \rightarrow \neg\alpha, \psi \rightarrow \beta\}$$

$$S_3 = \{\varphi, \psi, \varphi \rightarrow \alpha, \psi \rightarrow \beta\}$$

$$S_4 = \{\varphi, \psi, \varphi \rightarrow \neg\beta, \psi \rightarrow \neg\alpha\}$$

Only the first two scenarios are consistent with constraints. If this default theory has to be modelled in the preference logic, a preference relation has to be specified in such a way that $\{S_1, S_2\} = \mathcal{A}_\infty$. To determine the required preference relation on the hypotheses, combinations of two scenarios are considered. To assure that $S_1 \in \mathcal{A}_\infty$ and $S_3 \notin \mathcal{A}_\infty$, $\varphi \rightarrow \neg\beta$ has to be preferred to $\psi \rightarrow \beta$. To assure that $S_2 \in \mathcal{A}_\infty$ and $S_4 \notin \mathcal{A}_\infty$, $\psi \rightarrow \beta$ has to be preferred to $\varphi \rightarrow \neg\beta$. Hence, the preference relation would be reflexive, violating the requirement of irreflexivity in a strict partial order. This means that not every ordering of explanations in Poole's framework can be modelled, using the preference logic. Whether an ordering on the explanations that cannot be modelled, will make sense, is something that has to be investigated.

3.3 Preferred subtheories

In [6], G. Brewka describes a generalization of Poole's Framework. His generalization consists of defining a partial preference relation on the set of hypotheses. Furthermore, the set of facts are defined as the set of most preferred hypotheses. Following Rescher, Brewka determines preferred maximal consistent subsets of the set of hypotheses, and calls them *preferred subtheories*. The preferred subtheories correspond with the $\Delta \in \mathcal{A}_\infty$.

Brewka distinguishes between *weakly provable* and *strongly provable* formulas. The strongly provable formulas correspond with formulas that belong to B_∞. The weakly provable formulas correspond with formulas that follow from some $\Delta \in \mathcal{A}_\infty$.

Brewka does not define a semantics for his preferred subtheories. In Chapter 4 I will reformulate the semantics of section 2.5 to a semantics for Brewka's approach.

3.4 Default logic

The relation between the Default logic of R. Reiter [49] and the preference logic is only a moderated one. The default rules introduced by Reiter do not have something like a contraposition. In the preference logic a rule always has a contraposition. There is no way to block this. Therefore, only free defaults rules can be translated into the preference logic.

$$\frac{\mathsf{T} : \varphi(\overline{x})}{\varphi(\overline{x})}$$

This default rule can be represented by $\varphi(\overline{x})$ in the preference logic. Furthermore, for every premiss ψ:

$$\varphi(\overline{x}) \prec \psi.$$

In [3] P. Besnard showed that every extension of a normal default theory is a subset of an extension of a corresponding free default theory. This free default theory is created by replacing every normal default rule

$$\frac{\varphi(\overline{x}) \ : \ \psi(\overline{x})}{\psi(\overline{x})}$$

by

$$\frac{\mathsf{T} : \varphi(\overline{x}) \rightarrow \psi(\overline{x})}{\varphi(\overline{x}) \rightarrow \psi(\overline{x})}$$

Hence, every extension of a normal default theory is a subset of $Th(\Delta)$ for some $\Delta \in \mathcal{A}_\infty$. Here, $\mathcal{A}_\infty$ is the result of the corresponding theory in the preference logic.

3.5 Deriving new defaults

In the conditional logic of J. P. Delgrande [15] it is possible to derive new default rules from existing default rules, and from other information available. In the preference logic also new default rules can be derived. This is illustrated, by using an example of Delgrande [15].

premisses:

1. $Raven(x) \rightarrow Black(x)$
2. $Raven(x) \wedge Albino(x) \rightarrow \neg Black(x)$

preference relation: $1 \prec 2$

conclusion: $Raven(x) \rightarrow \neg Albino(x)$.

Here the second premiss is preferred to the first, because the first is more general than the second.

It is also possible to derive new defaults by using a transitive relation between premisses. From the premisses:

$$Bird(x) \rightarrow Can\_fly(x)$$

$$\forall x[Eagle(x) \rightarrow Bird(x)]$$

then the default 'eagles can fly' can be deduced.

$$Eagle(x) \rightarrow Can\_fly(x)$$

3.6 Interacting defaults

In the previous section the default 'eagles can fly' was derived as a result of a transitive relation between a default and an implication. The possibility to derive such a transitive relation is not always wanted. To avoid unwanted transitive relations among defaults and other implications, Reiter and Criscuolo [50] argued that beside normal defaults, semi normal defaults are required. A semi-normal default is a default with restrictions on its use. These restrictions make it possible to avoid unwanted transitive relations. In the preference logic we do not have something equivalent to a semi-normal default. We can, however, avoid the unwanted transitive relations by stating the implicit assumption described by the justifications of a default rule, explicit. How this is done, depends on the relation between the consequent and the justifications. Let

$$\frac{\alpha : \gamma, \beta_1, ..., \beta_n}{\gamma}$$

be a semi normal default rule. Then one of the following three situations may occur.

1. γ implies β_i. In this case we only have to choose the correct preference relation.
2. We implicitly assume that if γ, then β_i. In this case we should either state this implicit assumption explicit and add the correct preference relation or we should add β_i to the consequent of a rule, using a conjunction.

Example 3.2

- University students are normally adults.
- Adults are normally employed.

From these two sentences it can be concluded that university students are normally employed. One knows, however, that university students are normally unemployed. By adding this information with the correct preference relation, the unwanted transitive relation can be avoided.

premisses:

1. $Univ\_Stud(x) \rightarrow Adult(x)$
2. $Adult(x) \rightarrow Employed(x)$
3. $Univ\_Stud(x) \rightarrow \neg Employed(x)$

preference relations: $1 \succ 2$ and $3 \succ 2$

3.7 Circumscription

There is actually no relation between the preference logic and circumscription. The latter minimizes the relation for which a predicate is true, while the former selects

maximal consistent sets of premisses. Of course, in the preference logic we can try to get the same result as circumscription of an n-place predicate p by adding the premiss $\neg p(x_1, ..., x_n)$ and preferring every other premiss to it. The relation for which the predicate p is true, will only be minimal for those instances $\langle o_1, ..., o_n \rangle$ that can be denoted by a tuple of ground terms $\langle t_1, ..., t_n \rangle$. This is illustrated by the following example.

Example 3.3 Let L be a language with only one constant c and no functions. Furthermore, let $\mathcal{M} = \langle \{a, b\}, v \rangle$ be a semantic structure where $v(c) = a$ and $v(p) = \{b\}$. Clearly, given an empty set of premisses, $\mathcal{M}$ will not be a model for circumscription of p. It is, however, a model for $\Sigma = \{\neg p(x)\}$ in the preference logic, since the predicate p is false for any ground instance of the predicate that can be expressed in the language.

3.8 Inheritance networks

An area where the preference logic can be used, is the formalization of inheritance hierarchies with exceptions. As was argued by D. S. Touretzky [62], inheritance networks can be modelled by using only normal default rules and by defining a correct ordering on these default rules. The ordering Touretzky specifies, models his inferential distance algorithm [63]. As was shown by D. W. Etherington, all the facts returned by the inferential distance algorithm lay in a single extension of the corresponding default theory [20]. The ordering Touretzky specifies in [62], determines this extension. For the preference logic we can get a similar result.

The preference relation, specified in the following definition, is the preference relation which is required to model the inferential distance algorithm.

Definition 3.4 Let χ be a property of the class φ and ω be a property of the class ψ.
If φ is a subclass of ψ, then objects of the class φ are preferred to have the property χ to the property ω.

For each **premiss** $\varphi \rightarrow \chi, \psi \rightarrow \omega \in \Sigma$:
if $\varphi \rightarrow \psi \in B_\infty$, then $[\varphi \rightarrow \chi] \succ [\psi \rightarrow \omega]$

Using this preference relation, also relations which hold between two different inheritance hierarchies can be handled.

Example 3.5 Suppose that we know that royal elephants are elephants, that elephants do not like mice and that royal elephants like black mice.

1. $\forall x[\mathit{Royal\_Elephant}(x) \rightarrow \mathit{Elephant}(x)]$
2. $\mathit{Elephant}(x) \wedge \mathit{Mouse}(y) \rightarrow \neg \mathit{Like}(x, y)$
3. $\mathit{Royal\_Elephant}(x) \wedge \mathit{Mouse}(y) \wedge \mathit{White}(y) \rightarrow \mathit{Like}(x, y)$

Then the preference relation defined by Definition 3.4

$$2 \prec 3$$

let us conclude that if Clyde is a royal elephant and Micky is a black mouse, then Clyde likes Micky.

In [64] D. S. Touretzky, J. F. Horty and R. H. Thomason make a distinction between a sceptical and a credulous reasoner. A sceptical reasoner refuses to draw conclusions in ambiguous situations and a credulous reasoner tries to conclude as much as possible by generating multiple extensions. Translated into the preference logic, this means that B_∞ describes the belief set of a sceptical reasoner. A credulous reasoner is a reasoner that considers the deductive closure of every $\Delta \in \mathcal{A}_\infty$.

3.9 Truth maintenance systems

In preference logic justifications are introduced. Unlike the justification that used in the JTMS of J. Doyle [17] or the ATMS of J. de Kleer [31], the justifications in the preference logic are part of the logic. They follow directly from the requirement for a deduction process (section 2.1). The justifications are also different from the ones introduced by Doyle and de Kleer. In a(n) (A)TMS the justifications describe dependencies between formulas, while in the preference logic the in_justifications describe dependencies between formulas and premisses, and out_justifications describe dependencies among premisses. Therefore, the in_justifications of the preference logic can be compared with the labels in the ATMS [31]. Like a label, an in_justification describes from which premisses a formula is derived. The out_justifications have more or less the same function as the set **nogood** in ATMS. Like an element from the set **nogood**, the consequent and the antecedents of an out_justification may not be assumed to be true at the same time. Unlike an element of the set nogood, an out_justification describes which element has to be removed from the set of premisses (assumptions). Something like a justification containing non-monotonic antecedents, as used in Dolye's JTMS, does not occur in the preference logic.

Because in_justifications and labels are closely related, it is possible to describe an ATMS using a propositional preference logic. Let $\langle A, N, J \rangle$ be an ATMS where:

- A is a set of assumptions,
- N is a set of nodes, and
- J is a set of justifications.

We can model the ATMS in the preference logic using the following construction. Let $A \cup N$ be the set of propositions of the logic. Furthermore, let the set of premisses Σ be equal to $A \cup J$ where the justifications J are described by rules of the form:

$$p_1 \wedge \ldots \wedge p_n \rightarrow q.$$

Finally, let every justification be preferred to every assumption. Then the set $\mathcal{A}_\infty$ is equal to the set of maximal (under the inclusion relation) environments of an ATMS. Furthermore, the label for a node $n \in N$ is equal to the set:

$$\{P \mid P \Rightarrow n \in J_\infty \text{ and for no } Q \Rightarrow n \in J_\infty \colon Q \subset P\}.$$

The set of nogoods is equal to the set:

$$\{(P \cup \{p\}) \upharpoonright A \mid \begin{array}{l} P \not\Rightarrow p \in J_\infty \text{ and for no } Q \not\Rightarrow q \in J_\infty \colon \\ (Q \cup \{q\}) \upharpoonright A \subset (P \cup \{p\}) \upharpoonright A\}. \end{array}$$

For a practical implementation of the preference logic, an ATMS can be used to determine the in_justifications (the labels) and the out_justifications (nogoods). To determine the set $\mathcal{A}_i$ by using the out_justifications, a special monotonic TMS is needed. This special TMS must label all premisses for which we have a valid out_justification, *out*. A valid out_justification is a justification whose antecedents are labelled *in*. The other premisses for which we do not have a valid out_justification must be labelled *in*.

Although the deduction process of the preference logic differs from Goodwin's logical process theory [24], it is also based on his view that reasoning is a process of adopting new constraints on the current belief set. Every new constraint being adopted, causes a process of belief revision. In this way, as in Goodwin's logical process theory, an *inference finding process* is created.

3.10 The Yale shooting problem

In this section I will show how S. Hanks and D. McDermott's solution of the Yale shooting problem can be formulated in the preference logic. Since this solution cannot be formulated in some non-monotonic logics, it illustrates that the preference logic possesses more expressive power than these logics.

In [26] Hanks and McDermott described a temporal projection problem and showed that the non-monotonic logics they considered are too weak to model it. They specified their problem in a situation calculus, which I have reformulated for the preference logic.

premisses:

1. $\forall s[T(Loaded, Result(Load, s))]$
2. $\forall s[T(Loaded, s) \rightarrow T(Dead, Result(Shoot, s))]$
3. $\forall s[\neg(T(Alive, s) \wedge T(Dead, s))]$
4. $T(f, s) \rightarrow T(f, Result(e, s))$
5. $T(Alive, S_0)$

6. $S_1 = Result(Load, S_0)$
7. $S_2 = Result(Wait, S_1)$
8. $S_3 = Result(Shoot, S_2)$

preference relation: $4 \prec 1$, $4 \prec 2$, $4 \prec 3$, $4 \prec 5$, $4 \prec 6$, $4 \prec 7$ and $4 \prec 8$

From the premisses of the problem $T(Dead, S_3)$ and $T(Alive, S_3)$ can be derived, causing a contradiction. Because in both the deduction of $T(Alive, S_3)$ and $T(Dead, S_3)$ an instance of the same default 4 is used and because no preference relation between instances of the fourth premiss has been specified, we have to choose an instance that has to be removed. Hence $\mathcal{A}_\infty$ will contain two sets of premisses; one from which $T(Dead, S_3)$ and one from which $T(Alive, S_3)$ can be derived. About the same problem arises in some of the other non-monotonic reasoning logics. Hanks and McDermott suggested the following solution [26, page 393]. One should prefer the *chronological minimal* models. These are the models in which the normality assumptions are made in chronological order; i.e. those in which abnormality occurs as late as possible. To realize this solution in the preference logic, we must allow that a preference relation is specified on the instance of a formula containing free variables. With this extension of the logic, we can formulate Hanks and McDermott's solution of the shooting problem by using the following preference relation.

Definition 3.6 The new preference relation is the transitive closure of:

- $(\overline{\Sigma}, \prec)$
- Let $\varphi(f, e, s)$ denote $T(f, s) \rightarrow T(f, Result(e, s))$.
  For each pair of instances $\varphi(f, e, s)$ and $\varphi(f, e, Result(e, s))$:

  $$\varphi(f, e, s) \succ \varphi(f, e, Result(e, s)).$$

Using this preference relation, an abnormality will occur as late as possible. Hence, $T(Dead, S_3)$ will be an element of the belief set B_∞.

4

Preferred subtheories

In this chapter I will describe a semantics for the preferred subtheories of G. Brewka [6]. This semantics, based on the semantics developed in section 2.5, can also be used for Rescher's approach to deal with inconsistent knowledge [51].

