Artificial Intelligence 57 (1992) 69–103
Elsevier

A logic for reasoning with inconsistent knowledge *

Nico Roos

Research Institute for Knowledge Systems, Tongersestraat 6, P.O. Box 463, 6200 AL Maastricht, Netherlands

Received June 1991
Revised April 1992

Abstract

Roos, N., A logic for reasoning with inconsistent knowledge, Artificial Intelligence 57 (1992) 69–103.

In many situations humans have to reason with inconsistent knowledge. These inconsistencies may occur due to not fully reliable sources of information. In order 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 the set of premisses as a set of assumptions, however, it is possible to deduce useful conclusions from an inconsistent set of premisses. In this paper a logic for reasoning with inconsistent knowledge is described. This logic is a generalization of the work of Rescher [12]. In the logic a reliability 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 logic. For the logic a semantics based on the ideas of Shoham [14,15] is defined. It turns out that the semantics for the logic is a preferential semantics according to the definition of Kraus, Lehmann and Magidor [9]. Therefore the logic is a logic of system *P* and possesses all the properties of an ideal nonmonotonic logic.

1. Introduction

In many situations humans have to reason with inconsistent knowledge. These inconsistencies may occur due to sources of information which are not fully reliable. For example, in daylight information about the position of an object coming from your eyes is more reliable than the information

Correspondence to: Research Institute for Knowledge Systems, Tongersestraat 6, P.O. Box 463, 6200 AL Maastricht, Netherlands.
*The research reported on was carried out at Delft University of Technology (TU-Delft) and the National Aerospace Laboratory (NLR) in Amsterdam.

about the position of the object coming from your ears. But even reliable sources, such as domain experts, do not always agree.

To be able to reason with inconsistent knowledge it is not possible to view a set of premisses as absolute truths, as 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. As long as we do not have it proven otherwise, the premisses are assumed to be true statements about the world. When, however, a contradiction is derived, we can no longer make this assumption. To restore consistency, one of the premisses has to be removed. To be able to select a premiss to be removed, a reliability relation on the premisses will be used. This reliability relation denotes the relative reliability of the premisses.

In the following sections I will first describe the propositional case. After describing the propositional case, I will describe how to extend the logic to the first-order case.

2. Basic concepts

The language L, which will be used to express the propositions of the logic, consists of the propositions that can be generated using a set of atomic propositions and the logical operators $\neg$ and $\rightarrow$. When in this paper the operators $\wedge$ and $\vee$ are used, they should be interpreted as shortcuts: i.e. $\alpha \wedge \beta$ for $\neg(\alpha \rightarrow \neg\beta)$ and $\alpha \vee \beta$ for $\neg\alpha \rightarrow \beta$.

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 reliability relation. This reliability relation denotes the relative reliability of the premisses. It denotes that one premiss is more reliable than some other premiss. Clearly the relation must be irreflexive, asymmetric, and transitive. I do not demand this relation to be total, for a total reliability relation implies complete knowledge about the relative reliability of the premisses. This does not always have to be the case.

A set of premisses Σ is a subset of the language L. On the set of premisses Σ a partial reliability relation $\prec$ may be defined. Together they form a reliability theory.

Definition 2.1. A *reliability theory* is a tuple $\langle \Sigma, \prec \rangle$ where $\Sigma \subseteq L$ is a finite set of premisses and $\prec \subseteq (\Sigma \times \Sigma)$ is an irreflexive, asymmetric, and transitive partial reliability relation.

Using the reliability relation, we have to remove a least preferred premiss of the inconsistent set, thereby blocking the derivation of the contradiction.

Example 2.2. Let Σ denote a set of premisses,

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

and $\prec$ a reliability relation on Σ:

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

From Σ, ψ can be derived using 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 block the derivation of 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 an *in-justification*, describes the premisses from which a proposition is derived.

Secondly, a premiss that has been removed may have to be placed back because the contradiction causing its removal is also blocked by the removal of another premiss. This may occur because of some other contradiction being derived.

