REPORT 89-53

Preference logic:
a logic for reasoning with inconsistent knowledge

Nico Roos

TU Delft

Technische Universiteit Delft
Delft University of Technology

Faculteit der Technische Wiskunde en Informatica
Faculty of Technical Mathematics and Informatics

DELFT UNIVERSITY OF TECHNOLOGY

REPORT 89-53

Preference logic: a logic for reasoning with inconsistent knowledge

Nico Roos

ISSN 0922-5641

Reports of the Faculty of Technical Mathematics and Informatics no. 89-53

Delft 1989

Abstract

In many situations humans have to reason with inconsistent knowledge. These inconsistencies may occur due to not fully reliable sources of information or the use of general information like defaults. To be able to reason with inconsistent knowledge it is not possible to view a set of premisses as absolute truths as is done in predicate logic. Viewing a set of premisses as a set of assumptions, however, makes it possible to deduce useful conclusions from an inconsistent set of premisses. In this paper a preference logic for reasoning with inconsistent knowledge is described. This logic is based on the work of N. Rescher [16]. In this logic a preference relation is used to choose between incompatible assumptions. These choices are only made when a contradiction is derived. As long as no contradiction is derived, the knowledge is assumed to be consistent. This makes it possible to define an executable deduction process for the preference logic. As a special case of reasoning with inconsistent knowledge, the use of default rules is considered. A default rule can be described in the preference logic by an implication and appropriate preference relation. Hence the preference logic enables default reasoning with a deduction process. For the preference logic a semantics is defined. This semantics is based on the ideas of Y. Shoham [18, 19]. The models for the preference logic are those structures which satisfy more premisses with a higher preference than some other structure.

Contents

1 Introduction

Reasoning with inconsistent knowledge is very common in human reasoning. Sometimes inconsistencies may occur because not all sources of information are completely reliable. Different observations may result in different conclusions of what is observed. Also inconsistencies may occur if general rules are used. General rules are the result of some induction process. There may however exist specific cases, for which a general rule does not hold. In these cases the general rule contradicts with the specific information, causing a inconsistency. A special case of general knowledge are default rules.

1.1 Inconsistent knowledge

Reasoning with inconsistent knowledge can be traced back to Rescher [16]. In his book *Hypothetical Reasoning*, N. Rescher introduces *Preferred Maximal Mutually Compatible* subsets of an inconsistent set of premisses. These *PMMC* subsets are consistent subsets of the set of premisses, are preferred according to some condition and become inconsistent by adding more premisses to it. The conclusions which may be drawn from the premisses are the conclusions which follow from every *PMMC* subset. These conclusion are said to be *compatible-subset entailed.* The preference of some maximal consistent subset is based on a division of the premisses into modal categories. This division can be based on alethic modalities, epistemic modalities or probabilistic modalities. Rescher introduces this machinery to describe hypothetical (counterfactual) reasoning. Unfortunately he does not specify a deduction process for his logic.

An approach to deal with uncertain inconsistent knowledge is the system *Ponderosa* of J. R. Quinlan [13]. In this system every premiss has a probability value assigned to it. When selecting a consistent subset of the premisses the system tries to minimize the risk of removing the wrong premiss. A disadvantage of this approach is that one has to specify a probability value for every premiss.

A recent approach which deals with inconsistent knowledge is *a framework for default reasoning* of D. Poole [12]. Poole argues that non-monotonic reasoning should be viewed as 'scientific theory formation'. From a set of hypotheses a consistent subset has to be selected which together with a set of facts can explain some closed formula. To determine an explanation one has to be able to carry out a consistency check. This makes it impossible to apply the framework of Poole when using the full predicate logic.

1.2 Default reasoning

Over the past few years different approaches for default reasoning have been developed. Some of the formally sound approaches are *Circumscription* of J. McCarthy [10], *Default logic* of R. Reiter [14], *A Conditional logic* of J. P. Delgrande [1, 2] and *A logical framework for default reasoning* of D. Poole [12]. Each of these of these approaches have some drawbacks which can not be solved.

Circumscription is a theoretical elegant approach. To reason with *circumscription* one has to be able to collapse the second order *circumscription* formula into a first order formula. Since the completeness theorem does not hold for *circumscription* theories, there exist no general way to do this.

In the *Default logic* default rules can be formulated explicitly. The disadvantage of this logic is that it does not have a deduction process. This disadvantage can be overcome by combining *default logic* with the *Truth Maintenance System* of J. Doyle [3]. In combination with *TMS* an extension of a default theory can be approximated. *TMS* does not make things better. The labeling problem of a *TMS* graph is proven to be *NP-complete.*

In the *Conditional logic* of J. P. Delgrande new default rules can be derived. In [2] he describes how to reason with these default rules. To do this one has to be able to determine consistency. This, however, is an undecidable problem.

The *logical framework for default reasoning* of D. Poole which is already discussed in the previous section, is actually meant to model default reasoning. In this framework default reasoning is viewed as an special case of reasoning with inconsistent knowledge.