The semantics

The preferred subtheories are based on an enumeration of the premisses instead of a set of out_justifications. Therefore, the semantics described here is less complicated than the semantics described in section 2.5. Nevertheless, the models for the set of strongly derivable formulas Δ are also equal to those semantical structures that satisfy more premisses with a high preference ($\prec$) than some other semantical structure.

Definition 4.1 Let S and R be two sets of hypotheses.

The set S dominates the set R, $R \ll S$, if and only if $R \neq S$ and for every $\varphi \in (R - S)$, there exists a $\psi \in (S - R)$ such that:

$$\varphi \prec \psi.$$

Definition 4.2 Let $\mathcal{M}$ be a semantical structure.
Then the set of hypotheses $Hyp(\mathcal{M}) \subseteq \overline{\Sigma}$ satisfied by $\mathcal{M}$ is defined as:

$$Hyp(\mathcal{M}) = \{\varphi \mid \varphi \in \overline{\Sigma} \text{ and } \mathcal{M} \models \varphi\}$$

Definition 4.3 Let Str be the set of structures for the language L and let $(Str, \sqsubset)$ be a preference relation on these structures.
For each structure $\mathcal{M}, \mathcal{N}$ there holds:

$$\mathcal{N} \sqsubset \mathcal{M} \text{ if and only if } Hyp(\mathcal{N}) \ll Hyp(\mathcal{M}).$$

Given the preference relation between the structures, the set of models for the hypotheses can be defined.

Definition 4.4 Let Σ be a set of hypotheses. Furthermore, let $Mod_{\sqsubset}(\Sigma)$ denote the models for the hypotheses Σ.

$$\mathcal{M} \in Mod_{\sqsubset}(\Sigma)$$

if and only if there exists no structure $\mathcal{N}$ such that:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

Now the following important theorem, which guarantees the soundness and completeness of the preference logic, holds:

Theorem 4.5 Let $S^1, ..., S^n$ be the preferred subtheories of the hypotheses Σ with preference relation $(\Sigma, \prec)$.

$$Mod_{\sqsubset}(\Sigma) = \bigcup_{i=1}^{n} Mod(S^i) = Mod(\Delta)$$

where $Mod(X)$ denotes the set of classical models for a set of formulas X.

Proof From the definition of strongly provability, it follows immediately:

$$\bigcup_{i=1}^{n} Mod(S^i) = Mod(\Delta)$$

The proof of $Mod_{\sqsubset}(\Sigma) = \bigcup_{i=1}^{n} Mod(S_i)$ can be divided into a proof of soundness and a proof of completeness.

Completeness Suppose that for some S^i and some $\mathcal{M} \in Mod(S^i)$:

$$\mathcal{M} \notin Mod_{\sqsubset}(\Sigma).$$

Then there exists a structure $\mathcal{N}$:

$$\mathcal{M} \sqsubset \mathcal{N}.$$

Let $R = Hyp(\mathcal{N})$.
Since $\mathcal{M} \in Mod(S^i)$, $S^i = Hyp(\mathcal{M})$.
According to Definition 4.3:

$$\mathcal{M} \sqsubset \mathcal{N} \text{ if and only if } S^i \ll R.$$

From Definition 1.24 it follows:

$$S^i \not\subset R.$$

Hence, there exists a $\varphi \in (S^i - R)$.
By Definition 1.24 there exists an enumeration $\sigma_1, \sigma_2, \ldots$ of $\overline{\Sigma}$, which corresponds with S^i.
Let j be the lowest index such that $\sigma_j \in (S^i - R)$.
Now by Definition 4.1, there exists a $\psi \in (R - S^i)$ such that:

$$\sigma_j \prec \psi.$$

Since $\psi \notin S^i$ there exists a k such that $\sigma_k = \psi$, and $S^i_{k-1} \cup \{\psi\}$ are inconsistent.
Hence there exists a $\mu \in S^i_{k-1}$ and $\mu \notin R$.
So, $\mu \in (S^i - R)$ and because $\mu \in S^i_{k-1}$, there exists an index ℓ:

$$\sigma_\ell = \mu \text{ and } \ell < k < j.$$

Contradiction.

Hence, $\bigcup_i Mod(S^i) \subseteq Mod_{\sqsubset}(\Sigma)$.

Soundness Let $\mathcal{M}$ be any structure of $Mod_{\sqsubset}(\Sigma)$ and let $R = Hyp(\mathcal{M})$.
Firstly, $R \subseteq S^i$ for some i, is proven.

Let $H_1 = \{\varphi \in R \mid \forall \psi \in R : \varphi \not\prec \psi\}$.
Clearly, there exists an enumeration of $\overline{\Sigma}$ such that:

$$H_1 = \{\sigma_1 \ldots \sigma_j\}$$

for some $j \geqslant 1$.

Hence, H_1 is a subset of some preferred subtheory.

Proceeding inductively, let there exist an enumeration of $\overline{\Sigma}$ such that:

$$H_\ell = \{\sigma_1 ... \sigma_j\}$$

for some j: $1 \leqslant j$.
Furthermore, let $H_{\ell+1} = \{\varphi \in R \mid \forall \psi \in (R - H_\ell) : \varphi \not\prec \psi\}$.
Since for every $\varphi \in (H_{\ell+1} - H_\ell)$, there exists a $\psi \in H_\ell$ such that:

$$\varphi \prec \psi,$$

there exists an enumeration of $\overline{\Sigma}$ such that:

$$H_{\ell+1} = \{\sigma_1 ... \sigma_k\}$$

for some k: $1 \leqslant j < k$.

Hence, there exists a preferred subtheory S^i such that: $R \subseteq S^i$.

Suppose that there exists a $\varphi \in (S^i - R)$.
Because S^i is consistent, $R \cup \{\varphi\}$ is consistent.
Let $\mathcal{N}$ be a structure in $Mod(R \cup \{\varphi\})$.
By Definition 4.1, since $R \subset Hyp(\mathcal{N})$, there holds:

$$R \ll Hyp(\mathcal{N}).$$

Since $R = Mod(\mathcal{M})$, $\mathcal{M} \sqsubset \mathcal{N}$.
Hence,

$$\mathcal{M} \notin Mod_{\sqsubset}(\Sigma).$$

Contradiction.

Hence, for every structure $\mathcal{M} \in Mod_{\sqsubset}(\Sigma)$, there exists a preferred subtheory S^i such that:

$$S^i = Hyp(\mathcal{M}).$$

Therefore,

$$Mod_{\sqsubset}(\Sigma) \subseteq \bigcup_i Mod(S^i).$$

□

5

Evaluation

It has turned out that it is indeed possible to view default reasoning as a special case of reasoning with inconsistent knowledge. Furthermore, a deduction process has been developed, based on the ideas of Goodwin. As a result, the deduction process can be viewed as a logical process theory that will approximate the set of theorems of the premisses in the limit. It has also been proven that the preference logic is a logic of system **P**. Therefore, according to Kraus et al., the logic satisfies all properties an ideal non-monotonic logic should satisfy. Less successful has been the attempt to relate the preference logic with Gärdenfors's postulates for changing belief sets. The reason for this is because Gärdenfors does not assume a *base belief set* from which the belief set is generated. Hence, the belief set Gärdenfors considers, does not have an internal structure.

Unfortunately, there are two problems with the preference logic. These two problems are: firstly, that the determination of a belief set is NP-Hard and secondly, that the reasoning process is not very intuitive. To start with the former, if in case of a conflict there does not exist a least preferred element in the set of premisses on which the conflict is based, the time complexity can get out of control. In such a case we must either be satisfied with choosing randomly a premiss to be

removed, or ignore the inconsistency. Removing all least preferred premisses from the inconsistent set, however, can result in a totally wrong belief set. Although the determination of the belief set is an NP-Hard problem, the determination of a single *extension* $\Delta \in \mathcal{A}_i$ can be realized in polynomial time. This is still better than the determination of an extension of Reiter's default logic or Moore's autoepistemic logic. For these logics the determination of a single extension is already an NP-Hard problem.

The latter problem of the preference logic is a problem that can be found to some degree in every non-monotonic logic. In my opinion this problem is inherent to the use of logic as a tool for knowledge representation. In an attempt to overcome this problem in the next part of this thesis, I will propose an alternative approach not based on logic.

Part II

A proposal for an alternative way of reasoning

6

In defence of partial models

When reasoning with less accurate knowledge by using some special logic, there usually arise three problems, viz. (1) the intractability of the reasoning process; (2) the in-correctness of the conclusions derived, and (3) the counter intuitive way of reasoning.

To start with the first problem, all logic based reasoning systems explore a search space in which they try to find a sound proof path to some desired conclusion. Even for monotonic logics searching through this search space can be very inefficient, unless knowledge may only be expressed by using a limited subset of the language. This latter approach is usually implemented in rule based systems.

Using some non-monotonic logic, things become even worse. Somehow every non-monotonic logic bases its conclusions on a consistency check. Since in first order logic the consistency problem is undecidable, it become intractable to find a proof for some desired conclusion. To deal with this problem, Goodwin introduces the concept of a *current proof* [24]. In Goodwin's view, a current proof for a formula should be interpreted as: given the inferences that are made up till now, a formula can or cannot be proven to hold. Therefore, the inferences being made, have to be registrated. They function as a set of constraints on the formulas that can

currently be believed, the current belief set. A Reason Maintenance System is used to determine this current belief set. Although this concept can be used to approximate the set of correct conclusions, finding a *current proof* for some desired conclusion does not imply that this is a sound proof. In the preference logic, for example, we have to verify for each (intermediate) conclusion derived whether it can be overruled by its negation. In other logics, like default logic, we have to verify for every justification used in a default rule whether the denial of this justification cannot be derived.

For every new constraint on the current belief set being derived, Reason Maintenance has to be applied. When these constraints consist of justifications in a JTMS, determination of a current belief set is NP-Hard.

The second problem is caused by reasoning with less accurate knowledge. Unlike monotonic logics, finding a current proof for some proposition in a non-monotonic logic does not imply that this is an ultimately sound proof. It is only sound if it is also a proof for that proposition in the deductive closure of the premisses, i.e. its proof is not cancelled by proving another proposition. Hence, to determine whether a conclusion is correct, all logical consequences have to be known. Unfortunately, we can only approximate the deductive closure of a set of premisses. Therefore, we can never be sure whether the formulas of the current belief set also belong to the set of theorems of the premisses.

A similar problem arises with the uncertain conclusions defined in Chapter 8. Unlike logics for reasoning *with* uncertain propositions, the certainty measure used here can be compared with R. Carnap's *logical* probability. The certainty measures defined here express our ignorance.

As I already mentioned in the first part of this thesis, the third problem seems to be inherent to the use of logic as a tool for knowledge representation. When we are reasoning, we usually are only interested in determining properties of some specific objects. When, for example, we have to choose between two roads both leading to the place we want to visit, we are only interested in facts like the length of the road, the maximum speed, the chance of running into a traffic jam, etc. So, what we are interested in, is constructing a model containing all relevant facts of the two roads. To construct such a model, we must be able to extract the relevant facts from the knowledge available. Deductive reasoning processes, however, only combine pieces of knowledge to get a new piece of knowledge. They are not specially designed to extract the relevant facts from the knowledge available. Therefore, they are not very intuitive.

To tackle these problems, I propose to model a reasoning process as a process of constructing a partial model of the world we are reasoning about. In the next section I will motivate why we can base a reasoning process on the construction of a partial model and how at least two of the three problems can be solved.

6.1 Partial models

In his paper *In defence of logic* [25], P. J. Hayes did not, in the first place, defend the use of logic as a knowledge representation tool. Although he is in favour of logic, in his opinion we may choose any form of knowledge representation that is convenient for our purpose, as long as it possesses a model based semantics. According to Hayes, it is only then that we are able to evaluate the situations that satisfy the represented knowledge. Consider, for example, a representation that states that coffee will still be in a cup after having moved the cup. It is not difficult to imagine a situation that does not satisfy this knowledge. But to be able to verify this, we must know how situations (models) are related to the knowledge representation; i.e. it must have a model based semantics.

So what are these situations or models that we use to evaluate the represented knowledge? They represent some abstract description of the world. These descriptions of the world are different from the knowledge described by the knowledge representation we use. The former gives a description of the world as we *observe* it. The latter, however, gives a description of what must or should hold in any model of the world; i.e. it describes what we will or can observe in any world we look at. Hence, a model is the connection between our knowledge and our observations. It represents what we observe in and what we know about some world. Clearly, since we can observe the world only partially and since our knowledge about the world is far from complete, we can only have a partial model of the world. This partial model is assumed to be a finite approximation of some complete model, which can only be possessed by some ideal omniscient agent. It consists of a finite number of objects and of partially defined relations on these objects.

Since a partial model is the connection between our knowledge and our observations, I propose to give it a central place in a reasoning process by viewing the reasoning process as a process of constructing a partial model of the world we are reasoning about. With this view on reasoning, like observations, a reasoning step causes an expansion or a revision of our partial model of the world. To realize this view, it is necessary to introduce a syntactic representation of a partial model which we can manipulate. This syntactic representation can be viewed as a *conceptual model* of its semantic equivalent. I will, however, address this syntactic representation as a partial model.

The idea of using a partial model is not new. It can also be found in other research areas.

- In discourse theory, partial models are used by H. Kamp [30]. In his Discourse Representation Theory (DRT) partial models are used to represent the interpretation of a discourse. New sentences are added by updating this interpretation.

- In cognitive psychology, partial models are used by P. N. Johnson-Laird [29] to explain syllogistic reasoning by humans. According to Johnson-Laird, humans

do not reason by using some internal logic, but they reason by constructing a partial model. He distinguishes two different kinds of partial models, viz. an image and a conceptual model. An image is a three dimensional reconstruction of a scene while a conceptual model is an abstract model of the world, like the one described here.

- In knowledge representation also something like a partial model has been proposed by D. W. Etherington, A. Bogida, R. J. Brachman and H. Kautz. [21]. They propose the use of a *vivid* knowledge base to keep the reasoning process tractable. In this knowledge base they only store positive facts that have a one to one correspondence with objects and relations in the world. Other forms of knowledge are stored in a traditional knowledge base, which has a traditional inference engine.

If we view reasoning as a process of constructing a partial model, how can this help us to avoid the three problems mentioned at the beginning of this chapter? The first problem I mentioned in the introduction is the intractability of the reasoning process. I believe the reasoning process I propose here to be more tractable than traditional ones. Clearly, the information contained in a partial model can be read in polynomial time. So, the time complexity of the reasoning process depends on the process used to construct (update) the partial model by adding new information. Information like 'John goes to the movie or to the theatre tonight', for example, can be added in polynomial time. However, when information like 'most humans have brown eyes' is added or when default rules are involved, the time complexity of the reasoning process becomes less clear.

Since the consistency problem is decidable in a partial model, unlike non-monotonic logics, correctness of default conclusions can be guaranteed; i.e. conclusions based on a partial model are always correct with respect to the partial model and with respect to the information used to construct the model. This implies that if some conclusion that follows from a partial model, is not correct, then either some incorrect piece of information is used to construct the model or some relevant piece of information has not yet been used in the construction of the partial model. Hence, *we can discuss about the information used or not used in the construction of the partial model.* In logic based reasoning systems, however, we do not know whether all relevant deduction steps are made or whether some piece of information is either not available or not correct. Here, it is not possible to discuss the information used or not used since the correct answer can be implied by the knowledge but not yet be derived.