Example 2.3. Let Σ be a set of premisses,

$$\Sigma = \{\alpha, \neg\alpha \wedge \neg\beta, \beta\},$$

and let $\prec$ be a reliability relation on Σ given by

$$\alpha \prec (\neg\alpha \wedge \neg\beta) \prec \beta.$$

From α and $\neg\alpha \wedge \neg\beta$ we can derive a contradiction causing the removal of α. From $\neg\alpha \wedge \neg\beta$ and β we can also derive a contradiction causing the removal of $\neg\alpha \wedge \neg\beta$. When $\neg\alpha \wedge \neg\beta$ is removed, it is no longer necessary that α is also removed from the set of premisses to avoid the derivation of a contradiction.

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 reliable premiss in the set of premisses on which a contradiction is based. This can occur when no reliability relation between premisses is specified. In such a situation we have to consider the results of the removal of every alternative separately.

Choosing a premiss to be removed implies that we assume the alternative to be more reliable. Since the reliability relation is transitive, making such a choice influences the reliability relation defined on the premisses.

Example 2.4. Let $\Sigma = \{a, b, \neg a, \neg b\}$ be a set of premisses and let $\prec = \{(a, \neg b), (b, \neg a)\}$ be a reliability relation on Σ. Since a and $\neg a$ are in conflict and since there is no reliability relation defined between them, we have to choose a culprit. If we choose to remove a, $\neg a$ is assumed to be more reliable. Therefore, $\neg b$ is more reliable than b. Hence, since b and $\neg b$ are also in conflict, b must be removed.

As is illustrated in the example above, the premisses removed depend on the extension of the reliability relation. Therefore, in the logic described here, every (strict) linear extension of the reliability relation will be considered.

Different linear extensions of the reliability relation can result in different subsets of the premisses that are assumed to be true statements about the world (*that can be believed*). The set of theorems is defined as the intersection of all extensions of the logic.

As mentioned above two kinds of justifications, *in-justifications* and *out-justifications*, will be used. The in-justifications are used to denote that a proposition 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 2.5. Let Σ be a set of premisses. Then an *in-justification* is a formula,

$$P \Rightarrow \varphi,$$

where P is a subset of the set of premisses Σ and $\varphi \in L$ is a proposition.

An out-justification is a formula:

$$P \not\Rightarrow \varphi,$$

where P is a subset of the set of premisses Σ and φ is a premiss in Σ, but not in P.

3. Characterizing the set of theorems

In this section a characterization, based on the ideas of Rescher [12], is given for the set of theorems of a reliability theory. As is mentioned in the previous section, linear extensions of the reliability relation have to be considered. For each linear extension a set of premisses that can still be believed can be determined. This set can be determined by enumerating the premisses with respect to the linear extension of the reliability relation, starting with the most reliable premiss. Starting with an empty set D, if a premiss may consistently be added to the set D, it *should* be added. Otherwise it must be ignored. Because the most reliable premisses are added first, we get a *most reliable consistent set of premisses.*

Definition 3.1. Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Let $\sigma_1, \sigma_2, \ldots, \sigma_m$ be some enumeration of Σ such that for every $k < j$, $\sigma_j \prec \sigma_k$. Then D is a *most reliable consistent set of premisses* if and only if:

$$D = D_m, \qquad D_0 = \emptyset$$

and for $0 < i < m$

$$D_{i+1} = \begin{cases} D_i \cup \{\sigma_i\}, & \text{if } D_i \cup \{\sigma_i\} \text{ is consistent,} \\ D_i, & \text{otherwise.} \end{cases}$$

Let $\mathcal{A}$ be the set of all the most reliable consistent sets of premisses that can be determined.

Definition 3.2. Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Then the set $\mathcal{A}$ of *all the most reliable consistent sets of premisses* is defined by:

$$\mathcal{A} = \{D \mid D \text{ is a most reliable consistent set of premisses given some enumeration of } \Sigma\}.$$

The set of theorems of a reliability theory is defined as the set of those propositions that are logically entailed by every most reliable consistent set of premisses in $\mathcal{A}$.

Definition 3.3. Let $\langle \Sigma, \prec \rangle$ be a reliability theory and let $\mathcal{A}$ be the corresponding set of all the most reliable consistent sets of premisses. Then the *set of theorems* of $\langle \Sigma, \prec \rangle$ is defined as:

$$Th(\langle \Sigma, \prec \rangle) = \bigcap_{D \in \mathcal{A}} Th(D).$$

4. The deduction process

In this section a deduction process for a reliability theory is described. Given a strict linear extension $\prec'$ of the reliability relation $\prec$, the deduction process determines the set of premisses that can be believed.