2 The preference logic

To be able to reason with inconsistent knowledge in the preference logic premisses are considered as assumptions. These assumptions are considered to be true as long as no contradiction is derived from them. When, however, a contradiction is derived, one of the assumptions on which the contradiction is based has to be removed. The question is which assumption has to be removed. To select the premiss to be removed, a preference relation which is a strict partial ordering on the set of premisses, can be defined. If a contradiction is derived the set of premisses on which the contradiction is based, has to be determined. Using the preference relation a least preferred premiss is removed from this inconsistent set thereby blocking the derivation of the contradiction.

Example 1 *Let* Σ *denote a set of premisses,*

$$\Sigma = \{\varphi, \varphi \rightarrow \psi, \neg\psi\}$$

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

$$(\Sigma, \prec) = \{\neg\psi \prec \varphi, \neg\psi \prec \varphi \rightarrow \psi\}$$

Since $\neg\psi$ *is the least preferred premiss on which the inconsistency in* Σ *is based,* $\neg\psi$ *has to be removed.*

There are three problems which can arise when trying to remove an inconsistency.

- Firstly, one has to be able to determine the premisses on which an inconsistency is based. To solve this problem justifications are introduced. Such a justification which is called an *in_justification*, describes the premisses from which a formula is derived.

- Secondly, a premiss which has been removed, may have to be placed back because the contradiction causing his removal can not occur any more. This may happen due to some other contradiction which will be derived. To solve this problem an other 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.

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

  - Do nothing. The contradiction is not solved but this does not have to mean that the contradiction will not be solved at all [17].

- Remove all the minimal premisses on which a contradiction is based. Due to this policy it is possible that too much is removed from the set of premisses.
- Consider the results of the removal of every alternative apart. As a result of this policy different subsets of the set of premisses are considered. It is possible that these subsets will converge to one consistent subset of the set of premisses. If this happens the result of all three approaches will be the same.

There is a relation between the last two approaches. Result of the second approach is equal to the intersection of the set of premisses which are the result of the last approach. Since the result of the second approach can be determined in polynominal time, it can be used as an approximation of the last approach. It determines what certainly follows from the last approach.

As already was mentioned in the introduction, default reasoning can be treated as a special case of reasoning with inconsistent knowledge. Default rules are general rules which may contradict specific information. When this occurs specific information has to preferred above general information.

Example 2 *The specific information 'Tweety can not fly because it is a penguin' should be preferred above 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 who can not fly this premiss will be removed making it impossible to derive for any bird that it can fly. Since this is undesirable, an alternative approach for representing defaults is introduced here. In the preference logic an alternative semantics is given to a formula containing free variables. In the predicate logic a formula φ containing free variables $\overline{x}$ is equivalent to $\forall\overline{x}\varphi$. In the preference logic a formula φ cantaining free variables is interpreted as denoting a set of instances of this formula. To make it possible to derive new default rules, this set of instances is not limited to ground instances only.

Hence, a premiss containing free variables denotes a set of premisses. 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 3 *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 above the instance of the first premiss:

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

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

A question which 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 example.

Example 4 *Suppose 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 the set of premisses in the example is inconsistent. Now suppose that the second premiss is preferred above the first, then the whole set of premisses denoted by the first premiss has to be removed. Because with each instance of the first premiss a contradiction can be derived, the second premiss has to be preferred above 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 1 *Every instance of a premiss containing free variables should have the same preference as this premiss.*

3 Formal definitions

In the formal description of the preference logic, unification will be used [9]. To unify two formulas a substitution of terms for free variables may be required Such a substitution θ for the free variables is denoted placing $[\theta]$ behind a formula. Substitution which has to be carried out on every formula of a set of formulas or on every formula occurring in a justification, are 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 for a set of premisses Σ is defined as 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 $\overline{\Sigma}$ which also contains all instances, is introduced.

Definition 1 *Let S be a set of formulas. By $\overline{S}$ an extended set of formulas is denoted, which also contains all instances of the formulas of S.*

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

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 1 and should again be a strict partial order. The preference relation for the extended set of premisses is defined as:

Definition 2 *Let $(\overline{\Sigma}, \prec)$ denote the preference relation for $\overline{\Sigma}$. $(\overline{\Sigma}, \prec)$ is the smallest strict partial order containing $(\Sigma, \prec)$ which is closed under term substitution in the premisses of $\overline{\Sigma}$.*

One should notice that the preference relation $(\overline{\Sigma}, \prec)$ is not always defined as can be seen in the following example.

Example 5
$$\begin{array}{rcl} \Sigma & = & \{\varphi(x), \varphi(a), \psi\} \\ (\Sigma, \prec) & = & \{\varphi(x) \prec \psi, \psi \prec \varphi(a)\} \end{array}$$

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 3
$$\begin{aligned} \mathit{In\_Just} &= \{P \Rightarrow \varphi \mid P \subset \overline{\Sigma} \textit{ and } \varphi \in L\} \\ \mathit{Out\_Just} &= \{P \not\Rightarrow \varphi \mid P \subset \overline{\Sigma} \textit{ and } \varphi \in \overline{\Sigma}\} \end{aligned}$$

4 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, *Reason (Truth) Maintenance* is integrated in the deduction process. Therefore the preference logic can be viewed as a process logic. A deduction in the preference logic is a process of belief revision which occurs due to the addition of new justifications. See also [5]. This deduction process will finally terminate with a belief set which is the theory of the models of the set of premisses and the preference relation. How these model are defined can be found in section 6.