Another advantage of using a partial model as a basis for a reasoning process is that in a partial model we can *count* objects. In a first order logic we can, of course, also describe that we have a certain number of distinct objects for which some property is satisfied. This description, however, is a very clumsy one because we cannot count in a first order logic. Because objects can be counted in a partial

model, we can update the model with information describing that for a percentage of a class of objects some proposition is satisfied.

6.2 Defining a partial model

The issue I will address in this section is how to define a partial model. When the information, used to update a partial model, is described by a propositional language, a partial model can be defined, for example, by using Kleene's strong three valued logic. When, however, information is described by a first order language as will be assumed here, defining a partial model becomes much more complex. Before defining a partial model, I will first look at the things we want to represent in it.

A partial model of the world should contain objects that represent *observable* entities, and relations that represent *observable* events or structures. The question is how these objects and relations can be represented in a partial model. In traditional logics a model consists of a set of objects and a valuation function that assigns an object to every constant symbol and a relation to every predicate symbol. A partial model can be described by a set of objects and two valuation functions; one describing the instances of a relation that are known to be true, and one describing the instances of a relation that are known to be false. I will, however, introduce an alternative representation to describe a partial model. This representation, which has been chosen because it is more convenient in other definitions, consists of two sets of instances of relation; one containing the instances that are known to be true and one containing the instances that are known to be false.

Definition 6.1 Let O be a set of objects. An instance of an n-place relation $r \in O$ is an n+1 tuple

$$\langle r, o_1, ..., o_n \rangle$$

where $o_i \in O$ for $i = 1, ..., n$.

Since I do not use a valuation function and since I treat relation symbols in the same way as other objects, I need a method to name objects. Here I have chosen to introduce a set of special objects called *names*, which have the same function as constants in a first order logic. These names, which can be used for objects and relations, are assumed to be unique. Objects that are no names are called *anonymous* objects.

Definition 6.2 Let $Names$ be a finite set of names and let $\{o_i \mid i \in \mathbf{N}\}$ be an enumerable set of anonymous objects, $Names \cap \{o_i \mid i \in \mathbf{N}\} = \emptyset$. Then the domain of objects D for a partial model is defined as:

$$D = Names \cup \{o_i \mid i \in \mathbf{N}\}$$

Using this domain of objects, a partial model can be defined.

Definition 6.3 Let D be the domain of objects for a partial model. Then, a partial model $\mathcal{M}$ is a tuple $\langle O, V \rangle$ where:

- $O \subseteq D$ is the set of known objects,
- $V = \langle R^+, R^- \rangle$ is a view on the world where R^+ and R^- are sets of instances of relations $r \in O$, R^+ denoting a set of instances known to be true, and R^- denoting a set of instances known to be false.

The partial model defined above is the most simple realization of a partial model for a first order language. I will now investigate whether this partial model possesses sufficient expressive power to represent the information we want to represent by a partial model. Clearly, information like 'Pedro beats a donkey' can be represented in a partial model by the relations $\langle beat, Pedro, x \rangle$ and $\langle donkey, x \rangle$. Information containing a disjunction like for example 'John is 30 or 31 years old', however, cannot be expressed in a single partial model. Since a disjunction expresses different views on the world, two partial models are needed to represent this information; one stating that John is 30 and one stating that John is 31. Now, suppose that we want to represent the information 'a friend of John goes to Paris by bus or by train' by a set of partial models. Then also two partial models are needed; one in which the friend of John goes to Paris by car and one in which he goes by train. In each of these two partial models, John's friend will be represented by an anonymous object. Since there is no relation between the two partial models, there is no reason why the objects denoting John's friend in both models should denote the same person. They can just as well denote two different friends of John. This is, however, counter intuitive.

A related problem will arise when the definition of the uncertainty of a conclusion, which is given in Chapter 8, is based on the partial model of Definition 6.3. What will go wrong in that case is illustrated by the following two examples.

1. Suppose that John's bowling ball and two other bowling balls lay on the table. If one of them is red, what is the probability that this is John's bowling ball.

2. Suppose that John's, Paul's and Peter's bowling ball lay on the table. If one of them is red, what is the probability that this is John's bowling ball.

The information described by the first example can be represented by two partial models, one in which John's ball is red and one in which it is not. To represent the information described by the second example , three partial models are needed. One partial model in which John's ball is red, one in which Peter's ball is red and one in which Paul's ball is red. Since the first example can be represented by two partial models, while for the second example three partial models are needed, assuming that these partial models are equally likely, the probability that John's ball is red will be

different in both examples. To avoid these problems, a new definition of a partial model is needed.

This new partial model is a *partial epistemic* model. It consists of a set of views the reasoning agent can have on the world. Like in an epistemic model, ignorance whether an object o has a property r or s can be modelled, using different views.

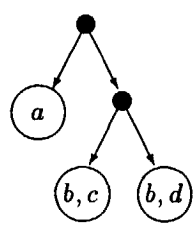
Definition 6.4 Let D be the domain of objects for a partial model. Then, a partial model $\mathcal{M}$ is a tuple $\langle O, V \rangle$ where:

- $O \subseteq D$ is the set of known objects,
- $V = \{V_1, ..., V_m\}$ is a set of views $V_i = \langle R^+, R^- \rangle$ where R^+ and R^- are sets of instances of relations $r \in O$, R^+ denoting a set of instances known to be true, and R^- denoting a set of instances known to be false.

Information expressed by a disjunction does not only express different views on the world, it also expresses our uncertainty about the world. For example, the disjunction $p = (a \vee b)$ expresses, in the absence of other information, two equally likely views on the world. If we represent this information in a partial model, two views are needed; one in which a holds and one in which b holds. Assuming that these views are equally likely, the partial model expresses the same uncertainty about the world.

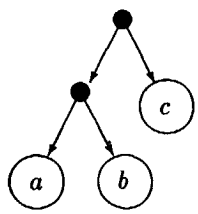
Now consider the situation in which we want to represent $q = (a \vee (b \wedge (c \vee d)))$ in a partial model. This disjunction describes that we are uncertain whether a or $(b \wedge (c \vee d))$ holds in the world we are reasoning about. If $(b \wedge (c \vee d))$ holds in the world, then we are uncertain whether $(b \wedge c)$ or $(b \wedge d)$ holds. These uncertainties cannot be expressed by the partial model defined above. Because q expresses three different views on the world, q can only be represented by a partial model containing three views. Assuming that all views are equally likely, this partial model does not express the uncertainty described by q. Therefore, again a new definition of a partial model is needed. In the new defined partial model, it should be possible to represent the uncertainty expressed by a formula.

One possible way of representing this uncertainty would be to use a *tree of views*. In such a tree the formula q will be represented by the root. This root must consist of two views; one representing the formula a and one representing the formulas b and $(c \vee d)$. Furthermore, the view representing $(c \vee d)$ must also consist of two views; one representing c, and one representing d.

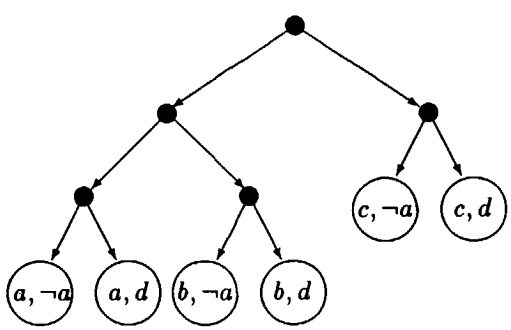


This way of defining a partial model has an important disadvantage. Representing the information of a single formula in such a partial model does not cause any problem. But representing the combined information of two or more formulas in a partial model does cause problems. To be able to represent this combined information, the uncertainties expressed by the formulas have to be combined. Because choices described by a disjunction can be eliminated by the information of other formulas, using a tree of views can result in an incorrect representation of the uncertainties expressed by the combined information.

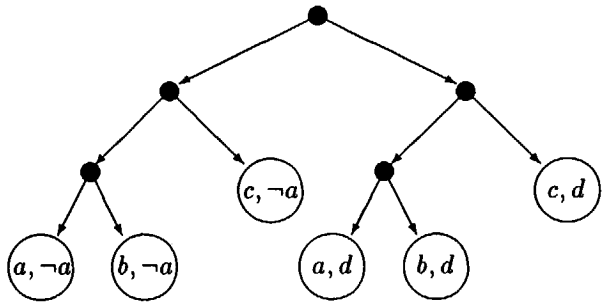
To illustrate this, consider an empty partial model to which we add the information $p = ((a \vee b) \vee c)$. This will result in the following tree of views.



Suppose that after adding p to the partial model, we add the information $q = (\neg a \vee d)$ to the partial model. Since this information describes two equally likely alternatives, this should also be expressed by the new partial model. If we would distribute this information over the information contained in the partial model, the new partial model will contain the following tree of views.

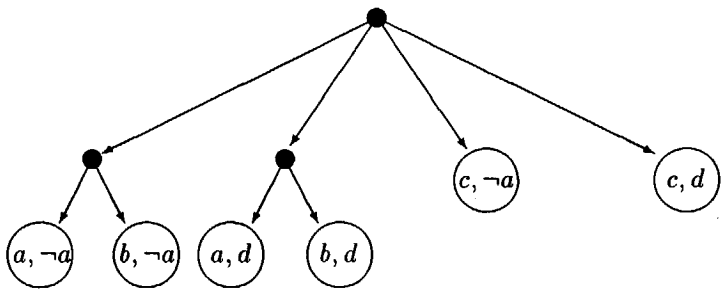


Since the view $\langle a, \neg a \rangle$ is inconsistent, it has to be eliminated from the tree. Then, the probability of the view $\langle a, \neg d \rangle$ will be equal to $\frac{1}{4}$. Now suppose that we first add q and after that p to the partial model. This will result in the following tree of views.

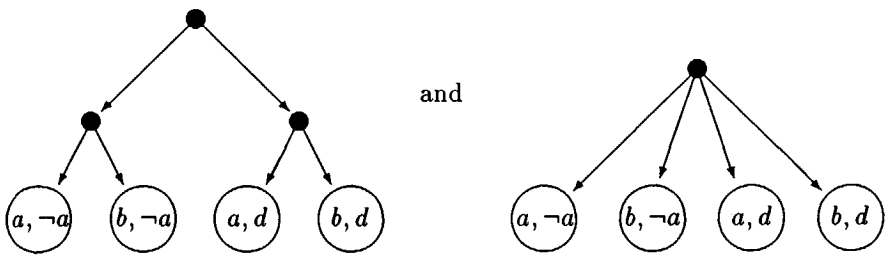


Here, the probability of the view $\langle a, \neg d \rangle$ is equal to $\frac{1}{8}$. Because this result is counter intuitive, the distribution of uncertainty is impossible.

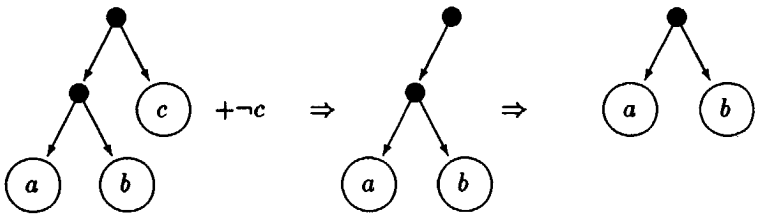
Another possibility is to determine the product of the choices described by the partial model and the new information. For the example described above, this results in:



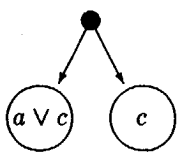
Now the order in which p and q are added to the partial model does not influence the result. But if we also added $\neg c$ to the partial model, again the result will depend on the order in which the information is added.



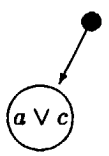
It should be clear that the right tree represents the uncertainty correctly. The reason why we can also get the incorrect partial model is because by adding $\neg c$, we eliminate a complete level of the tree of views.



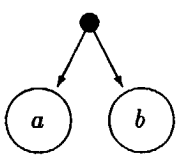
If the views in the partial model of Definition 6.4 may contain formulas instead of only instances of relations, the problem I have illustrated above can be avoided. Such a partial model can also be viewed as representing a tree of views. Here, however, the levels of the tree that do not have the root as a parent, are represented implicitly by the formulas. For example $((a \vee b) \vee c)$ is represented as:



If we add $\neg c$ to this 'tree' of views, we get:



Since the formula $(a \vee b)$ holds in every view of the partial model, we may expand it. So we get:



Another way of looking at the formulas in a view is to consider them to be constraints that have to be satisfied by this view on the world. Since formulas have not been defined yet, I will postpone the definition of the partial model till the formulas have been defined.

7

The reasoning process

In this chapter I will characterize the reasoning process needed to update a partial model with new information. Before I will describe the actual reasoning process, first I will define the information used to update a partial model. For this a language is needed in which the knowledge used can be represented. In this language it must be possible to represent quantification over a description of objects (a quantified description). By a quantified description I mean expressions like *all farmers* and *a donkey*. It is not possible to limit ourselves to quantifiers like *all* and *a*, when discussing the derivation of uncertain conclusions. Uncertainty can also arise by using quantifiers like *most* or *many*.

7.1 New information

In this section I will describe the representation for the information used to update a partial model. The main problem discussed here is how to represent a quantified description. Firstly, it is not possible to describe a class of objects directly. Secondly, in first order logic only universal quantification is possible. Extending first order logic by adding more quantifiers will not solve the problem. For example, the sentence

'Most apples are red' is not equivalent to $(most\ x)[apple(x) \rightarrow red(x)]$. Here the quantifier *most* ranges over the whole domain of objects instead of over the class of apples. Because of this, objects not being apples are also counted for the percentage *most*, for the implication also holds when the antecedent is false. We might, of course, extend the logic with a device for quantifying over classes of objects. For example: $(most\ x)[apple(x)] : red(x)$.

An alternative approach is the knowledge representation language OMEGA [27, 1]. This knowledge representation language was specially designed to reason with indefinite descriptions. Here, classes are represented explicitly. They are used to create the descriptions. Therefore, it is not difficult to extend the language such that quantification over classes becomes possible. Another extension which is needed, is the extension of the number of relations between objects. In OMEGA the only relation between objects is the relation *is*. With these extensions we are able to describe properties of objects of a class like the class of *persons*. Sometimes, however, we want to describe subclasses like *persons who own a driving licence*. Again we must extend OMEGA to make the description of subclasses possible.

In the partial models defined in section 6.2, we do not have classes. If a class of objects has to be represented in a partial model, a relation has to be used. We can, of course, also specify that the relation has a property that says it is a class. Since a class is represented by a relation in a partial model, we can also do this in OMEGA. But then we get more or less the extended predicate logic mentioned above.

In the following section I will introduce a representation based on the extended predicate logic discussed here.

7.2 Formal definitions

In this section I will define the formulas used to describe new information and to describe constraints used in a partial model. After that, the partial model introduced at the end of section 6.2 will be defined.