Remark 4.1. Instead of starting a deduction process for every strict linear extension of $\prec$, we can also create different extensions of $\prec$ when a contradiction not based on a single least reliable premiss is derived. This approach results in one deduction tree instead of a deduction sequence for every linear extension of $\prec$.

Instead of deriving new propositions, only new justifications are derived. These justifications are generated by the inference rules. The reason why justifications instead of propositions are derived, is that the propositions that can be believed (the belief set) depend on the set of premisses that can still be believed. Since this set of premisses may change because of new information derived, the belief set can change in a nonmonotonic way. The justifications, however, do not depend on the information derived. Furthermore, they contain all the information needed to determine the premisses that can still be believed and the corresponding belief set.

Starting with an initial set of justifications J_0, the deduction process generates a sequence of sets of justifications:

$$J_0, J_1, J_2, \ldots.$$

With each set of justifications J_i there corresponds a belief set B_i. So we get a sequence of belief sets:

$$B_0, B_1, B_2, \ldots.$$

Although for the sets of justifications there holds:

$$J_i \subseteq J_{i+1},$$

such a property does not hold for the belief sets. Because a belief set B_i is determined by a *reason maintenance system* using the justifications J_i, the belief set can change in a nonmonotonic way. Goodwin has called this the process nonmonotonicity of the deduction process [8]. According to Goodwin this process nonmonotonicity is just an other aspect of nonmonotonic logics.

In the limit, when the set of all justifications J_∞ has been derived, the corresponding belief set B_∞ will be equal to an extension of the reliability theory. Goodwin called this process of deriving the set of theorems the *logical process theory* of a logic [8]. The logical process theory focuses on the deduction process of 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 follows from the premisses.

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 proposition is believed if the corresponding premiss is believed.

Definition 4.2. 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, $i > 0$, is generated from the set J_{i-1} by adding a new justification. How these justifications are determined depends on the deduction system used. In the following description of the deduction process I will use an axiomatic deduction system for the language L, only containing the logical operators $\rightarrow$ and $\neg$.

Axioms. The *logical axioms* are the tautologies of a propositional logic.

Because an axiomatic approach is used, justifications for the axioms have to be introduced. Since an axiom is always valid, it must have an in-justification with an antecedent equal to the empty set. An axiom is introduced by the following axiom rule.

Rule 4.3. *An axiom φ gets an in-justification $\emptyset \Rightarrow \varphi$.*

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

Rule 4.4. *Let φ and $\varphi \rightarrow \psi$ be two propositions with justifications*

$$P \Rightarrow \varphi \quad \textit{and} \quad Q \Rightarrow (\varphi \rightarrow \psi),$$

respectively. Then the proposition ψ gets an in-justification $(P \cup Q) \Rightarrow \psi$.

While modus ponens introduces a new in-justification, the contradiction rule introduces a new out-justification to eliminate a contradiction.

Rule 4.5. *Let φ and $\neg\varphi$ be propositions with justifications $P \Rightarrow \varphi$ and $Q \Rightarrow \neg\varphi$, and let $\eta = min(P \cup Q)$ where the function min selects the minimal element given the extended preference relation $\prec'$. Then the premiss η gets an out-justification $((P \cup Q)/\eta) \not\Rightarrow \eta$.*

In order to guarantee that the current set of believed premisses will approximate a most reliable consistent set of premisses, we have to guarantee that the process creating new justifications is fair, i.e. the process does not forever defer the addition of some possible justification to the set of justifications.

Assumption 4.6. *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 proofs are given in Appendix A.) The first theorem guarantees the soundness of the in-justifications, i.e. the antecedent of an in-justification logically entails the consequent of the in-justification. The second theorem guarantees the completeness of the in-justifications, i.e. if a proposition is logically entailed by a subset of the premisses, then there exists a corresponding in-justification. Finally, the third and fourth theorems guarantee respectively the soundness and the completeness of the out-justifications.

Theorem 4.7 (Soundness). *For each $i \geqslant 0$:*

if $P \Rightarrow \varphi \in J_i$, then

$P \subseteq \Sigma$ *and* $P \models \varphi$.