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

Definition 4 $J_0 = \{\{\varphi\} \Rightarrow \varphi \mid \varphi \in \Sigma\}$

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 which is used. In the following description of the preference logic it is assumed that an axiomatic deduction system for a language L which only contains the logical operators $\rightarrow$ and $\neg$ and the quantor $\forall$, is used. The following axiom scheme will be used.

1. Tautologies
2. $\forall x \varphi(x) \rightarrow \varphi[\theta]$ where θ denotes a substitution for x
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

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

Rule 1 *An instance φ of the an axiom scheme gets an in_justification $\emptyset \Rightarrow \varphi$.*

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 *Let the formulas φ and $\psi \rightarrow \mu$ have an in-justification of respectively $P \Rightarrow \varphi$ and $Q \Rightarrow (\psi \rightarrow \mu)$. If φ and ψ can be unified with a the 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 3 *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 the most general unifier. If $R = min(P \cup Q)$ —minimal under the preference relation $(\overline{\Sigma}, \prec)$—, then each premiss $\eta \in R$ gets an out_justification $(((P \cup Q)/\eta) \not\Rightarrow \eta)[\theta]$.*

It is assumed that the process which creates a set of justifications J_{k+1} from the set J_k is fair. This means that this process does not forever defer the addition of some possible justification to the set of justifications. If a fair process is used, the following theorems hold. The first theorem guarantees the soundness of the in_justifications. This means that for every in_justification there exists a corresponding deduction from the premisses in the antecedent to the formula in the consequent. The second theorem guarantees the completeness of the in_justifications. This means that for each deduction of a formula from the premisses there exists a corresponding in_justification. Finally the third and the fourth theorem guarantee respectively the soundness and the completeness of the out_justifications.

Theorem 1
For each $i \geq 0$: if $P \Rightarrow \varphi \in J_i$, then for each substitution θ: $P[\theta] \subseteq \overline{\Sigma}$ and $P[\theta] \vdash \varphi[\theta]$.

Theorem 2
For each $P \subseteq \overline{\Sigma}$: if $P \vdash \varphi$, then for some $i \geq 0$: $Q \Rightarrow \psi \in J_i$ and for some substitution θ: $Q[\theta] \subseteq P$ and $\psi[\theta] = \varphi$.

Theorem 3
For each $i \geq 0$: 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 inconsistent and for each $\psi \in P[\theta]$ there holds: $\psi \not\prec \varphi[\theta]$.

Theorem 4
For each $P \subseteq \overline{\Sigma}$ if P is a minimal inconsistent set and $Q = min(P)$, then for some $i \geq 0$ there holds for each $\varphi \in Q$: $R \not\Rightarrow \psi \in J_i$, and for some substitution θ: $P/\varphi = R[\theta]$ and $\varphi = \psi[\theta]$.

Given a set of justifications there may exists one or more subsets of the set of premisses which can be believed. Such a subset contains the premisses which do not have to be withdrawn due to an out-justification. Since the out-justifications which can be applied depend on the set of premisses which are not withdrawn, the following fixed point definition should be satisfied.

Definition 5 *Let $\mathcal{A}_i$ denote the set containing all the subsets of the premisses which can be believed given the out_justifications in J_i.*

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

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

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

Definition 6 *Let J_i be a set of justifications and $\mathcal{A}_i$ be the corresponding sets of premisses. The set of formulas B_i which can be believed is defined as:*

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

Observation 1 *For each $\varphi \in B_i$: $[\Delta \vdash \varphi$ for each $\Delta \in \mathcal{A}_i]$*

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

Definition 7 $J_\infty = \bigcup_{i\to\infty} J_i$

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

Observation 2 *For each* $\Delta \in \mathcal{A}_\infty$*:* $\Delta \subseteq \overline{\Sigma}$.

Observation 3 *For each* $\Delta \in \mathcal{A}_\infty$*:* Δ *is maximal consistent.*

Observation 4 *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 an other subset in which it is not a minimal element, then* $|\mathcal{A}_\infty| = 1$.

Observation 5

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

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

5 Determination of the belief set

In this section two algorithms are described. The first algorithm determines a single set Δ from $\mathcal{A}_i$ given the justifications J_i. The second algorithm determines the intersection —$\bigcap \mathcal{A}_i$— of all the sets of $\mathcal{A}_i$. The result of this algorithm can be used as an approximation of the set $\mathcal{A}_i$. Both algorithms are polynomial time algorithms.