Before the formulas are defined, I will discuss the representation of a disjunction used here. Unlike the predicate logic, a formula representing a disjunction will not be a binary operator. If the operator $\vee$ would be a binary operator, the formula $a \vee b \vee c$ can be read in two different ways; either as $(a \vee b) \vee c$ or as $a \vee (b \vee c)$. As was argued in Section 6.2, both formulas represent two different choices. The first formula represents a choice between $(a \vee b)$ and c, while the second formula represents a choice between a and $(b \vee c)$. If, however, we want to describe three or more alternatives, n-place disjunction operators are needed.

Since the information described by a formula may refer to the objects in the current partial model, these objects have to be available. These objects to which we can refer, are called *external objects*. They bind the objects that occur free in a formula.

Definition 7.1 Let O_{ex} be the set of external objects for the formulas to be defined here.

The formulas are defined as follows.

- $\varphi = \langle r, o_1, ..., o_n \rangle$ is an n-place relation where $r \in O_{ex}$ is a relation symbol and $o_1, ..., o_n \in O_{ex}$ are the arguments of the relation.
- $\varphi = \langle \neg, \psi \rangle$ is the negation of the formula ψ with external objects, the set O_{ex}.
- $\varphi = \langle \wedge, \psi_1, ..., \psi_n \rangle$ with $n \geqslant 2$ is a conjunction of formulas ψ_i with external objects, the set O_{ex}.
- $\varphi = \langle \vee, \psi_1, ..., \psi_n \rangle$ with $n \geqslant 2$ is a disjunction of formulas ψ_i with external objects, the set O_{ex}.
- $\varphi = \langle \#, n, o, \psi(o) \rangle$ is a description of a group of objects where $n \in \mathbb{N}$ is a natural number, $o \in D$ is an object and $\psi = \psi(o)$ is a formula with external objects, the set $O_{ex} \cup \{o\}$.
- $\varphi = \langle \mathbf{a}, o, \psi_1(o), \psi_2(o) \rangle$ is an indefinite description where $o \in D$ is an object that occurs in both the formulas ψ_1 and ψ_2, and where $\psi_1(o) = \psi_1$ and $\psi_2(o) = \psi_2$. The set of external objects for the formulas ψ_1 and ψ_2 is the set $O_{ex} \cup \{o\}$.
- $\varphi = \langle \%, p, o, \psi_1(o), \psi_2(o) \rangle$ is a quantified description where the percentage $p \in [0, 1]$ is a rational number, $o \in D$ is an object that occurs in both the formulas ψ_1 and ψ_2, and where $\psi_1(o) = \psi_1$ and $\psi_2(o) = \psi_2$. The set of external objects for the formulas ψ_1 and ψ_2 is the set $O_{ex} \cup \{o\}$.

Now the formulas have been defined, the definition of the partial model can be given. This partial model consists of a set of objects and a set of views. Each view consists of a set of formulas that have to be satisfied by it.

Definition 7.2 Let D be the domain of objects for a partial model. Then, a partial model $\mathcal{M}$ is a tuple $\langle O, V \rangle$ where:

- $O \subseteq D$ is the set of known objects,
- $V = \{V_1, ..., V_m\}$ is a set of views V_i where V_i is a set of formulas that have to be satisfied by this view.

7.3 Semantics

In this section I will describe two semantics for the formulas. Before I do so, I will discuss the meaning of a description. A description does not denote a specific object. It only describes an object and, therefore, this object can be mapped on any object in the partial model that satisfies the description. Take, for example,

the sentence: 'most humans have brown eyes'. This sentence can be represented as: $\langle \%, p, o, \varphi_1(o), \varphi_2(o) \rangle$ where $\varphi_1(o) = \langle human, o \rangle$ and $\varphi_2 = \langle has\_brown\_eyes, o \rangle$. Here every possible mapping of the object o on the objects of the partial model has to be considered. Given these mappings, the number of different objects in the partial model on which o can be mapped such that respectively φ_1 and both φ_1 and φ_2 are satisfied, have to be counted. The ratio between the two numbers should be equal to p. For any world we can verify whether it satisfies such a description. If the cardinality of the class of humans is unknown in a partial model, any cardinality c is possible as long as $p \cdot c$ is a natural number. Since there are infinitely many different cardinalities that satisfy this requirement, it becomes impossible to describe in a finite partial model the result of updating a partial model with the information of a quantified description. Therefore, in the semantics below I will introduce the condition that the cardinality of a class of objects described by a partial model has to be known.

As mentioned above, two different semantics will be defined, a *strong* and a *weak* semantics. The weak semantics is intended for normal use, i.e. when one wishes to know whether a formula is true or false in a partial model. The strong semantics has a totally different purpose. This semantics is introduced to be able to characterize the reasoning process. It will be used to define the conditions under which a partial model has incorporated the information described by some formula.

Firstly, I will define the weak semantics. Two satisfiability relations will be used. Since we only have a *partial* model, formulas not satisfied by the model do not need to be *false*. Therefore, it is not sufficient to define a satisfiability relation for the formulas that are *true* ($\models^+$). We also need a satisfiability relation that defines when a formula is *false* ($\models^-$); i.e. a formula that cannot be satisfied by any consistent extension of the partial model.

Definition 7.3 Let φ be a formula and let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_1, ..., V_m\} \rangle$ be a partial model.

A formula φ is *true* in $\mathcal{M}$, $\mathcal{M} \models^+ \varphi$, if and only if φ is true in each view of $\mathcal{M}$, i.e. for each V_i:

$$\langle O_{\mathcal{M}}, V_i \rangle \models^+ \varphi.$$

$\langle O_{\mathcal{M}}, V_i \rangle \models^+ \varphi$ if and only if either $\varphi \in V_i$ or one of the following conditions is satisfied.

- $\langle O_{\mathcal{M}}, V_i \rangle \models^+ \langle \neg, \psi \rangle$ if and only if:

  $$\langle O_{\mathcal{M}}, V_i \rangle \models^- \psi.$$

- $\langle O_{\mathcal{M}}, V_i \rangle \models^+ \langle \wedge, \psi_1, ..., \psi_n \rangle$ if and only if for each ψ_j:

  $$\langle O_{\mathcal{M}}, V_i \rangle \models^+ \psi_j.$$

- $\langle O_{\mathcal{M}}, V_i\rangle \models^{+} \langle \vee, \psi_1, ..., \psi_n\rangle$ if and only if for some ψ_j:

  $\langle O_{\mathcal{M}}, V_i\rangle \models^{+} \psi_j$.

- $\langle O_{\mathcal{M}}, V\rangle \models^{+} \langle \mathbf{a}, x, \psi_1(x), \psi_2(x)\rangle$ if and only if for some $o \in O_{\mathcal{M}}$:

  $\langle O_{\mathcal{M}}, V_i\rangle \models^{+} \psi_1(o)$ and $\langle O_{\mathcal{M}}, V_i\rangle \models^{+} \psi_2(o)$.

A formula φ is *false* in $\mathcal{M}$, $\mathcal{M} \models^{-} \varphi$, if and only if φ is false in each view of $\mathcal{M}$, i.e. for each V_i:

$\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \varphi$.

$\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \varphi$ if and only if either $\langle \neg, \varphi\rangle \in C$ or one of the following conditions is satisfied.

- $\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \langle \neg, \psi\rangle$ if and only if:

  $\langle O_{\mathcal{M}}, V_i\rangle \models^{+} \psi$.

- $\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \langle \wedge, \psi_1, ..., \psi_n\rangle$ if and only if for some ψ_j:

  $\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \psi_j$.

- $\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \langle \vee, \psi_1, ..., \psi_n\rangle$ if and only if for each ψ_j:

  $\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \psi_j$.

- $\langle O_{\mathcal{M}}, V_i\rangle \models^{-} \langle \mathbf{a}, x, \psi_1(x), \psi_2(x)\rangle$ if and only if

  $\langle O_{\mathcal{M}}, V_i\rangle \models^{+} \langle \%, 0, x, \psi_1(x), \psi_2(x)\rangle$.

Notice that a quantified description and a description of a group of objects φ can only be true or false in a view if respectively $\varphi \in V_i$ or $\langle \neg, \varphi\rangle \in V_i$. It is not possible to define their meanings in terms of their constituents. The reason for this is illustrated by the following example. Suppose that we know 5 balls to be red. Does this imply that there only exist 5 red balls? Since partial models are incomplete, we cannot answer such a question.

The following definition defines the strong satisfiability relation that is used to characterize the reasoning process. Since it will be used to characterize the reasoning process, here the meaning of a quantified description and a description of a group of objects will be defined.

Definition 7.4 Let φ be a formula and let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_1, ..., V_m\}\rangle$ be a partial model.

$\mathcal{M}$ *strongly* satisfies φ, $\mathcal{M} \models \varphi$, if and only if the following conditions are satisfied.

- $O_{\mathcal{M}} \models \langle r, o_1, ..., o_n\rangle$ if and only if for each V_i:

$\langle O_{\mathcal{M}}, V_i \rangle \models \langle r, o_1, ..., o_n \rangle$.

- $\mathcal{M} \models \langle \neg, \psi \rangle$ if and only if for each V_i:

  $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \neg, \psi \rangle$.

- $\mathcal{M} \models \langle \wedge, \psi_1, ..., \psi_n \rangle$ if and only if for each ψ_j:

  $\mathcal{M} \models \psi_j$.

- $\mathcal{M} \models \langle \vee, \psi_1, ..., \psi_n \rangle$ if and only if each V_i there exists a ψ_j:

  $\langle O_{\mathcal{M}}, V_i \rangle \models \psi_j$.

- $\mathcal{M} \models \langle \#, n, x, \psi(x) \rangle$ if and only if there exists a $O \subseteq O_{\mathcal{M}}$ such that:
  1. for each $o \in O$:
     $\mathcal{M} \models \psi(o)$.
  2. $|O| = n$,
  3. for each $o, o' \in O$:
     $\mathcal{M} \models \langle \neg, \langle =, o, o' \rangle \rangle$,
  4. and for each $o \in (O_{\mathcal{M}} - O)$ such that:
     $\mathcal{M} \models \psi(o)$
     there holds:
     $\mathcal{M} \models \langle =, o, o' \rangle$
     for some $o' \in O$.

- $\mathcal{M} \models \langle \mathbf{a}, x, \psi_1(x), \psi_2(x) \rangle$ if and only if for some $o \in O_{\mathcal{M}}$:

  $\mathcal{M} \models \psi_1(o)$ and $\mathcal{M} \models \psi_2(o)$.

- $\mathcal{M} \models \langle \%, p, x, \psi_1(x), \psi_2(x) \rangle$ if and only if for some natural number n:
  1. $\mathcal{M} \models \langle \#, n, x, \psi_1(x) \rangle$,
  2. there exists an $O_1 \subseteq O_{\mathcal{M}}$ with $|O_1| = n$ such that for every $o, o' \in O_1$:
     $\mathcal{M} \models \langle \neg, \langle =, o, o' \rangle \rangle$,
  3. $0 \leqslant k = p \cdot n \leqslant n$, and
  4. for each V_i there exists a $\{o_1, ..., o_k\} \subseteq O_1$ such that for each $o \in \{o_1, ..., o_k\}$:
     $\langle O_{\mathcal{M}}, V_i \rangle \models \psi_2(o)$
     and for each $o \in (O_1 - \{o_1, ..., o_k\})$:
     $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \neg, \psi(o) \rangle$.

A formula is satisfied by a view, $\langle O_{\mathcal{M}}, V_i \rangle \models \varphi$, if and only if the following conditions are satisfied.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle r, o_1, ..., o_n \rangle$ if and only if:

$\langle r, o_1, ..., o_n \rangle \in V_i$.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \neg, \psi \rangle$ if and only if:

  $\langle \neg, \psi \rangle \in V_i$.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \wedge, \psi_1, ..., \psi_n \rangle$ if and only if for each ψ_j:

  $\langle O_{\mathcal{M}}, V_i \rangle \models \psi_j$.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \vee, \psi_1, ..., \psi_n \rangle$ if and only if:

  $\langle \vee, \psi_1, ..., \psi_n \rangle \in V_i$.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \#, n, x, \psi(x) \rangle$ if and only if:

  $\langle \#, n, x, \psi(x) \rangle \in V_i$.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \mathbf{a}, x, \psi_1(x), \psi_2(x) \rangle$ if and only if for some $o \in O_{\mathcal{M}}$:

  $\langle O_{\mathcal{M}}, V_i \rangle \models \psi_1(o)$ and $\langle O_{\mathcal{M}}, V_i \rangle \models \psi_2(o)$.

- $\langle O_{\mathcal{M}}, V_i \rangle \models \langle \%, p, x, \psi_1(x), \psi_2(x) \rangle$ if and only if

  $\langle \%, p, x, \psi_1(x), \psi_2(x) \rangle \in V_i$.

Using the weak satisfiability relation, we can define whether a partial model is consistent.

Definition 7.5 Let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_1, ..., V_m\} \rangle$ be a partial model.

$\mathcal{M}$ is consistent if and only if for no view V_i there holds for some $\varphi \in V$:

$$\mathcal{M} \models^{-} \varphi.$$

7.4 The reasoning process

To be able to reason by using partial models, an updating function has to be defined. This updating function maps a partial model $\mathcal{M}$ and some piece of new information φ on a new partial model $\mathcal{N}$. This new partial model, which is called an extension of $\mathcal{M}$, must contain all the information described by $\mathcal{M}$ and must strongly satisfy φ. So, first of all I have to define when an extension $\mathcal{N}$ is at least as informative as $\mathcal{M}$. The idea behind the ordering relation defined here is that if a partial model $\mathcal{M}$ can be embedded in a partial model $\mathcal{N}$, the latter is at least as informative as the former. Since the anonymous objects of a partial model can denote any entity in the world we are reasoning about, in another partial model other objects may have been used to denote the same entity. It is even possible that an object in a partial model is represented by more than one object in another partial model. Therefore, a partial model $\mathcal{N}$ can only be at least as informative as a partial model $\mathcal{M}$ with

respect to some *interpretation function*. This interpretation function, which is a partial function from D to D, interprets the objects of $\mathcal{M}$ in $\mathcal{N}$.

Since objects of a partial model $\mathcal{M}$ can be unified in a more informative partial model $\mathcal{N}$ by an interpretation function, i.e. they will be represented by one object, the equality relation may become useless. So, if $\mathcal{M} \models^+ \langle =, a, b\rangle$ and $f(a) = f(b)$ for some interpretation function f, $\langle =, f(a), f(b)\rangle$ need not occur in $\mathcal{N}$. Therefore, only if $\mathcal{N}$ is extended with these equality relations, $\mathcal{M}$ can be embedded in $\mathcal{N}$.

Furthermore, if the partial model $\mathcal{N}$ satisfies $\langle =, a, b\rangle$ one of the objects a, b can be removed. Since an interpretation function can only map objects of the partial model $\mathcal{M}$ to one of the objects a, b, these objects have to be unified. Therefore, a unification function will be used.

Definition 7.6 Let φ be a formula, let O_{ex} be the set of external objects for φ and let $f : D \to D$ be a partial interpretation function.

Then $\varphi[f]$ denotes the result of substituting every object o that occurs in φ by $f(o)$.