Theorem 4.8 (Completeness). *For each $P \subseteq \Sigma$:*

if $P \models \varphi$, then

there exists a $Q \subseteq P$ such that for some $i \geqslant 0$: $Q \Rightarrow \varphi \in J_i$.

Theorem 4.9 (Soundness). *For each $i \geqslant 0$:*

if $P \not\Rightarrow \varphi \in J_i$, then

$(P \cup \{\varphi\}) \subseteq \Sigma$ *and* $(P \cup \{\varphi\})$ *is not satisfiable.*

Theorem 4.10 (Completeness). *For each $P \subseteq \Sigma$:*

if P is a minimal unsatisfiable set of premisses and $\varphi = min(P)$, where the function min selects the minimal element given the extended preference relation $\prec'$, then *for some* $i \geqslant 0$: $P/\varphi \not\Rightarrow \varphi \in J_i$.

Given a set of justifications, there exists a set of the premisses that can still be believed. Such a set 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 $\Delta \subseteq \Sigma$ is the set of the premisses that are assumed to be true by the reasoning agent. Then for each premiss φ such that for some out-justification $P \not\Rightarrow \varphi \in J_i$ there holds that $P \subseteq \Delta$, one may not believe φ. The set of premisses that may not be believed given a set of justifications J_i, is denoted by $Out_i(\Delta)$.

Definition 4.11.

$$Out_i(S) = \{\varphi \mid P \not\Rightarrow \varphi \in J_i, \text{ and } P \subseteq S\}.$$

The set of premisses Δ must, of course, be equal to the set of premisses obtained after removing all the premisses we may not believe, i.e. $\Delta = \Sigma - Out_i(\Delta)$. The set of premisses that satisfy this requirement is defined by the following fixed point definition.

Definition 4.12. Let Σ be a set of premisses and let J_i be a set of justifications. Then the *set of premisses* Δ_i that can be assumed to be true is defined as:

$$\Delta_i = \Sigma - Out_i(\Delta_i).$$

Property 4.13. *For every* i, *the set* Δ_i *exists and is unique.*

After determining the set of premisses that can be believed, the set of derived propositions that can be believed can be derived from the in-justifications. This set is defined as:

Definition 4.14. Let J_i be a set of justifications and Δ_i be the corresponding set of premisses that may assumed to be true. The set of propositions B_i that can be believed, *the belief set*, is defined as:

$$B_i = \{\psi \mid P \Rightarrow \psi \in J_i \text{ and } P \subseteq \Delta\}.$$

Property 4.15. *For each* $\varphi \in B_i$:

$$\Delta_i \vdash \varphi.$$

Let J_∞ be the set of all justifications that can be derived.

Definition 4.16.

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

The corresponding set of premisses that can be believed and the belief set will be denoted by Δ_∞ and B_∞, respectively.

Property 4.17. *Δ_∞ is maximal consistent.*

Property 4.18.

$$B_\infty = Th(\Delta_\infty),$$

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

The following theorem implies that the characterization of the theorems of the logic, given in the previous section, is equivalent to the intersection of the belief sets that can be derived.

Theorem 4.19. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Then there holds:*

$$\mathcal{A} = \{\Delta_\infty \mid \textit{for some linear extension of} \prec, \Delta_\infty \textit{ can be derived}\}.$$

Corollary 4.20.

$$Th(\langle \Sigma, \prec \rangle) = \bigcap \{B_\infty \mid \textit{for some linear extension of} \prec\}.$$

5. Determination of the belief set

In this section I will describe the algorithms that determine the set of premisses that can be believed and the belief set, given a set of out-justifications. The first algorithm determines the set Δ_i given the sets of justifications J_i. To understand how the algorithm works, recall that the consequent of an out-justification is less reliable than the premisses in the antecedent. Therefore, if the consequent of an out-justification $P \not\twoheadrightarrow \varphi$ is the most reliable premiss to be removed, because we still belief the premisses in the antecedent P, removing φ will never have to be undone. After having removed φ we can turn to the next most reliable consequent of an out-justification.