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 $\psi \prec \varphi$;
$delta$:= $prem$;
repeat
  $P \not\Rightarrow \varphi \in max(min\_out\_just)$;

$min\_out\_just := min\_out\_just - \{P \not\Rightarrow \varphi\}$;
if $P \subseteq delta$
then $delta := delta - \{\psi \mid \psi$ is an instance of φ and $\psi \in prem\}$;
until $min\_out\_just = \emptyset$;
return $delta$;
end.

The algorithm above determines a set *delta* such that: $\overline{delta} \in \mathcal{A}_i$. It is not difficult to modify the algorithm so that it determines every element of $\mathcal{A}_i$. The following algorithm determines a set $delta^*$ such that: $\overline{delta^*} = \bigcap \mathcal{A}_i$.

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 $\psi \prec \varphi$;
$delta^* := prem$;
repeat
$R := \emptyset$;
$S := max(min\_out\_just)$;
$min\_out\_just := min\_out\_just - S$;
repeat
$P \not\Rightarrow \varphi \in S$;
$S := S - \{P \not\Rightarrow \varphi\}$;
if $P \subseteq delta^*$
then $R := R \cup \{\psi \mid \psi$ is an instance of φ and $\psi \in prem\}$;
until $S = \emptyset$;
$delta^* := delta^* - R$;
until $min\_out\_just = \emptyset$;
return $delta^*$;
end.

6 The semantics for the logic

The semantics of the preference logic is based on the ideas of Y. Shoham [18, 19]. In [18, 19] Shoham argues that the difference between monotonic logic and non-monotonic logic is a difference in the definition of the entailment relation. In 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 the work of Shoham. 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 every structure of the language is defined. Given such a preference relation on the structures, the models of preference logic are the most preferred structures. Hence, an appropriate preference relation on the structures has to be defined. In the preference logic a structure which satisfies more premisses with a higher preference ($\prec$) than some other structure, is preferred ($\sqsubset$) above this structure.

In the preference logic choices between premisses are made 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 above 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 which belongs by this choice should also satisfy Condition 1. 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 1. The following definitions describe this formally.

Definition 8 *Let $Prem(\mathcal{M})$ denote the subset of the premisses $\overline{\Sigma}$ which are satisfied by the model $\mathcal{M}$.*

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

Definition 9 *Let Str denote the set of structures for the language L and let $(Str, \sqsubset)$ denote a preference relation on these structures. For each 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 $\varphi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$, there is a $\psi \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$*

such that for every linear extension of $(\overline{\Sigma}, \prec)$ *which satifies Condition 1:* $\varphi \prec \psi$ *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 10 *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 which guarantees the soundness and the completeness of the preference logic, holds:

Theorem 5

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

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

7 Related work

In this section some related approaches are discussed.

7.1 Hypothetical reasoning

The preference logic which is presented in this paper is closely related with the work of N. Rescher [16]. The modal categories which Rescher introduces can be expressed with a partial preference relation. When these modal categories are described by a preference relation, the sets Δ of $\mathcal{A}_\infty$ are equal to the *Preferred Maximal Mutually Compatible* subsets of Rescher. Further the *compatible-subset entailed* formulas of Rescher are the formulas of B_∞.

7.2 A framework for default reasoning

The preference logic is also related with the work of D. Poole [12]. Poole introduces two sets of premisses, facts and hypothesis. The set of facts is always consistent and can not be removed. The set of hypothesis, however, may be inconsistent. Further a hypothesis may contain free variables. Each hypothesis which contains a free variable denotes a set of 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 of Poole can be represented in the preference logic by preferring each fact above each hypothesis. A consistent set of hypothesis which in the framework of Poole can explain some closed formula, corresponds with set Δ of $\mathcal{A}_\infty$ which entails this closed formula.

Although the framework of Poole can be expressed in the preference logic, one should realize that the philosophy behind the two approaches are quite differently. The work of Poole is based on the idea that default reasoning is a process of selecting consistent sets of hypothesis 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 assumption may change due to new information.

In the framework of Poole constraints can be added to denote that some set of hypotheses may not be used as an explanation. What these constraints denote is that some explanations are preferred above others. This is realized by making the latter explanation inconsistent through the addition of the constraints. Because constraints are implemented with formulas which may not be used in an explanation, in the opinion of the author constraints are rather ad hoc.

As was argued above the framework of Poole without constraints can be modeled in the preference logic. When constraints are added, one denotes that some explanations are preferred above others. Since in the preference logic a preference relation on the premisses generates a preference relation on consistent subsets of the premisses, one can ask if the converse also holds. Unfortunately the answer is 'no'. This means that not every ordering of explanations in the framework of Poole can be modeled using the preference logic. If an ordering on the explanations which can not be modeled, make sense, is something which has to be investigated.

7.3 Default logic

There also exist a relation between the Default logic of R. Reiter [14] and the preference logic. If a default theory only contains normal default rules, this default theory can be described in the preference logic. To describe such a default theory in the preference logic, one has to replace every default rule

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

by a premiss containing free variables

$$\varphi(\overline{x}) \rightarrow \psi(\overline{x})$$

For the default 'Birds can fly' this becomes:

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

Further default rules are given a lower preference than the other premisses. This assures that an instance of a default rule is withdrawn when it contradicts specific information. When a default theory which contains only normal defaults is described in this way with the preference logic, the set of extensions of the default theory is equal to the set $\{Th(\Delta) \mid \Delta \in \mathcal{A}_\infty\}$.