- $\langle r, o_1, ..., o_n\rangle[f] = \langle f(r), f(o_1), ..., f(o_n)\rangle$.
- $\langle \neg, \psi\rangle[f] = \langle \neg, \psi[f]\rangle$.
- $\langle \wedge, \psi_1, ..., \psi_n\rangle[f] = \langle \wedge, \psi_1[f], ..., \psi_n[f]\rangle$.
- $\langle \vee, \psi_1, ..., \psi_n\rangle[f] = \langle \vee, \psi_1[f], ..., \psi_n[f]\rangle$.
- $\langle \#, n, o, \psi(o)\rangle[f] = \langle \#, n, o', \psi[f'](o')\rangle$
  where $o' \in D$, $o' \notin f(O_{ex})$ and $f' = f[^{o'}/_{o}]$.
- $\langle \mathbf{a}, o, \psi_1(o), \psi_2(o)\rangle[f] = \langle \mathbf{a}, o', \psi_1[f'](o'), \psi_2[f'](o')\rangle$
  where $o' \in D$, $o' \notin f(O_{ex})$ and $f' = f[^{o'}/_{o}]$.
- $\langle \%, p, o, \psi_1(o), \psi_2(o)\rangle[f] = \langle \%, p, o', \psi_1[f'](o'), \psi_2[f'](o')\rangle$
  where $o' \in D$, $o' \notin f(O_{ex})$ and $f' = f[^{o'}/_{o}]$.

Definition 7.7 Let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_{\mathcal{M},1}, ..., V_{\mathcal{M},m}\}\rangle$ and $\mathcal{N} = \langle O_{\mathcal{N}}, \{V_{\mathcal{N},1}, ..., V_{\mathcal{N},n}\}\rangle$ be two partial models and let $h : D \to D$ be an interpretation function. h is a partial function, which is at least undefined for those objects not belonging to $O_{\mathcal{M}}$. Furthermore, let $Eq(O)$ be the set of equality relations for the objects O and let $u : O_{\mathcal{N}} \to O_{\mathcal{N}}$ be a unification function. Eq is defined as:

$$Eq(O) = \{\langle =, o, o\rangle \mid o \in O\}.$$

Let $[o]_=$ be the equivalent class of equal objects in which o occurs; i.e.:

$o' \in [o]_=$ if and only if there exists a sequence $o_1, ..., o_n$, $o = o_1$, $o' = o_n$ and for each $1 \leqslant i < n$:

$$\mathcal{N} \models^+ \langle =, o_i, o_{i+1}\rangle \text{ or } \mathcal{N} \models^+ \langle =, o_{i+1}, o_i\rangle.$$

Then u is defined as:

for each $o \in O_{\mathcal{N}}$ there exists a $o' \in [o]_=$ such that:

$$[o]_= = \{o'' \in O_{\mathcal{N}} \mid u(o'') = o'\}.$$

$\mathcal{N}$ is at least as informative as $\mathcal{M}$ under the partial interpretation function h, $\mathcal{M} \leqslant_h \mathcal{N}$, if and only if for each $o \in (O_{\mathcal{M}} \cap Names)$:

$$h(o) = o,$$

and for each $V_{\mathcal{N},j}$ there exists a $V_{\mathcal{M},i}$ such that:

$$V_{\mathcal{M},i}[u \circ h] \subseteq (V_{\mathcal{N},j}[u] \cup Eq(O_{\mathcal{N}})).$$

Lemma 7.8 Let $\mathcal{L}, \mathcal{M}, \mathcal{N}$ be partial models and let f and g be interpretation functions such that:

$$\mathcal{L} \leqslant_f \mathcal{M} \leqslant_g \mathcal{N}$$

Then $\mathcal{L} \leqslant_{g \circ f} \mathcal{N}$.

Proof Let $\mathcal{L} = \langle O_{\mathcal{L}}, \{V_{\mathcal{L},1}, ..., V_{\mathcal{L},\ell}\}\rangle$, $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_{\mathcal{M},1}, ..., V_{\mathcal{M},m}\}\rangle$ and $\mathcal{N} = \langle O_{\mathcal{N}}, \{V_{\mathcal{N},1}, ..., V_{\mathcal{N},n}\}\rangle$.
Since $\mathcal{M} \leqslant_g \mathcal{N}$, for each $V_{\mathcal{N},k}$ there exists a $V_{\mathcal{M},j}$ such that:

$$V_{\mathcal{M},j}[u_{\mathcal{N}} \circ g] \subseteq (V_{\mathcal{N},k}[u_{\mathcal{N}}] \cup Eq(O_{\mathcal{N}})).$$

Since $\mathcal{L} \leqslant_f \mathcal{M}$, for each $V_{\mathcal{M},j}$ there exists a $V_{\mathcal{L},i}$ such that:

$$V_{\mathcal{L},i}[u_{\mathcal{M}} \circ f] \subseteq (V_{\mathcal{M},j}[u_{\mathcal{M}}] \cup Eq(O_{\mathcal{M}})).$$

Therefore, for each $V_{\mathcal{N},k}$ there exists a $V_{\mathcal{L},i}$ such that:

$$V_{\mathcal{L},i}[u_{\mathcal{N}} \circ g \circ f] \subseteq (V_{\mathcal{N},j}[u_{\mathcal{N}}] \cup Eq(O_{\mathcal{N}})).$$

Hence, $\mathcal{L} \leqslant_{g \circ f} \mathcal{N}$. □

Definition 7.9 Let $\mathcal{M}$ and $\mathcal{N}$ be two partial models.

$\mathcal{N}$ and $\mathcal{M}$ are equivalent partial models, $\mathcal{M} \equiv \mathcal{N}$, if and only if for some interpretation function f:

$$\mathcal{M} \leqslant_f \mathcal{N}$$

and some interpretation function g:

$$\mathcal{N} \leqslant_g \mathcal{M}.$$

Using this ordering on the partial models, the set of possible extensions of $\mathcal{M}$ satisfying φ can be defined. These partial models must be at least as informative as $\mathcal{M}$, must satisfy φ and must not be more informative than strictly necessary.

Definition 7.10 $\mathcal{N} \in Ex(\mathcal{M}, \varphi)$ if and only if $\mathcal{N}$ is consistent, there exists f:

$$\mathcal{M} \leqslant_f \mathcal{N} \text{ and } \mathcal{N} \models \varphi[f]$$

and for every consistent $\mathcal{L}$ such that for some g, h:

$$\mathcal{M} \leqslant_g \mathcal{L} \leqslant_h \mathcal{N}$$

there holds:

$$\mathcal{L} \models \varphi[g] \text{ implies } \mathcal{L} \equiv \mathcal{M}.$$

For this set $Ex(\mathcal{M}, \varphi)$ we can prove that all the partial models in this set are equivalent.

Theorem 7.11 Let $\mathcal{M}$ be a partial model and let φ be a formula.

For every $\mathcal{L}, \mathcal{N} \in Ex(\mathcal{M}, \varphi)$ there holds: $\mathcal{L} \equiv \mathcal{N}$.

Proof Let $\mathcal{L}, \mathcal{N} \in Ex(\mathcal{M}, \varphi)$.

$$\mathcal{L} = \langle O_{\mathcal{L}}, \{V_{\mathcal{L},1}, \ldots, V_{\mathcal{L},m}\}\rangle.$$

$$\mathcal{N} = \langle O_{\mathcal{N}}, \{V_{\mathcal{N},1}, \ldots, V_{\mathcal{N},n}\}\rangle.$$

Let f and g be two interpretation functions such that:

$$\mathcal{M} \leqslant_f \mathcal{L} \text{ and } \mathcal{M} \leqslant_g \mathcal{N}.$$

Let $h : (O_{\mathcal{L}} - f(O_{\mathcal{M}})) \rightarrow D$ be a bijective function such that:

$$h(O_{\mathcal{L}} - f(O_{\mathcal{M}})) \cap O_{\mathcal{M}} \subseteq Names$$

and

$$h(O_{\mathcal{L}} - f(O_{\mathcal{M}})) \cap O_{\mathcal{N}} \subseteq Names.$$

This function introduces new objects for objects of the partial model $\mathcal{L}$ not originating from the partial model M.
Furthermore, let a and b be two interpretation functions that are defined as:

$$a(o) = \begin{cases} f(o) & \text{if } o \in O_{\mathcal{M}} \\ h^{-1}(o) & \text{if } o \in h(O_{\mathcal{L}} - f(O_{\mathcal{M}})) \end{cases}$$

and

$$b(o) = \begin{cases} g(o) & \text{if } o \in O_{\mathcal{M}} \\ o & \text{if } o \in (O_{\mathcal{N}} - g(O_{\mathcal{M}})). \end{cases}$$

Given the functions a and b, we can construct a partial model

$$\mathcal{L}' = \langle O_{\mathcal{L}'}, \{V_{\mathcal{L}',1}, \ldots, V_{\mathcal{L}',m}\}\rangle.$$

and a partial model

$$\mathcal{N}' = \langle O_{\mathcal{N}'}, \{V_{\mathcal{N}',1}, \ldots, V_{\mathcal{N}',n}\}\rangle.$$

where

$$O_{\mathcal{L}'} = (O_{\mathcal{M}} \cup h(O_{\mathcal{L}} - f(O_{\mathcal{M}}))),$$

$$O_{\mathcal{N}'} = (O_{\mathcal{M}} \cup (O_{\mathcal{N}} - g(O_{\mathcal{M}}))),$$

$$V_{\mathcal{L}',i} = \{\varphi \mid \psi \in V_{\mathcal{L},i}, \varphi[a] = \psi\} \cup \{\langle =, o, o'\rangle \mid o, o' \in O_{\mathcal{M}}, f(o) = f(o')\}$$

and

$$V_{\mathcal{N}',i} = \{\varphi \mid \psi \in V_{\mathcal{N},i}, \varphi[b] = \psi\} \cup \{\langle =, o, o'\rangle \mid o, o' \in O_{\mathcal{M}}, g(o) = g(o')\}.$$

For these partial models $\mathcal{L}'$ and $\mathcal{N}'$ there holds:

$$(O_{\mathcal{L}'} - O_{\mathcal{N}'}) \subseteq (O_{\mathcal{M}} \cup Names),$$

$$O_{\mathcal{M}} \subseteq O_{\mathcal{L}'} \text{ and } O_{\mathcal{M}} \subseteq O_{\mathcal{N}'},$$

$O_{\mathcal{M}} \leqslant_{id} O_{\mathcal{L}'}$ and $O_{\mathcal{M}} \leqslant_{id} O_{\mathcal{N}'}$,

$O_{\mathcal{L}} \equiv O_{\mathcal{L}'}$ and $O_{\mathcal{N}} \equiv O_{\mathcal{N}'}$

and

$O_{\mathcal{L}'} \models \varphi$ and $O_{\mathcal{N}'} \models \varphi$.

Furthermore, we can construct a partial model

$$\mathcal{P} = \langle O_{\mathcal{L}'} \cup O_{\mathcal{N}}, \{V_{\mathcal{L}',1}, ..., V_{\mathcal{L}',m}, V_{\mathcal{N},1}, ..., V_{\mathcal{N},n}\}\rangle.$$

Since $V_{\mathcal{P}} \supseteq V_{\mathcal{L}'}$ and $O_{\mathcal{P}} \supseteq O_{\mathcal{L}'}$ we have:

$\mathcal{P} \leqslant_{id} \mathcal{L}'$.

Hence by Lemma 7.8:

$\mathcal{P} \leqslant_{idoa} \mathcal{L}$.

Similarly, since $V_{\mathcal{P}} \supseteq V_{\mathcal{N}}$ and $O_{\mathcal{P}} \supseteq O_{\mathcal{N}}$, we have:

$\mathcal{P} \leqslant_{id} \mathcal{N}'$.

Hence by Lemma 7.8:

$\mathcal{P} \leqslant_{idob} \mathcal{N}$.

Furthermore, since $\mathcal{M} \leqslant_{id} \mathcal{L}'$ and $\mathcal{M} \leqslant_{id} \mathcal{N}'$,

$\mathcal{M} \leqslant_{id} \mathcal{P}$.

Because for each $V_{\mathcal{L}',i}$:

$\langle O_{\mathcal{L}'} \cup O_{\mathcal{N}'}, V_{\mathcal{L}',i}\rangle \models \varphi$

and for each $V_{\mathcal{N}',j}$:

$\langle O_{\mathcal{L}'} \cup O_{\mathcal{N}'}, V_{\mathcal{N}',j}\rangle \models \varphi$,

there holds:

$$\mathcal{P} \models \varphi.$$

Since $\mathcal{L} \in Ex(\mathcal{M}, \varphi)$ and $\mathcal{P} \models \varphi$, it follows from Definition 7.10 that

$$\mathcal{L} \leqslant_{a'} \mathcal{P}$$

for some a' and therefore:

$$\mathcal{L} \leqslant_{boa'} \mathcal{N}.$$

Likewise, $\mathcal{N} \in Ex(\mathcal{M}, \varphi)$ and $\mathcal{P} \models \varphi$, implies that

$$\mathcal{N} \leqslant_{b'} \mathcal{P}$$

for some b' and therefore:

$$\mathcal{N} \leqslant_{aob'} \mathcal{L}.$$

Then, by Definition 7.9:

$$\mathcal{L} \equiv \mathcal{N}.$$

□

Remark 7.12 Although all partial models are equivalent, there are important differences between them. Here I will define two conditions in order to choose a canonical form.

Since objects in a partial model can denote the same objects in a world, there is no upper bound on the number of objects in a partial model. Therefore, the only non arbitrary number of objects in a partial model is the smallest number of objects.

Definition 7.13 Let $\mathcal{M}$ be a partial model and let φ be a formula. Furthermore, let $\mathcal{L}, \mathcal{N} \in Ex(\mathcal{M}, \varphi)$.

$\mathcal{L} <_1 \mathcal{M}$ if and only if $|O_{\mathcal{L}}| > |O_{\mathcal{N}}|$.

There is yet another condition that has to be satisfied. It is possible that a partial model contains redundant views. These are views that describe at least the same information as some other view. Therefore, a partial model containing more views than some equivalent partial model, must contain redundant views. Of course, we prefer those partial models that do not contain redundant views.

Definition 7.14 Let $\mathcal{M}$ be a partial model and let φ be a formula. Furthermore, let $\mathcal{L}, \mathcal{N} \in Ex(\mathcal{M}, \varphi)$.

$$\mathcal{L} <_2 \mathcal{N} \text{ if and only if } |V_{\mathcal{L}}| > |V_{\mathcal{N}}|.$$

Clearly, because $<_1$ and $<_2$ are strict partial orderings on $Ex(\mathcal{M}, \varphi)$, so is the combined ordering, $<_{1,2} = (<_1 \text{ or } <_2)$. Using the ordering $<_{1,2}$, the preferred partial model of $Ex(\mathcal{M}, \varphi)$ can be determined. Whether these preferred partial models are the models we are looking for, is still an open question.