The time complexity of the algorithm below depends on the **for** and the **repeat** loop. The former loop can be executed in $O(n)$ steps where n in the number of out-justifications. The latter loop can be executed in $O(m)$ steps where m in the number of premisses in Σ. Therefore, the whole algorithm can be executed in $O(n \cdot m)$ steps.

```
begin
  Δ_i := Σ;
  repeat
    φ ∈ max(Σ);
    Σ := Σ/φ;
    for each P ⇏ φ ∈ J_i do
      if  P ⊆ Δ_i
      then Δ_i := Δ_i/φ;
  until Σ = ∅;
  return Δ_i;
end.
```

Using the in-justfications, the belief set B_i can be determined in a straightforward way. Clearly, B_i can be determined in $O(n)$ steps where n is the number of in-justifications.

```
begin
  B_i = ∅;
  repeat
    P ⇒ φ ∈ J_i;
    J_i := J_i − {P ⇒ φ};
    if  P ⊆ Δ_i
    then B_i := B_i ∪ {φ};
  until J_i = ∅;
  return B_i;
end.
```

6. The semantics for the logic

The semantics for the logic is based on the ideas of Shoham [14,15]. In [14,15] he argues that the difference between monotonic logic and nonmonotonic logic is a difference in the definition of the entailment relation. In a monotonic logic a proposition is entailed by the premisses if it is true in every model for the premisses. In a nonmonotonic logic, however, a proposition 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 logic differs slightly from Shoham's approach. Since the set of premisses may be inconsistent, the set of models for these premisses may be empty. Therefore, instead of defining a preference relation on the models of the premisses, a partial preference relation on the set of semantical interpretations for the language is defined. Given such a preference relation on the interpretations, the models for a reliability theory are the most preferred semantical interpretations. The preference relation used here 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 reliability than a premiss with a high reliability.

Therefore, an interpretation satisfying more premisses with a higher reliability ($\prec$) than some other interpretation, is preferred ($\sqsubset$).

Example 6.1. Let $\mathcal{M}$ and $\mathcal{N}$ be two interpretations. Furthermore, let $\mathcal{M}$ satisfy α and β, and let $\mathcal{N}$ satisfy β and γ. Finally, let α be more reliable than γ, $\gamma \prec \alpha$. Clearly, we cannot compare $\mathcal{M}$ and $\mathcal{N}$ using the premiss β. $\mathcal{M}$ and $\mathcal{N}$ can, however, be compared using the premisses α and γ. Since α is more reliable than γ, $\mathcal{N}$ does not satisfy α, and $\mathcal{M}$ does not satisfy γ, we find that $\mathcal{M}$ must be preferred to $\mathcal{N}$.

Definition 6.2. An *interpretation* $\mathcal{M}$ is a set containing the atomic propositions that are true in this interpretation.

Definition 6.3. Let $\mathcal{M}$ be a semantical interpretation and let Σ be a set of premisses. Then the premisses $Prem(\mathcal{M}) \subseteq \Sigma$ that are satisfied by $\mathcal{M}$ are defined as:

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

Definition 6.4. Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Furthermore, let $\sqsubset$ be a preference relation on the interpretations. For all interpretations $\mathcal{M}$ and $\mathcal{N}$ there holds:

> $\mathcal{M} \sqsubset \mathcal{N}$ if and only if $Prem(\mathcal{M}) \neq Prem(\mathcal{N})$ and for every $\varphi \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$, there is a $\psi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$ such that $\varphi \prec \psi$.

Given the preference relation on the interpretations, the set of models for the premisses can be defined.

Definition 6.5. Let $\langle \Sigma, \prec \rangle$ be a reliability theory and let $Mod_{\sqsubset}(\langle \Sigma, \prec \rangle)$ denote the models for the reliability theory. Then

> $\mathcal{M} \in Mod_{\sqsubset}(\langle \Sigma, \prec \rangle)$ if and only if there exists no interpretation $\mathcal{N}$ such that $\mathcal{M} \sqsubset \mathcal{N}$.

Now the following important theorem, guarantees the soundness and the completeness of the logic.

Theorem 6.6. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Furthermore, let $\mathcal{A}$ be the corresponding set of all most reliable consistent sets of premisses. Then:*

$$Mod_{\sqsubset}(\langle \Sigma, \prec \rangle) = \bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty),$$

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

7. 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 nonmonotonic logics described by Kraus, Lehmann, and Magidor [9]. Secondly, I will compare the behaviour of the logic when new information is added with Gärdenfors' theory for belief revision [6].