Example 6

premisses:
1. $Bird(Donald)$
2. $Bird(Woody)$
3. $Lazy\_duck(Donald)$
4. $Bird(x) \rightarrow Can\_fly(x)$
5. $\forall x[Lazy\_duck(x) \rightarrow \neg Can\_fly(x)]$

preference relations: $1 \succ 4$, $2 \succ 4$, $3 \succ 4$ *and* $5 \succ 4$

conclusions: $Can\_fly(Woody)$
$\neg Can\_fly(Donald)$

In the preference logic new default rules can be deduced. This is illustrated using an example of Delgrande [1].

Example 7

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

preference relation: $1 \prec 2$

conclusion: $Raven(x) \wedge \neg Albino(x) \rightarrow Black(x)$

In the example above the second premiss is preferred above the first because the first is more general than the second.

It is also possible to derive new defaults using a transitive relation between premisses. If the premiss 'every eagle is a bird' is added to Example 6,

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

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

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

This default can be derived form the premisses 'every eagle is a bird' and 'birds can fly'.

In the deduction of the the default 'eagles can fly', a transitive relation between a default and an implication is derived. 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 [15] argued that it is not enough to have only normal defaults. They argue that semi normal defaults are required. A semi normal default is used to describe a default with exceptions on the application of this default. This makes it possible to avoid unwanted transitive relations. The preference logic does not have something equivalent to a semi normal default. In this logic unwanted transitive relations are avoided by the addition of new defaults and by defining the right preference relation.

Example 8

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

From these two sentences one can conclude 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* $3 \succ 2$

7.4 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 [20] inheritance networks can be modeled by using only normal default rules and by defining a correct ordering on these default rules. The ordering which Touretzky specifies, models his inferential distance algorithm [21]. 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 [4]. The ordering Touretzky specifies [20], selects this extension. Since normal default theories can be transformed into the preference logic, and since preferences between default rules can be specified in the preference logic, the preference logic can be used to model inheritance networks in which the inferential distance algorithm is used.

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

Definition 11 *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 9

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

preference relation: $2 \prec 3$

In [7] J. F. Horty, R. H. Thomason and D. S. Touretzky introduce an alternative approach to inheritance networks. This approach can not be modeled with the preference logic. In their approach two conflicting paths can neutralize each other. This may enable other inheritance paths which otherwise were not possible. Conflicting paths which neutralize each other, can not be modeled using the preference logic.

7.5 Truth maintenance systems

In preference logic justifications are introduced. Unlike the justification which occur in the JTMS of J. Doyle [3] or the ATMS of J. de Kleer [8], the justifications in the preference logic are a part of the logic. In this logic the justifications follow directly from the requirement for a deduction process (section 2). Also the justifications are different from the ones introduced by Doyle and de Kleer. Both Doyle and de Kleer introduce local justifications while in the preference logic only global justifications are used. The in_justifications of the preference logic can be compared with the labels which de Kleer introduces in the ATMS [8]. Like a label an in_justification describes from which premisses a formula is derived. An out_justification has 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 nogood set an out_justification describes which element has to be removed from the set of premisses (assumptions). Something like an non-monotonic justification as are used in the JTMS of Doyle does not occur in the preference logic.

8 The Yale shooting problem

In this section it is shown how the Yale shooting problem can be solved. Since this problem can not be solved by most approaches for default reasoning, it illustrates that the preference logic possesses more expressive power than these other approaches.

In [6] Hanks and McDermott describe a temporal projection problem for which they showed that the non-monotonic logic they considered are to weak to model it. They specified their problem in a situation calculus which has been reformulated in 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 these 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, one has to choose which instance has to be removed. Hence $\mathcal{A}_\infty$ will contain two set of premisses, one from which $T(Dead, S_3)$ and one from which $T(Alive, S_3)$ can be derived. About the same problem arise when some other form of non-monotonic reasoning is used. Hanks and McDermott suggested the following solution [6, 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. This solution can be realized in the preference logic by specifying a preference relation on the instances of the default rule. By preferring one instance above another instance when the time constant in the former is lower than the time constant in the latter, abnormality will occur as late as possible. The Yale shooting problem can be solved by extending the preference relation:

Definition 12 *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* $\varphi(f, e, Result(e, s))$*:*
  $\varphi(f, e, s) \succ \varphi(f, e, Result(e, s))$.

9 Conclusion

The preference logic presented in this paper generalizes the work of N. Rescher in four different ways. Firstly instead of a linear order of modal categories, a partial preference relation is used. Secondly, the preference logic has a deduction process. Thirdly, a semantics is defined for the logic. Fourthly, due to the new interpretation of formulas containing free variable, default reasoning can be modeled with this logic.