Definition 7.15 Let $Ex(\mathcal{M}, \varphi)$ be the set of minimal extensions of $\mathcal{M}$ that satisfy φ. Then, the set of preferred minimal extensions of $\mathcal{M}$ is defined as follows:

$$Ex_p(\mathcal{M}, \varphi) = max(Ex(\mathcal{M}, \varphi), <_{1,2}).$$

When a partial model $\mathcal{M}$ is being updated with a formula φ, the partial model that is actually constructed by the updating function, must be an element of the set $Ex_p(\mathcal{M}, \varphi)$. This partial model is denoted by $\mathcal{M}[\varphi]$.

8

Uncertain conclusions

If a formula is satisfied by every view of a partial model, we are certain that it holds. It may be possible that a formula is satisfied by only some of the views. In that case we cannot be certain whether it holds. To express this uncertainty, I will define a *probability* and a *likelihood* measure in this chapter. The probability measure will be used for conclusions that express an *expectation*, and the likelihood measure will be used for conclusions that express an *explanation*. Both of these certainty measures express our ignorance with respect to our partial models. Examples of both kinds of conclusions are respectively:

- 'John probably loves a woman',
- 'It is likely that the patient has a brain tumor'.

How these conclusions can be derived, I will describe in the following sections.

The contents of this chapter was presented at the IPMU conference in 1990 and will be published [54].

8.1 Expectations

An expectation like: 'John probably loves a woman' can be derived from the quantified description: 'Most men love a woman'. The question is how to formalize this deduction. One possibility is to assume that John is randomly chosen from the class of all men and that the quantified description is independent of other known information, e.g. that it is not overruled by some other quantified description. The main objection against this approach is the demand that John has to be randomly chosen. When John is a colleague of yours, he cannot be considered to be randomly chosen from the class of all men.

The approach taken here, is based on the ideas of R. Carnap [8]. A measure for the expectation of a formula will only be derived from the information available. So we have to define how the probability measure is based on the set of views. To motivate the definition below, first consider the following situation. Suppose that some of the views of a partial model satisfy the formula for which we want to determine a probability measure. Then the definition of the probability measure should satisfy the following conditions.

- The probability measure should be proportional to the number of views that satisfy the formula.
- The measure should be inversely proportional to the total number of views.
- Since we have no reason to prefer one view to another, the insufficient reason argument of Bernoulli and Laplace [22] can be applied on the views.

The probability measure defined below satisfies these requirements.

Definition 8.1 Let $\mathcal{M} = \langle O, \{V_1, ..., V_n\}\rangle$ be a partial model and let a formula φ be known in every view V_i, i.e. for every view $\langle O, V_i\rangle$: either $\langle O, V_i\rangle \models^+ \varphi$ or $\langle O, V_i\rangle \models^- \varphi$. Then the probability measure $Pr(\varphi \mid \mathcal{M})$ for a formula φ with respect to a partial model $\mathcal{M}$ is defined as follows:

$$Pr(\varphi \mid \langle O, \{V_1, ..., V_n\}\rangle) = \frac{1}{n} \cdot \sum_i Pr(\varphi \mid V_i)$$

and

$$Pr(\varphi \mid \langle O, V_i\rangle) = \begin{cases} 1 & \text{if } \langle O, V_i\rangle \models^+ \varphi \\ 0 & \text{if } \langle O, V_i\rangle \models^- \varphi \end{cases}$$

Given this definition, the following observation can be made.

Observation 8.2 Let $\mathcal{M} = \langle O, \{V_1, ..., V_n\}\rangle$ be a partial model and let φ and ψ be formulas that are known in $\mathcal{M}$, i.e. for each V_i:

$$\langle O_{\mathcal{M}}, V_i \rangle \models \varphi \text{ or } \langle O_{\mathcal{M}}, V_i \rangle \models \neg\varphi$$

and

$$\langle O_{\mathcal{M}}, V_i \rangle \models \psi \text{ or } \langle O_{\mathcal{M}}, V_i \rangle \models \neg\psi.$$

Then the axioms of probability are satisfied. These axioms are:

- $0 \leqslant Pr(\varphi \mid \mathcal{M}) \leqslant 1$.
- $Pr(\varphi \vee \neg\varphi) = 1$.
- If $\mathcal{M} \models \neg(\varphi \wedge \psi)$, then:

$$Pr(\varphi \vee \psi \mid \mathcal{M}) = Pr(\varphi \mid \mathcal{M}) + Pr(\varphi \mid \mathcal{M}).$$

Using the definition, it is possible to derive that John probably loves a woman if the conceptual model contains the fact 'John is a man' and the quantified description 'Most men love a woman' is used.

Property 8.3 Let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V\} \rangle$ denote a partial model and let φ denote the quantified description: $\langle \%, p, x, \psi_1(x), \psi_2(x) \rangle$. Furthermore, let all the objects described by the class ψ_1 be known in the partial model, let *ob* be one of these objects and let ψ_2 be unknown for these objects.
Then:

$$Pr(\psi_2(ob) \mid \mathcal{M}[\langle \%, p, x, \psi_1(x), \psi_2(x) \rangle]) = p$$

Proof Let $\mathcal{N} = \mathcal{M}[\varphi]$, let

$$X = \{o \in O_{\mathcal{M}} \mid \mathcal{M} \models^{+} \psi_1(o)\},$$

$$Y = \{o \in X \mid \mathcal{M} \models^{+} \psi_2(o)\},$$

$$Z = \{o \in X \mid \mathcal{M} \models^{-} \psi_2(o)\}.$$

and let $|X| = n$. Since *ob* belongs to the class described by ψ_1 and since ψ_2 is unknown for the objects of this class,

$$ob \in X \text{ and } Y = Z = \varnothing$$

- Suppose that $p = 0$. Then for no $o \in X$:

$$\mathcal{N} \models^{+} \psi_2(o).$$

Hence

$$Pr(\psi_2(ob) \mid \mathcal{M}[\varphi]) = 0$$

- Suppose that $p = 1$. Then for every $o \in X$:

  $$\mathcal{N} \models^+ \psi_2(o).$$

  Hence

  $$Pr(\psi_2(ob) \mid \mathcal{M}[\varphi]) = 1$$

- Suppose that $0 < p < 1$. Then we can choose $k = \binom{n}{p \cdot n}$ different subsets of X containing $p \cdot n$ objects.
  For every subset of X containing $p \cdot n$ different objects o, there must exist a view V_i for which there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^+ \psi_2(o)$$

  while for the other objects o in X there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^- \psi_2(o).$$

  Therefore $\{V_1, ..., V_k\}$. is the set of views of $\mathcal{N}$.
  In $\binom{n-1}{p \cdot n-1}$ of these views V_i, there holds for the object ob:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^+ \psi_2(ob),$$

  while for the other views V_i there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^- \psi_2(ob).$$

  Hence,

  $$Pr(\psi_2(ob) \mid \mathcal{M}[\varphi]) = \frac{\binom{n-1}{p \cdot n-1}}{\binom{n}{p \cdot n}} = p.$$

□

Now I will discuss two examples that have been used to defend respectively the Dempster-Shafer theory [56] and the transferable belief model [60]. In [56] G. Shafer discusses an example illustrating that a Bayesian cannot always assign a consistent probability measure.

Example 8.4 *Life near Sirius?* Are there or are there not living things in the orbit of the star Sirius? Some scientists may have evidence on this question, but most of us will profess complete ignorance about it. So, let α denote the possibility that there is such life, then we know that $\alpha \vee \neg\alpha$ will hold. When we incorporate this information in the partial model,

$$Pr(\alpha \mid \mathcal{M}_1) = \frac{1}{2}.$$

We can also consider the question in the context of a more refined set of possibilities. We might raise the question whether there exist planets around Sirius. Let this be denoted by β. Shafer considers three possibilities in his example, viz. α, $\neg\alpha \wedge \beta$ and $\neg\alpha \wedge \neg\beta$. If we update the partial model with $\alpha \vee (\neg\alpha \wedge \beta) \vee (\neg\alpha \wedge \neg\beta)$, we get the same inconsistent probability measures for α as Shafer does. The formula $\alpha \vee (\neg\alpha \wedge \beta) \vee (\neg\alpha \wedge \neg\beta)$ states that we consider three distinct possibilities. This is not what we actually considered. What we did consider, however, was life or no life and planets or no planets.

When we update $\mathcal{M}_1$ with $\beta \vee \neg\beta$, the probability of α in the resulting partial model $\mathcal{M}_2$ will be:

$$Pr(\alpha \mid \mathcal{M}_2) = \frac{1}{2}.$$

The next example was used by P. Smets to illustrate the difference between his transferable belief model and the Bayesian approach when new information is received [60]. Smets's transferable belief model is based on the Dempster-Shafer theory and is intended to model changes of belief when new information comes available. In the transferable belief model, there is a distinction between a *credal* level and a *pignistic* (betting) level. At the credal level, belief masses are assigned to subsets of the frame of discernment. A belief mass assigned to a set can be transferred to its subsets when new information is received. When one is asked to make a bet, the belief masses assigned to a set, have to be divided over its elements using the insufficient reason argument; i.e. the probabilities at the pignistic level are being determined.

Example 8.5 *Mr Jone's murdering* Big Boss has decided that Mr Jone has to be murdered by one of the three persons present in his waiting room and whose names are Peter, Paul and Mary. Big Boss has decided that the killer on duty will be selected according to the result of a dice tossing experiment: if the result is even, the killer will be a female: if the result is odd, the killer will be a male. We, the judges, know who were in the waiting room and know about the story of the dice tossing experiment, but ignore what was the result and who was selected. We also ignore how Big Boss would have decided between Peter and Paul if the result given by the dice had been odd.

If we update a partial model using the information available, the probability that the killer is a female and the probability that the killer is a male will both be equal to 0.5.

Then we learn that if Peter was not the killer, he would go to the police station at the time of the killing in order to get a perfect alibi. Peter indeed went to the police station, so he is not the killer. Now the question is what is the probability that the killer is a female and what is the probability that the killer is a male, given this new information.

If we update the partial model using this new information, the probability of Peter being the killer and the probability of Paul being the killer will change. The probability that the killer is a female and the probability that the killer is a male, however, will still both be equal to 0.5.

This example shows that the probability measure defined here, like the transferable belief model, but unlike the Bayesian model, results in intuitively sound conclusions. Since the same results follow from the probability measure defined here without the need of using two levels, the probability measure defined here seems to be preferable to the transferable belief model.

8.2 Inheritance networks

A totally different approach toward default reasoning is based on the view that (some) default rules are actually quantified descriptions. In his article *In defence of probability* [12], P. Cheeseman claims that all default rules are actually quantified descriptions. It is, however, not difficult to find counter examples for this claim. More modest claims are presented by F. Bacchus [2] and by L. Shastri [57]. Both authors claim that some default relations are actually quantified descriptions. Bacchus distinguishes between quantified descriptions and relations describing typical properties. Shastri, however, distinguishes two areas in an inheritance hierarchy. According to Shastri the top of the inheritance hierarchy is an ontological tree, only used for classification. The other part of the inheritance hierarchy, which need not be a tree, contains quantified descriptions.

In most of the AI literature quantified descriptions are confused with typical properties. For example in [64, page 480] a counter example for off-path pre-emption is described, which clearly is a problem of reasoning with quantified descriptions. Touretzky et al. discuss whether George is a beer drinking marine chaplain when one knows that George is a marine and a chaplain, and also that men are beer drinkers and chaplains are not. In the discussion of this problem they say: 'the most relevant missing bit of information is the rate of beer drinking among marines'. They continue with: 'If this rate is far higher than the rate of abstention among chaplains, one would be better off assuming George is a beer drinker than not'. Also R. Reiter [49] treats quantified descriptions and typical relations identically. In his

view, properties that hold for most objects of some class, are preferred to hold for any object in this class. So, a sentence like 'most φ are ψ', he represents as:

$$\frac{\varphi(x) : \psi(x)}{\psi(x)}$$

It is not difficult to recognize a quantified description when words like 'all', 'most', 'many', 'always', 'usually', 'often', etc. are used. These words denote a percentage of a set of objects or time intervals. It becomes more difficult to recognize a quantified description in case these words are not used in the sentence. To verify whether a sentence like: 'Republicans are non pacifists' is describing a typical relation, using Shastri's view on inheritance hierarchies, we wonder if this relation can be used to classify someone as a republican. If the answer is no, then the sentence must describe a quantified description.

When a property of a type is described by a quantified description, the expectation that an object of the type has this property, can be determined. Bacchus [2] and L. Shastri [57] both describe a model for inheritance reasoning using quantified descriptions. Bacchus recognizes that a number of default relations are actually relations of the form: 'most men are beer drinkers'. He represents this kind of relations with a special kind of implication. With this approach, however, he cannot distinguish between different frequencies.

A different approach is used by Shastri [57]. In his approach frequencies are used. Actually, Shastri uses the number of objects which belong to a class. In his results, however, only the ratios between the number of objects that belong to a class are relevant. Shastri determines a conclusion by comparing the number of worlds in which the conclusion holds with the number of worlds in which it does not hold. Here a world denotes a distribution of objects over properties such that the constraints, i.e. the quantified descriptions, are met. By comparing the number of worlds that satisfy a formula of interest, with the number of worlds that do not, inheritance problems are solved. To solve the problem of multiple inheritance, Shastri uses knowledge about the distribution of objects of a type over its subtypes. He illustrates this with the Nixon diamond. To determine whether Nixon has pacifist beliefs, knowledge about the distribution of pacifist beliefs over all person is needed. More knowledge is needed in case the types involved in the multiple inheritance do not have a common parent, but only a common ancestor. Shastri argues that this knowledge is needed, because otherwise the problem is underconstrained. It seems counter intuitive that this additional knowledge is needed to solve multiple inheritance problems. The actual reason why he needs the additional knowledge is that he determines the number of objects which satisfy a property. In the example of the Nixon diamond he determines the number of republican-quakers who have pacifist and who have non-pacifist beliefs. So, not only the ratio between pacifist and non-pacifist republican-quakers is estimated, but also the number of republican-quakers. We are, however, only interested in the former and this can be determined without additional knowledge.

The probability measure defined in this chapter can also be used to justify inheritance reasoning in a similar way as is proposed by Shastri [57]. This is illustrated by the following two theorems.

Theorem 8.6 Let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_1, ..., V_n\}\rangle$ be a partial model and let $\varphi = \langle \%, p, x, \alpha(x), \gamma(x)\rangle$ and $\psi = \langle \%, q, x, \beta(x), \gamma(x)\rangle$ be two quantified descriptions. Furthermore, let all the objects described by the classes α and β be known in the partial model, let the class described by α be a subclass of the class described by β and let γ be unknown for any object of the class described by β.

Then for any object *ob* that belongs to the class described by α there holds:

$$Pr(\gamma(ob) \mid \mathcal{M}[\varphi \wedge \psi]) = Pr(\gamma(ob) \mid \mathcal{M}[\varphi]).$$

Proof Let $\mathcal{N} = \mathcal{M}[\varphi \wedge \psi]$, let

$$A = \{o \in O_{\mathcal{M}} \mid \mathcal{M} \models^{+} \alpha(o)\},$$

$$B = \{o \in O_{\mathcal{M}} \mid \mathcal{M} \models^{+} \beta(o)\},$$

$$C = \{o \in B \mid \mathcal{M} \models^{+} \gamma(o)\},$$

$$D = \{o \in B \mid \mathcal{M} \models^{-} \gamma(o)\}$$

and let $a = |A|$ and $b = |B|$.
Since the class of objects described by α is a subclass of the class described by β, $A \subset B$.
Since $\gamma(o)$ is unknown for any object $o \in B$, $B \cap C = B \cap D = \varnothing$.