At this moment it is an open question if the preference logic is also a generalization of the framework for default reasoning of D. Poole. To answer this question, one has to determine first which ordering of the explanation in the framework of Poole make sense. It will of course be nice if the ordering which make sense are those which can be describe by specifying a preference relation on set of hypotheses.

Acknowledgement

I want to thank C. Witteveen for reading the early drafts of this paper an for providing them with critical comments. Further I want to thank H. Hellendoorn for proofreading the final version.

References

[1] Delgrande, J. P., First-order conditional logic for prototypical properties, *Artificial Intelligence* **33** (1987) 105-130.

[2] Delgrande, J. P., An approach to default reasoning based on first-order conditional logic: revised report, *Artificial Intelligence* **36** (1988) 63-90.

[3] Doyle, J., A truth maintenance system, *Artificial Intelligence* **12** (1979) 231-272.

[4] Etherington, D. W., More on inheritance hierarchies with exceptions; default theories and inferential distance, *AAAI-87* (1987) 352-357.

[5] Goodwin, J. W., *A theory and system for non-monotonic reasoning*, Department of Computer and Information Science, Linköping University, Linköping, Sweden (1987).

[6] Hanks, S., McDermott, D., Nonmonotonic logic and temporal projection, *Artificial Intelligence* **33** (1987) 379-412.

[7] Horty, J. F., Thomason, R. H., Touretzky, D. S., A skeptical theory of inheritance in non-monotonic semantics networks, *AAAI-87* (1987) 358-363.

[8] Kleer, J. de, An assumption based TMS, *Artificial Intelligence* **28** (1986) 127-162.

[9] Lloyd, J. W., *Foundations of Logic Programming*, Springer-Verlag, Berlin, (1984).

[10] McCarthy, J., Circumscription - a form of non-monotonic reasoning, *Artificial Intelligence* **13** (1980) 27-39.

[11] McCarthy, J., Applications of circumscription, *Artificial Intelligence* **28** (1986) 89-116.

[12] Poole, D., A logical framework for default reasoning, *Artificial Intelligence* **36** (1988) 27-47.

[13] Quinlan, J. R., Internal consistency in plausible reasoning systems, *New Generation Computing* **3** (1985) 157-180.

[14] Reiter, R., A logic for default reasoning, *Artificial Intelligence* **13** (1980) 81-132.

[15] Reiter, R., Criscuolo, G., On interacting defaults, *Proceedings of the 7th IJCAI* (1981) 270-276.

[16] Rescher, N., *Hypothetical Reasoning* North-Holland Publishing Company, Amsterdam (1964).

[17] Roos, N., *A preference logic for non-monotonic reasoning*, Report 88-94, Department of computer science, Delft University of Technology, Delft, the Netherlands (1988).

[18] Shoham, Y., *Reasoning about change*, Dissertation Yale University, Department of computer science (1986) 85-109.

[19] Shoham, Y., Non-monotonic logic: meaning and utility, *IJCAI-87* (1987) 388-393.

[20] Touretzky, D. S., Implicit ordering of defaults in inheritance systems, *AAAI-84* (1984) 322-325.

[21] Touretzky, D. S., *The mathematics of inheritance systems*, Pitman, London, Morgan Kaufman Publishers Inc., Los Altos California (1986).

Appendix

Theorem 1 *For each $i \geq 0$: if $P \Rightarrow \varphi \in J_i$, then for each substitution θ: $P[\theta] \subseteq \overline{\Sigma}$ and $P[\theta] \vdash \varphi[\theta]$.*

Proof The theorem will be proven by induction on the index of the sets of justifications. For $i = 0$: $\{\varphi\} \Rightarrow \varphi \in J_0$ if and only if $\varphi \in \Sigma$. Since $\overline{\Sigma}$ is closed under termsubstitution, for each substitution θ: $\varphi[\theta] \in \overline{\Sigma}$. Therefore for each substitution θ: $\{\varphi[\theta]\} \vdash \varphi[\theta]$.

Proceeding inductively, suppose that $P \Rightarrow \varphi \in J_{k+1}$. $P \Rightarrow \varphi \in J_{k+1}$ if and only if $P \Rightarrow \varphi \in J_k$ or $P \Rightarrow \varphi$ has been added by Rule 1 or 2.

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 1, then it is an axiom. Therefore $P = \emptyset$ and for each substitution θ: $\vdash \varphi[\theta]$.

If $P \Rightarrow \varphi$ is introduced by Rule 2, then there is a $Q \Rightarrow \alpha \in J_k$, $R \Rightarrow (\beta \rightarrow \psi) \in J_k$, α and β are unifiable with a most general unifier θ such that: $P = (Q \cup R)[\theta]$ 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: $P[\zeta] \subseteq \overline{\Sigma}$ and $P[\zeta] \vdash \varphi[\zeta]$.

Theorem 2 *For each $P \subseteq \overline{\Sigma}$: if $P \vdash \varphi$, then for some $i \geq 0$: $Q \Rightarrow \psi \in J_i$ and for some substitution θ: $Q[\theta] \subseteq P$ and $\psi[\theta] = \varphi$.*

Proof Let $P \subseteq \overline{\Sigma}$ and $P \vdash \varphi$. 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 $j \leq n$: 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 to the length of the deduction sequence.

Let $\langle \varphi_0 \rangle$ be a deduction sequence for $P \vdash \varphi$.

If $\varphi_0 \in P$, then $\varphi_0 \in \overline{\Sigma}$ and there exists a $\psi \in \Sigma$ such that for some substitution θ: $\psi[\theta] = \varphi$.

If φ_0 is an axiom, then there exists some $i_0 \geq 0$ such that $J_{i_0} = J_{i_0-1} \cup \{\emptyset \Rightarrow \varphi_0\}$ and $\emptyset \Rightarrow \varphi_0$ is added by Rule 1.

Hence the theorem holds for a deduction sequence 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 $J_{i_{m+1}} = J_{i_{m+1}-1} \cup \{\emptyset \Rightarrow \varphi_{m+1}\}$ and $\emptyset \Rightarrow \varphi_{m+1}$ is added by Rule 1.

If there exists a φ_k and a φ_l with $k, l \leq m+1$ 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: $Q[\theta] \subseteq P$ and $\varphi_k = \alpha[\theta]$, and for some substitution ζ: $R[\zeta] \subseteq P$ and $\varphi_l = (\beta \rightarrow \psi)[\zeta]$. Since $\alpha[\theta] = \beta[\zeta] = \varphi_k$, α and β are unifiable. Let ξ be the most general unifier. Now there exists some i_{m+1} with $i_k, i_l < i_{m+1}$ such that $S \Rightarrow \psi \in J_{i_{m+1}}$, $S = (Q \cup R)[\xi]$ and for some substitution σ: $\varphi_{m+1} = \psi[\sigma]$.

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 3 *For each $i \geq 0$: 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 inconsistent and for each $\psi \in P[\theta]$ there holds: $\psi \not\vdash \varphi[\theta]$.*

Proof The theorem is proven using induction to the index of the set of justifications.

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 3.

If $P \not\Rightarrow \varphi \in J_k$, then by the induction hypothesis for substitution θ: $(P \cup \{\varphi\})[\theta] \subseteq \overline{\Sigma}$, $(P \cup \{\varphi\})[\theta]$ is inconsistent and for each $\psi \in P[\theta]$ there holds: $\psi \not\prec \varphi[\theta]$.

If $P \not\Rightarrow \varphi$ is introduced by Rule 3, then there is a $R \Rightarrow \alpha \in J_k$, $Q \Rightarrow \neg\beta \in J_k$ and α and β are unifiable. If ζ is the most general unifier, then $\varphi \in min((Q \cup R)[\zeta])$ and $P = ((R \cup Q)[\zeta])/\varphi$. By Theorem 1 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 substitutuion θ: $(P \cup \{\varphi\})[\zeta \circ \theta] \subseteq \overline{\Sigma}$, $(P \cup \{\varphi\})[\zeta \circ \theta]$ is inconsistent and for each $\psi \in P[\theta]$: $\psi \prec \varphi[\theta]$.

Theorem 4 *For each $P \subseteq \overline{\Sigma}$ if P is a minimal inconsistent set and $Q = min(P)$, then for some $i \geq 0$ there holds for each $\varphi \in Q$: $R \not\Rightarrow \psi \in J_i$, and for some substitution θ: $P = Q[\theta]$ and $\varphi = \psi[\theta]$.*

Proof Let P be a minimal inconsistent subset of $\overline{\Sigma}$ with $Q = min(P)$. Since P is inconsistent there exists a formula α such that $P \vdash \alpha$ and $P \vdash \neg\alpha$. By Theorem 2 there exists a $j \geq 0$: $S \Rightarrow \beta \in J_j$ and for some substitution ζ $S[\zeta] \subseteq P$ and $\alpha = \beta[\zeta]$. Also by Theorem 2 there exits a $k \geq 0$: $T \Rightarrow \neg\gamma \in J_k$ and for some substitution ξ: $T[\xi] \subseteq P$ and $\alpha = \gamma[\xi]$. Since β and γ are unifiable, there exists a most general unifier σ. Hence, for some substitution θ: $P = (S \cup T)[\sigma \circ \theta]$ and $Q = min(P) = min((S \cup T)[\sigma \circ \theta])$. Therefore, there exists a $l > j, k$ such that for each $\varphi \in Q$, there is a $\psi \in min(S \cup T)$: $((R \cup S)/\psi)[\sigma] \not\Rightarrow \psi[\sigma] \in J_l$, $\varphi = \psi[\theta]$ and $R = P/\varphi = (((S \cup T)/\psi)[\sigma])[\theta]$.

Hence for some $i \geq 0$: $P/\varphi \not\Rightarrow \psi \in J_i$ and for some substitution θ: $P = Q[\theta]$ and $\psi[\theta] = \varphi$.

Observation 1 *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: $\varphi[\theta] = \psi$ and $P[\theta] \subseteq \Delta$. Therefore by theorem 1: $P \vdash \varphi$ and $P[\theta] \subseteq \Delta$. Hence, for each $\Delta \in \mathcal{A}_i$: $\Delta \vdash \varphi$.