- Suppose that $p = 1$ and $q \leqslant 1$.
  Then for every $o \in A$:

  $$\mathcal{N} \models^{+} \gamma(o).$$

  Since $ob \in A$,

  $$Pr(\gamma(o) \mid \mathcal{M}[\varphi \wedge \psi]) = 1 = Pr(\gamma(o) \mid \mathcal{M}[\varphi]).$$

- Suppose that $p = 0$ and $q \geqslant 0$.
  Then for every $o \in A$:

  $$\mathcal{N} \models^{-} \gamma(o).$$

Since $ob \in A$,

$$Pr(\gamma(o) \mid \mathcal{M}[\varphi \wedge \psi]) = 0 = Pr(\gamma(o) \mid \mathcal{M}[\varphi]).$$

- Suppose that $0 < p < 1$ and $0 < q < 1$.
  Then we can choose $\binom{a}{p \cdot a}$ different subsets of A containing $p \cdot a$ objects.
  For each subset, $\mathcal{N}$ must contain a view such that for the objects o in the subset there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^+ \gamma(o)$$

  and for the other objects in A there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^- \gamma(o).$$

  Furthermore, given some subset of A, we can also choose $\binom{b-a}{q \cdot b - p \cdot a}$ different subsets of $(B - A)$ containing $(q \cdot b - p \cdot a)$ objects.
  For each subset, $\mathcal{N}$ must also contain a view such that for the objects o in the subset there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^+ \gamma(o)$$

  and for the other objects in B there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^- \gamma(o).$$

  Therefore,

  $$\mathcal{N} = \langle O_{\mathcal{N}}, \{V_1, \ldots V_k\} \rangle$$

  with $k = \binom{a}{p \cdot a} \cdot \binom{b-a}{q \cdot b - p \cdot a}$.
  In $\binom{a-1}{p \cdot a - 1} \cdot \binom{b-a}{q \cdot b - p \cdot c}$ of these views V_i there holds for an object $ob \in A$:

  $$O_{\mathcal{N}}, V_i \models^+ \gamma(ob),$$

  while in the other views V_i there holds:

  $$O_{\mathcal{N}}, V_i \models^- \gamma(ob).$$

  Hence,

  $$Pr(\gamma(ob) \mid \mathcal{M}[\varphi \wedge \psi]) = \frac{\binom{a-1}{p \cdot a - 1} \cdot \binom{b-a}{q \cdot b - p \cdot a}}{\binom{a}{p \cdot a} \cdot \binom{b-a}{q \cdot b - p \cdot a}} = p = Pr(\gamma(o) \mid \mathcal{M}[\varphi])$$

□

Example 8.7 Let Pierre be a Quebecois. Furthermore, let it be known that most Quebecios are not native English speakers and that every Quebecois is a Canadian. Finally, let it be known that most Canadians are native English speakers. Then, according to Theorem 8.6, Pierre is probably not a native English speaker.

In his article *Objective probabilities* [36], H. E. Kyburg describes a model for assigning probabilities to formulas. To determine these probabilities, he introduces reference classes of objects in his model. For a reference class, one can specify what the percentage of objects is that satisfies some property. To be able to determine the probability that an object possesses this property, Kyburg introduces an axiom that is essentially the same as the theorem described above. Kyburg, however, neither takes into account the number of objects in a reference class nor the number of objects that are known to possess the property. Therefore, he must implicitly assume that a reference class contains infinitely many objects. Another axiom Kyburg introduces is that equivalent formulas must have the same probability. Clearly, this also holds for the model described here.

In case of multiple inheritance an object inherits conflicting properties of two unrelated classes. The following theorem confirms our intuitions about multiple inheritance. It shows that we need not know the distribution of objects in some superclass of the two classes from which the conflicting properties are inherited. We simply can make a decision by comparing the percentages of objects of the classes for which the conflicting properties hold.

Theorem 8.8 Let $\mathcal{M} = \langle O_{\mathcal{M}}, \{V_1, ..., V_n\}\rangle$ be a partial model and let $\varphi = \langle \%, p, x, \alpha(x), \gamma(x)\rangle$ and $\psi = \langle \%, q, x, \beta(x), \langle \neg, \gamma(x)\rangle\rangle$ be two quantified descriptions. Furthermore, let all the objects described by the classes α and β be known in the partial model, let α and β be two unrelated classes, let ob be the only object known to belong to both classes and let γ be unknown for any object of these classes.

Then $p > q$ implies:

$$Pr(\gamma(ob) \mid \mathcal{M}[\varphi \wedge \psi]) > \frac{1}{2}$$

.

Proof Let $\mathcal{N} = \mathcal{M}[\varphi \wedge \psi]$, let

$$A = \{o \in O_{\mathcal{M}} \mid \mathcal{M} \models^{+} \alpha(o)\},$$

$$B = \{o \in O_{\mathcal{M}} \mid \mathcal{M} \models^{+} \alpha(o)\},$$

$$C = \{o \in A \mid \mathcal{M} \models^{+} \gamma(o)\},$$

$$D = \{o \in A \mid \mathcal{M} \models^{-} \gamma(o)\},$$

$$E = \{o \in B \mid \mathcal{M} \models^{+} \gamma(o)\},$$

$$F = \{o \in B \mid \mathcal{M} \models^{-} \gamma(o)\}$$

and let $a = |A|$ and $b = |B|$.
Since ob is the only known object in the intersection of A and B, $\{ob\} = (A \cap B)$.
Since $\gamma(o)$ is unknown for any object $o \in (A \cup B)$, $A \cap C = A \cap D = B \cap E = B \cap F = \varnothing$.
Notice that it is not necessary to make an assumption about the objects in the intersection of the two classes A and B. Treating all the objects as distinct objects, will also cover each possible number of objects in the intersection of the two classes. Hence ob the is only object in the intersection of the classes.

- Suppose that $p = 1$ and $q > 0$.
  Then for every $o \in A$:

  $$\mathcal{N} \models^{+} \gamma(o).$$

  Since $ob \in A$,

  $$Pr(\gamma(ob) \mid \mathcal{M}[\varphi \wedge \psi]) = 1 > \frac{1}{2}.$$

- Suppose that $p = 0$ and $q < 1$.
  Then for every $o \in A$:

  $$\mathcal{N} \models^{-} \gamma(o).$$

  Since $ob \in A$,

  $$Pr(\gamma(o) \mid \mathcal{M}[\varphi \wedge \psi]) = 0 < \frac{1}{2}.$$

- Suppose that $p < 1$ and $q = 0$.
  Then for every $o \in B$:

  $$\mathcal{N} \models^{-} \gamma(o).$$

  Since $ob \in B$,

$$Pr(\gamma(o) \mid \mathcal{M}[\varphi \wedge \psi]) = 1 > \frac{1}{2}.$$

- Suppose that $p = 0$ and $q < 1$.
  Then for every $o \in B$:

  $$\mathcal{N} \models^{+} \gamma(o).$$

  Since $ob \in B$,

  $$Pr(\gamma(o) \mid \mathcal{M}[\varphi \wedge \psi]) = 0 < \frac{1}{2}.$$

- Suppose that $0 < p < 1$ and $0 < q < 1$.
  Consider those views in which $\gamma(ob)$.
  Then, we can choose $\binom{a-1}{p \cdot a - 1}$ different subsets of (A/ob) containing $(p \cdot a - 1)$ objects.
  For each subset, $\mathcal{N}$ must contain a view V_i such that for the objects o in the subset there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^{+} \gamma(o)$$

  and for the other objects in A there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^{-} \gamma(o).$$

  Furthermore, we can choose $\binom{b-1}{q \cdot b}$ different subsets of (B/ob) containing $(q \cdot b - p \cdot a)$ objects.
  For each subset, $\mathcal{N}$ must also contain a view V_i such that for the objects o in the subset there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^{+} \gamma(o)$$

  and for the other objects in B there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^{-} \gamma(o).$$

  Now consider those views in which $\gamma(ob)$ is false. Then, we can choose $\binom{a-1}{p \cdot a}$ different subsets of (A/ob) containing $p \cdot a$ objects.
  For each subset, $\mathcal{N}$ must contain a view V_i such that for the objects o in the subset there holds:

  $$\langle O_{\mathcal{N}}, V_i \rangle \models^{+} \gamma(o)$$

and for the other objects in A there holds:

$$\langle O_{\mathcal{N}}, V_i \rangle \models^{-} \gamma(o).$$

Furthermore, we can choose $\binom{b-1}{q \cdot b-1}$ different subsets of (B/ob) containing $(q \cdot b-1)$ objects.
For each subset, $\mathcal{N}$ must also contain a view V_i such that for the objects o in the subset there holds:

$$\langle O_{\mathcal{N}}, V_i \rangle \models^{+} \gamma(o)$$

and for the other objects in B there holds:

$$\langle O_{\mathcal{N}}, V_i \rangle \models^{-} \gamma(o).$$

Therefore,

$$\mathcal{N} = \langle O_{\mathcal{N}}, \{V_1, \ldots V_k\} \rangle$$

with

$$k = \binom{a-1}{p \cdot a-1} \cdot \binom{b-1}{q \cdot b} + \binom{a-1}{p \cdot a} \cdot \binom{b-1}{q \cdot b-1}.$$

In $\binom{a-1}{p \cdot a-1} \cdot \binom{b-1}{q \cdot b}$ of these views V_i, there holds:

$$\langle O_{\mathcal{N}}, V_i \rangle \models^{+} \gamma(ob)$$

while in the other views V_i there holds:

$$\langle O_{\mathcal{N}}, V_i \rangle \models^{-} \gamma(ob).$$

Therefore,

$$Pr(\gamma(ob) \mid \mathcal{M}[\varphi \wedge \psi]) = \frac{\binom{a-1}{p \cdot a-1} \cdot \binom{b-1}{q \cdot b}}{\binom{a-1}{p \cdot a-1} \cdot \binom{b-1}{q \cdot b} + \binom{a-1}{p \cdot a} \cdot \binom{b-1}{q \cdot b-1}} = \frac{1}{1 + \frac{q \cdot (1-p)}{p \cdot (1-q)}}$$

Hence, if $p > q$, then

$$Pr(\gamma(ob) \mid \mathcal{M}[\varphi \wedge \psi]) > \frac{1}{2}.$$

□

The following example has been described by D. S. Touretzky, J. F. Horty and R. H. Thomason [64]. They gave this example as a counter example for off-path pre-emption. The solution they suggested to derive the correct conclusion, is being described in the example below.

Example 8.9 Let John be a marine chaplain. Furthermore, let it be known that the percentage of beer drinking marines is p and that the percentage of beer drinking chaplains is q. According to Theorem 8.8, if $p > (1 - q)$, it is more likely for John to be a beer drinker than not.

8.3 Explanations

An explanation tries to describe a *cause* (a disease, a malfunction) for anomalies (symptoms) observed. There may, however, exist more than one cause that can explain the anomalies observed. Since we are interested in the actual cause, we need a method to discriminate between the possible causes. One possibility to discriminate between the possible causes is to determine their probabilities. To be able to do so, we have to know their a priori probabilities. As was argued by J. T. Nutter [47], it is often not possible to know these probabilities. Furthermore, in a study carried out by A. Tversky and D. Kahneman [65], it was observed that humans do not use a priori probabilities either. Although this cannot be used as an argument for neglecting a priori probabilities, it is an indication that a priori probabilities may not be necessary for explanations. A stronger argument to neglect a priori probabilities is that there are cases in which the use of a priori probabilities can result in wrong decisions. Consider, for example, the situation in which two diseases can explain the same symptoms. If one of the diseases is common and requires no medication, while the other disease is very rare and will kill a patient when no medication is given, then, by using a priori probabilities, the latter disease will never be considered.

Although a priori probabilities are not used here, this does not imply that they cannot be used at all when they are known. When reasoning on a meta level about the likelihood measures, it is still possible to use these a priori probabilities. So, defining a likelihood measure independent of the a priori probabilities, enables us to reason about possible causes with or without using the a priori probabilities.

Instead of using a probability measure, here the compatibility of a possible cause with the current state of knowledge is determined. This means that we have to determine the views in which we cannot believe in the possible cause. This compatibility is expressed by an unlikelihood measure. Now a likely cause for the anomalies observed can be determined by showing that all other possible causes are unlikely. This approach can be viewed as a generalization of the *falsification principle*.

Like the probability measure, the unlikelihood measure of a formula is also determined by considering the set of views of a partial model.

- The unlikelihood measure should be proportional to the number of views in

which the formula is false.

- The measure should be inversely proportional to the total number of views.
- Since we have no reason to prefer one view to another, the insufficient reason argument of Bernoulli and Laplace [22] should be applied on the set of views.

Definition 8.10 Let $\mathcal{M} = \langle O, \{V_1, ..., V_n\}\rangle$ be a partial model. The unlikelihood measure $UL(\varphi \mid \mathcal{M})$ for a formula φ with respect to a partial model $\mathcal{M}$ is defined as follows:

$$UL(\varphi \mid \langle O, \{V_1, ..., V_n\}\rangle) = \frac{1}{n} \cdot \sum_i UL(\varphi \mid V_i)$$

and

$$UL(\varphi \mid \langle O, V_i\rangle) = \begin{cases} 1 & \text{if } \langle O, V_i\rangle \models^{-} \varphi \\ 0 & \text{otherwise.} \end{cases}$$

For this likelihood measure the following observations can be proven.

Observation 8.11 Let $\mathcal{M}$ be a partial model. If $Pr(\varphi \mid \mathcal{M})$ is determined, then

$$Pr(\varphi \mid \mathcal{M}) = 1 - UL(\varphi \mid \mathcal{M}).$$

Using the unlikelihood measure, an efficient diagnostic reasoning process can be realized when we only try to determine one likely cause. For this diagnostic reasoning process an abstraction hierarchy of possible causes (diseases) is needed. The abstraction hierarchy of possible causes is used as a search tree. In this search tree the unlikelihood measure is used as an evaluation function. In an ideal situation we can find a specific cause in $\mathcal{O}(\log n)$ steps, where n denotes the number of possible causes. The worst case is, of course, $\mathcal{O}(n)$ steps. The use of an abstraction hierarchy is not only motivated by the wish to realize an efficient diagnostic reasoning process. In a study of existing expert systems, carried out by W. Clancey [13], it was observed that such a hierarchy is implicitly implemented in these systems.

Given an abstraction hierarchy for the possible causes, the following diagnostic reasoning process can be used to determine a specific cause.

1. Start with the most abstract possible cause.
2. Determine the most abstract refinements of the possible cause.
3. Determine for each refinement the unlikelihood measure.
4. For each refinement which is not proven to be unlikely, i.e. its unlikelihood measure is above some threshold value, one can repeat step 2 until one reaches the specific causes, or until there are no possible causes which are not proven to be unlikely.