Observation 2 *For each $\Delta \in \mathcal{A}_\infty$: $\Delta \subseteq \overline{\Sigma}$.*

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

Observation 3 *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 4 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.

Observation 4 *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 an other subset in which it is not a minimal element, then $\mid \mathcal{A}_\infty \mid = 1$.*

Proof Suppose that $\mid \mathcal{A}_\infty \mid > \infty$. Then there exists 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 4 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, $\psi[\theta] \prec \eta[\theta]$ for every $\eta \in P$. Since $\varphi \notin \Delta$ and $\varphi \in \Delta'$, there exists a $\eta \in P$: $\eta[\theta] \in \Delta$ and $\eta[\theta] \in \Delta'$ Hence there exists an infinite sequence of minimal inconsistent subsets such a minimal element of one subset is a non minimal element an other subset.

Hence Δ_∞ is unique.

Observation 5

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

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

Proof For each $\Delta \in \mathcal{A}_\infty$: $B_\infty \subseteq Th(\Delta_\infty)$ because according to Observation 1: if $\varphi \in B_\infty$, then for each $\Delta \in \mathcal{A}_\infty$: $\Delta \vdash \varphi$.

Suppose there exists a φ such that: $\varphi \notin B_\infty$ 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 for each $\Delta \in \mathcal{A}_\infty$ there exists a $P \Rightarrow \varphi \in J_i$ and $P \subseteq \Delta$. Therefore $P \Rightarrow \varphi \in J_\infty$, and $P \subseteq \Delta$. Hence $\varphi \in B_\infty$. Contradiction.

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

Theorem 5

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

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

Proof From Observation 5 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 devided into the proof of the soundness and the proof of the completeness of the preference logic. Firstly the completeness is proven.

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 Observation 3 $Prem(\mathcal{M}) \not\subset Prem(\mathcal{N})$. Hence, there exists a $\varphi \in (\Delta - Prem(\mathcal{N}))$ — $Prem(\mathcal{M}) = \Delta$—. Now by Definition 9 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: $P[\theta] \subseteq \Delta$ and $\psi = \eta[\theta]$. Now, $P[\theta] \not\subseteq Prem(\mathcal{N})$, otherwise $Prem(\mathcal{N})$ would be inconsistent. Hence there exists a $\mu \in P[\theta]$, $\mu \in (\Delta - Prem(\mathcal{N}))$ and $\mu \not\prec \psi$. Therefore there exists a linear extension of $(\overline{\Sigma}, \prec)$ such that $\varphi \prec \psi \prec \mu$, contradiction.

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

Now the completeness has been proven, soundness is proven.

Suppose there exists 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 and 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 $\varphi \in Prem(\mathcal{M})$ 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: $P[\theta] \subseteq Prem(\mathcal{M})$ 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 4 there exists a $R \not\Rightarrow \psi$ and for some substitution θ: $P/\varphi = R[\theta]$ 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 MI the union of all the sets $min(P)$ for each minimal inconsistent subset P of $Prem(\mathcal{M}) \cup \{\varphi\}$. For each $\eta \in MI$ there holds $\eta \prec \varphi$. Clearly the set $(Prem(\mathcal{M}) \cup \{\varphi\}) - MI$ is consistent. Let $\mathcal{N} \in Mod((Prem(\mathcal{M}) \cup \{\varphi\}) - MI)$. 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 $\overline{\Sigma} - Out_\infty(Prem(\mathcal{M})) \subseteq Prem(\mathcal{M})$.

The following reports have appeared in this series:

89-42	Jacco Wesselius en Frans Ververs	Software Quality Assurance: naar een meetbare kwaliteitsverbetering
89-43	J.M. Gambaudo, S. van Strien et C. Tresser	Vers un ordre de Sarkovskii pour les plongements du disque preservant l'orientation
89-44	C. Roos and J.Ph. Vial	Long steps with the logarithmic penalty barrier function in linear programming
89-45	Freerk A. Lootsma	Parallel non-linear optimization
89-46	Maria Polkowska-Semeniuk and Lech Polkowski	An introduction to Montague grammar
89-47	J.A.H. Alkemade and P.L. Vermeer	Multiresolution and time-frequency representations of signals
89-48	René F. Swarttouw	On a third q-Bessel function
89-49	Roger Cooke and Hans Meeuwissen	Uncertainty analysis of the accident effect area of a BLEVE
89-50	J.J. Kalker	Elastic and viscoelastic analysis of two multiply layered cylinders rolling over each other with coulomb friction
89-51	Roelof Koekoek	Generalizations of a q-analogue of Laguerre polynomials
89-52	L. Stakenborg	Guide to RMS-literature
89-53	Nico Roos	Preference logic: a logic for reasoning with inconsistent knowledge

Copies of these reports may be obtained from the bureau of the Faculty of Technical Mathematics and Informatics, Julianalaan 132, 2628 BL DELFT, phone 015-784568.

893724