The correctness of this reasoning process depends on the following theorem.

Theorem 8.12 The unlikelihood measure for some abstract possible cause $\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle$ is a lower bound for each of its refinements. Here a refinement is either $\langle \mathbf{a}, o, \langle cl', o\rangle, \langle d, o\rangle\rangle$ or $\langle d, inst\rangle$ where cl' subclass of class cl and $inst \in Names$ is an instance of the class cl for any partial model $\mathcal{M}$.

$$UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \mathcal{M}) \leqslant UL(\langle \mathbf{a}, o, \langle cl', o\rangle, \langle d, o\rangle\rangle \mid \mathcal{M})$$

and

$$UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \mathcal{M}) \leqslant UL(\langle d, inst\rangle \mid \mathcal{M})$$

Proof By Definition 8.10:

$$UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \langle O, \{V_1, ..., V_n\}\rangle) =$$

$$\frac{1}{n} \cdot \sum_i UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \langle O, V_i\rangle)$$

and

$$UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \langle O, V_i\rangle) = \begin{cases} 1 & \text{if } \langle O, V_i\rangle \models^{-} \langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \\ 0 & \text{otherwise.} \end{cases}$$

Since any view that satisfies $\langle \mathbf{a}, o, \langle cl', o\rangle, \langle d, o\rangle\rangle$ or $\langle d, inst\rangle$ will also satisfy $\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle$, there holds:

$$UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \mathcal{M}) \leqslant UL(\langle \mathbf{a}, o, \langle cl', o\rangle, \langle d, o\rangle\rangle \mid \mathcal{M})$$

and

$$UL(\langle \mathbf{a}, o, \langle cl, o\rangle, \langle d, o\rangle\rangle \mid \mathcal{M}) \leqslant UL(\langle d, inst\rangle \mid \mathcal{M})$$

□

This theorem confirms our intuition that if a possible cause turns out to be unlikely, then each of its refinements will also be unlikely. For example, if a lung disease is unlikely, tuberculosis will also be. The strategy behind the diagnostic reasoning process was already suggested by B. Chandrasekaran and M. C. Tanner [10]. Here, this strategy is given a sound foundation. Notice that because only lower bounds are used for the unlikelihood measure, one can use linguistic percentages, e.g. 'most', 'many', etc., and linguistic unlikelihood measures, e.g. 'impossible' and 'unlikely', instead of numbers. These linguistic percentages and unlikelihood measures can be linked using a simple table.

9

Evaluation

In this part of my thesis I have proposed an alternative way of looking at a reasoning process, namely to view a reasoning process as a process of constructing a partial model. This way of looking at a reasoning process has some important advantages over the traditional view on a reasoning process. In a traditional reasoning process information is being derived by combining formulas. Here, however, information is extracted from a formula by using it in the construction of a partial model of the world we are reasoning about. When these formulas do not contain quantifiers, the information they describe can be added to a partial model in polynomial time. When, however, quantifiers are involved, there are cases in which the time complexity can become exponential. Here, more research is needed.

Another subject for further research is the definition of the *updating function*, which adds new information to a partial model. In this thesis I have only formulated the conditions that have to be satisfied by such an updating function.

To illustrate the possibilities of this reasoning process, I have shown how uncertain conclusions can be derived from a partial model. Two different certainty measures have been defined, one expressing the *probability* of a formula and one expressing the *unlikelihood* of a formula. The former measure is used for conclusions

expressing an *expectation* and the latter for conclusions expressing an *explanation*. Using the probability measure, pre-emption and multiple inheritance in an inheritance network can be dealt with correctly. Using the unlikelihood measure, an efficient heuristic diagnostic reasoning process can be realized. Since the diagnostic reasoning process described in this thesis is only suited for diagnostic problems in which there is only one cause for the anomalies observed, more research is needed. Furthermore, the combination of heuristic and diagnostic reasoning has to be investigated. In my opinion the reasoning process described here is suited for this.

Compared with other certainty measures in literature, the measures defined here have some important advantages. For example, in the Bayesian probability theory probabilities are either viewed as representing relative frequencies or as representing subjective belief values. In the former view we have to be able to assign a correct a priori probability value to every proposition. As J. T. Nutter remarks, it is not always possible to know these probabilities [47]. In the latter view a probability describes the belief in a proposition of a reasoning agent [11, 12]. There is, however, no inter-subjective interpretation of such a belief value. Hence, there is no reason why two persons with the same knowledge should agree on a belief value assigned to a statement. The certainty measures defined here, however, do not depend on a priori probabilities. Furthermore, the measures defined here have an inter-subjective interpretation. Two agents possessing the same knowledge, i.e. they have the same partial model about the world, will assign the same certainty measures to a formula. Therefore, the measures defined here do not suffer from these drawbacks. Other certainty measures like the belief in Dempster-Shafer theory [56] and the certainty factors in the certainty factor model of EMYCIN [7] do not use a priori probabilities either. Unfortunately, these measures have a subjective but no inter-subjective interpretation.

Two other approaches based on the probability theory are the probabilistic logics of J. Los [40] and of N. J. Nilsson [46]. Both authors define a probability distribution over a set of models. The probability of a formula is defined as the sum of the probabilities of the models that satisfy the formula. This way of defining a probability measure is related with the measures defined in the preceding chapter. This relation also needs to be investigated.

A very important property of a partial model is the decidability of the consistency problem. The consistency problem is not only decidable, but it can be solved in polynomial time as well. This suggests that it may be possible to realize an efficient non-monotonic reasoning process without the process non-monotonicity; i.e. a partial model is always correct with respect to the defaults used to create it. In [53] I have shown that if we limit ourselves to normal default rules only, such a reasoning process is possible. Normal default rules, however, do not possess sufficient expressive power. Therefore, I propose to define a preference relation on normative rules. Unfortunately, lack of time has prevented me from investigating this idea in detail. Therefore, it is not reported in this thesis.

Despite of the fact that there are a number of important problems that have to

be solved yet, in my opinion it is an important new idea to view a reasoning process as a process of constructing a partial model. Hopefully, others will share this opinion and will start to investigate this idea.

References

[1] Attardi, G., Simi, M., Consistency and completeness of OMEGA, a logic for knowledge representation, *Proceedings of the 7-th IJCAI* (1981) 504-510.

[2] Bacchus, F., A modest, but semantically well founded inheritance reasoner, *IJCAI-89* (1989) 1104-1109.

[3] Besnard, P., *A Introduction to Default Logic*, Springer-Verlag, Berlin (1989).

[4] Bosch, K. O. ten, Redeneren met de preferentielogica, Master thesis (1990).

[5] Brewka, G., The logic of inheritance in frame systems, *IJCAI-87* (1987) 483-488.

[6] Brewka, G., Preferred subtheories: an extended logical framework for default reasoning, *IJCAI-89* (1989) 1043-1048

[7] Buchanan, B. G., Shortliffe, E. H., *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project,* Addison-Wesley Publishing Company (1984).

[8] Carnap, R., *Logical foundations of probability,* The University of Chicago Press, Chicago (1950).

[9] Chandrasekaran, B., Generic tasks in knowledge based reasoning; high-level building blocks for expert system design, *IEEE EXPERT* (1986) 23-30.

[10] Chandrasekaran, B., Tanner, M. C., Uncertainty handling in expert systems, in: Kanal, L. N., Lemmer, J. F. (eds), *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam (1986) 35-46.

[11] Charniak, E., The Bayesian basis of common sense medical diagnosis, *AAAI-83* (1983) 70-73.

[12] Cheeseman, P., In defense of probability, *IJCAI-85* (1985) 1002-1009.

[13] Clancey, W. C., Classification problem solving, *AAAI-84* (1984) 49-55.

[14] Clancey, W,. C., Heuristic classification, *Artificial Intelligence* **27** (1985) 289-350.

[15] Delgrande, J. P., First-order conditional logic for prototypical properties, *Artificial Intelligence* **33** (1987) 105-130.

[16] Delgrande, J. P., An approach to default reasoning based on first-order conditional logic: revised report, *Artificial Intelligence* **36** (1988) 63-90.

[17] Doyle, J., A truth maintenance system, *Artificial Intelligence* **12** (1979) 231-272.

[18] Enderton, H, B., *A mathematical introduction to logic*, Academic Press, New York (1972).

[19] Etherington, D. W., Reiter, R., On inheritance hierarchies with exceptions, *AAAI-83* (1983) 104-108.

[20] Etherington, D. W., A semantics for default logic, *IJCAI-87* (1987) 495-498.

[21] Etherington, D. W., Brogida, A., Brachman, R. J., Kautz, H., Vivid knowledge and tractable reasoning, *IJCAI-90* (1989) 1146-1152.

[22] Fine, T. N., *Theories of probability*, Academic Press, New York (1973).

[23] Gärdenfors, P., *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Bradford Books, MIT Press, Cambridge MA (1988).

[24] Goodwin, J. W., *A theory and system for non-monotonic reasoning*, Department of Computer and Information Science, Linköping University, Linköping, Sweden (1987).

[25] Hayes, P. J., In defence of logic, *IJCAI-77* (1977) 559-565.

[26] Hanks, S., McDermott, D., Nonmonotonic logic and temporal projection, *Artificial Intelligence* **33** (1987) 379-412.

[27] Hewitt, C., Attardi, G., Simi, M., Knowledge embedding in the description system OMEGA, *Proc. of First National Annual Conference on Artificial Intelligence* (1980) 157-163.

[28] Horty, J. F., Thomason, R. H., Touretzky, D. S., A skeptical theory of inheritance in non-monotonic semantics networks, *AAAI-87* (1987) 358-363.

[29] Johnson-Laird, P. N., *Mental models, Toward a cognitive science of language inferences and consciousness*, Cambridge University Press, Cambridge (1983).

[30] Kamp, H., A theory of truth and semantic representation,
in: Groenendijk, J. A. G., Jansen, T. M. V., Stokhof, M. B. J. (eds), *Formal methodes in the study of language*, Mathematical Centre Tracts, Amsterdam (1981) 277-322.

[31] Kleer, J. de, An assumption based TMS, *Artificial Intelligence* **28** (1986) 127-162.

[32] Konolige, K., On the relation between default logic and autoepistemic logic, *Artificial Intelligence* **35** (1988) 343-382.

[33] Kraus, S., Lehmann, D., Magidor, M., Nonmonotonic reasoning, preferential models and cumulative logics, *Artificial Intelligence* **44** (1990) 167-207.

[34] Krishnaprasad, T., Kifer, M., An evidential framework for a theory of inheritance, *IJCAI-89* (1989) 1093-1098.

[35] Krishnaprasad, T., Kifer, M., Warren, D. S., On the declarative semantics of inheritance networks, *IJCAI-89* (1989) 1099-1103.

[36] Kyburg, H. E., Objective probabilities, *AAAI-83* (1983) 902-904.

[37] Lifschitz, V., Computing circumscription, *IJCAI-85* (1985) 121-127.

[38] Lifschitz, V., Piontwise circumscription, *AAAI-86* (1986) 406-410.

[39] Lloyd, J. W., *Foundations of Logic Programming*, Springer-Verlag, Berlin, (1984).

[40] Los, J., Semantic representation of the probability of formulas in formal theories, *Studia Logica* **14** (1963) 183-196.

[41] McCarthy, J., Circumscription - a form of non-monotonic reasoning, *Artificial Intelligence* **13** (1980) 27-39.

[42] McCarthy, J., Applications of circumscription, *Artificial Intelligence* **28** (1986) 89-116.

[43] McDermott, D.,Doyle, J., Non-montonic logic I, *Artificial intelligence* **13** (1980) 41-72.

[44] Moore, R. C., Semantical considerations on non-monotonic logic, *Artificial Intelligence* **25** (1985) 75-94.

[45] Moore, R. C., Autoepistemic logic, in: Smets, Ph., Mamdani, E., H., Dubois, D., Prade, H. (eds), *Non-Standard Logics for Automated Reasoning*, Academic Press, London (1988) 105-136.

[46] Nilsson, N. J., Probabilistic logic, *Artificial Intelligence* **28** (1986) 71-87.

[47] Nutter, J. T., Uncertainty and probability, *IJCAI-87* (1987) 373-379.

[48] Poole, D., A logical framework for default reasoning, *Artificial Intelligence* **36** (1988) 27-47.

[49] Reiter, R., A logic for default reasoning, *Artificial Intelligence* **13** (1980) 81-132.

[50] Reiter, R., Criscuolo, G., On interacting defaults, *Proceedings of the 7th IJCAI* (1981) 270-276.

[51] Rescher, N., *Hypothetical Reasoning* North-Holland Publishing Company, Amsterdam (1964).

[52] Roos, N., *A preference logic for non-monotonic reasoning*, Report 88-94, Department of computer science, Delft University of Technology, Delft, the Netherlands (1988).

[53] Roos, N., Redeneren met behulp van normatieve regels, *NAIC'90* (1990) 241-248.

[54] Roos, N., *How to reason with uncertain knowledge*, to appear in a volume of series Lecture Notes in Computer Science of Springer Verlag, Springer Verlag, Berlin (1991).

[55] Sandewall, E., A functional approach to non-monotonic logic, *IJCAI-85* (1985) 100-106.

[56] Shafer, G., *A mathematical theory of evidence*, Princeton University Press, Princeton, London (1976).

[57] Shastri, L., Default reasoning in semantic networks: a formalization of recognition and inheritance, *Artificial Intelligence* **39** (1989) 283-355.

[58] Shoham, Y., *Reasoning about change*, Dissertation Yale University, Department of computer science (1986) 85-109.

[59] Shoham, Y., Non-monotonic logic: meaning and utility, *IJCAI-87* (1987) 388-393.

[60] Smets, P., Transferable belief model versus bayesian model, *ECAI-88* (1988) 495-500.

[61] Stein, L. A., Skeptical inheritance: computing the intersection of credulous extensions, *IJCAI-89* (1989) 1153-1158.

[62] Touretzky, D. S., Implicit ordering of defaults in inheritance systems, *AAAI-84* (1984) 322-325.

[63] Touretzky, D. S., *The mathematics of inheritance systems*, Pitman, London, Morgan Kaufman Publishers Inc., Los Altos California (1986).

[64] Touretzky, D. S., Horty, J. F., Thomason, R. H., A clash of intuitions: The current state of non-monotonic multiple inheritance systems, *IJCAI-87* (1987) 476-482.

[65] Tversky, A., Kahneman, D., Judgement under uncertainty: heuristics and biases, In: Wendt, D., Vlek, C. (eds), *Utility, probability and human decision making*, D. Reidel Publishing Company, Dordrecht Netherlands (1982) 141-162.

Curriculum vitae

The author was born on the may the 3-rd 1959 in Rotterdam.
In 1978 he started his study for electrical engineer at the HTS in Breda. He completed it successfully in 1982.
The same year he started his master study in computer science at the Delft University of Technology (TU-Delft). In 1987 he completed this study.
From February 1987 he worked in a research project of the section of theoretical computer science at the TU-Delft and of the section of computer science at the National Aerospace Laboratory (NLR). This thesis reports on the results of the research carried out during the four years of the project.