7.1. *Preferential models and cumulative logics*

In [9] Kraus et al. describe a general framework for the study of nonmonotonic 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 logic discussed here is the preferential consequence relation of system $\boldsymbol{P}$. I will show that the preference relation on the semantic interpretations, described in the previous section, corresponds to a preferential model as described by Kraus et al.

Lemma 7.1. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Furthermore, let*

$$\widehat{\alpha} = \{\mathcal{M} \mid \mathcal{M} \models \alpha\},$$

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

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

Then $\mathcal{M} \in Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$ if and only if $\mathcal{M} \in \widehat{\alpha}$ and for no $\mathcal{N} \in \widehat{\alpha}$: $\mathcal{M} \sqsubset \mathcal{N}$.

Theorem 7.2. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. $\langle S, l, < \rangle$ is a preferential model for $\langle \Sigma, \prec \rangle$ if and only if S is the set of all possible interpretations for the language L, $l : S \to S$ is the identity function, and for each $\mathcal{M}, \mathcal{N} \in S$:*

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

Now I will relate the consequence relation of system $\boldsymbol{P}$ to the logic. To motivate the relation I will describe below, recall that $\alpha \mid\!\sim \beta$ 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. Furthermore, α must be more reliable than any other premiss, otherwise we cannot guarantee that α is an element of the set of theorems $Th(\langle \Sigma, \prec \rangle)$. It is possible that α is an element of the original set of premisses. In that case we should use the most reliable knowledge source for a premiss, i.e. the assumption that α is true beyond any doubt. If α is indeed an element of B_∞, we must prove that β will also be an element of $Th(\langle \Sigma, \prec \rangle)$.

Theorem 7.3. *Let $W = \langle S, l, < \rangle$ be a preferential model for $\langle \Sigma, \prec \rangle$. Then the following equivalence holds:*

$$\begin{aligned}
&\alpha \mid\!\sim_W \beta \quad \text{if and only if} \\
&\Sigma' = \Sigma \cup \{\alpha\}, \\
&\prec' = (\prec \upharpoonright (\Sigma/\alpha \times \Sigma/\alpha)) \cup \{\langle \varphi, \alpha \rangle \mid \varphi \in \Sigma/\alpha\}, \\
&\beta \in Th(\langle \Sigma', \prec' \rangle).
\end{aligned}$$

Corollary 7.4. *Let $W = \langle S, l, < \rangle$ be a preferential model for $\langle \Sigma, \prec \rangle$. Then:*

$$Th(\langle \Sigma, \prec \rangle) = \{\alpha \mid \;\mid\!\sim_W \alpha\}.$$

7.2. Belief revision

In [6], 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 proposition. Revision is a more complex form of adding a new proposition. 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 proposition. 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 logic. To be able to do this, the set of theorems of a reliability theory is identified as a belief set as defined by Gärdenfors. Here expansion, revision, and contraction of the belief set K, with respect to a proposition α, will be denoted by $K^{+}[\alpha]$, $K^{*}[\alpha]$, and $K^{-}[\alpha]$, respectively.

7.2.1. Expansion

To expand a belief set K with respect to a proposition α, α should be added to the set of premisses that generate the belief set. Since the logic does not allow an inconsistent belief set, α can be added if the belief set does not already contain $\neg\alpha$. 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 reliability theory.

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

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

Clearly, adding α to Σ does not result in believing α. Hence, the second postulate for 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 logic. The reason for this is that the reasons for believing a proposition 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.

7.2.2. Revision

For revision of a belief set K with respect to a proposition 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 logic.

Theorem 7.5. *Let belief set $K = Th(\langle\Sigma, \prec\rangle)$ be the set of theorems of the reliability theory $\langle\Sigma, \prec\rangle$. Suppose that $K^{*}[\alpha]$ is the belief set of the premisses $\Sigma \cup \{\alpha\}$ with reliability relation:*

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

i.e. $K^{*}[\alpha] = \{\beta \mid \alpha \mathrel{|\!\sim}_W \beta\}$ where W is a preferential model for $\langle\Sigma, \prec\rangle$. Then the following postulates are satisfied:

(1) *$K^{*}[\alpha]$ is a belief set.*
(2) $\alpha \in K^{*}[\alpha]$.
(6) *If $\vdash \alpha \leftrightarrow \beta$, then $K^{*}[\alpha] = K^{*}[\beta]$.*

Contraction

It is not possible to realize contraction of a belief set in the logic in a straightforward way. To be able to contract a proposition α from a belief set K, we have to determine the premisses on which belief in this proposition is based. This information can be found in the applicable in-justification that supports the proposition α. When we have determined these premisses, we have to remove some of them, i.e. for each linear extension of the reliability relation, we must add the following out-justifications to J_∞

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

Unfortunately, this solution, which requires a modification of the logic, can only be applied after J_∞ has been determined. Furthermore, only the most trivial postulates (1), (3), (4) and (6) will be satisfied.

8. Extension to first order logic

The logic described in the previous sections can be extended to a first-order logic. To realize this we have to replace the propositional language L by a first-order language, which only contains the logical operators $\neg$ and $\rightarrow$, and the quantifier $\forall$. Furthermore we have to replace the logical axioms for a propositional logic by the logical axioms for a first-order logic with modus ponens as the only inference rule. We can for example use the following axiom scheme, which originate from [5].

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

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

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

Finally, we have to replace the definition of the semantical interpretations by a definition for the semantical interpretations of a first-order logic.

When these modifications are made we have a first-order logic for reasoning with inconsistent knowledge. For this first-order logic all the results that can be found in the preceding section also hold.

9. Related work

In this section I will discuss some related approaches. Firstly, the relation with of Rescher's work will be discussed. Rescher's work is closely related to the logic described here. Secondly, the relation with Poole's framework for default reasoning, which is a special case of Rescher's work, will be discussed. Thirdly, the difference between paraconsistent logics and the logic described here will be discussed. Finally, the relation with truth maintenance systems, and especially de Kleer's ATMS, will be discussed.

9.1. Hypothetical reasoning

In his book *Hypothetical Reasoning* [12], Rescher describes how to reason with 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 the case of counterfactual reasoning, we make an assumption that we know to be false. For example, let us suppose that Plato lived in the Middle Ages. To be able to make such a counterfactual assumption, we, temporally, have to give up some of our beliefs to maintain consistency. It is, however, not always clear which of our beliefs we have to give up. The following example gives an illustration.

Example 9.1.
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 wish to withdraw (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 $M_0, \ldots, M_n$ 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.
(3) S_n is a PMMC subset.

Given these PMMC subsets, Rescher defines two entailment relations.

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

It is not difficult to see that Rescher's modal categories can be represented by a partial reliability relation on the premisses. For all modal categories M_i and M_j, $i < j$, there must hold that each premiss in M_i is more reliable than any premiss in M_j. Given this ordering, from Definition 3.1 it follows immediately that the PMMC subsets are equal to the most reliable consistent sets of premisses.

9.2. A framework for default reasoning

The central idea behind Poole's approach is that default reasoning should be viewed as *scientific theory formation* [10]. 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* has 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 independently of the other instances in an explanation. An explanation for a proposition φ is a maximal (with respect to the inclusion relation) *scenario* that implies φ. Here a scenario is a consistent set containing all the facts and some ground instances of the hypotheses.

This framework can be viewed as a special case of Rescher's work. Poole's framework consists of only two modal categories, the facts M_0 and the hypotheses M_1. Since Rescher's work is a special case of the logic described in this paper, so is Poole's framework. Poole, however, extends his framework with constraints. These constraints are added to be able to eliminate some scenarios as possible explanations for a formula φ. A scenario is eliminated when it is not consistent with the constraints.

The constraints can be interpreted as describing that some scenarios are preferred to others. Since in the logic described in this paper a reliability relation on the premisses generates a preference relation on consistent subsets of the premisses, an obvious question is whether the preference relation described by the constraints can be modeled with an appropriate reliability relation. Unfortunately, the answer is "no". This is illustrated by the following example.

Example 9.2.
Facts: φ and ψ.
Defaults: $\varphi \to \alpha$, $\varphi \to \neg\beta$, $\psi \to \neg\alpha$, $\psi \to \beta$.
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.

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

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

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

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

Only the first two scenarios are consistent with constraints. If this default theory has to be modeled in the logic, a reliability relation has to be specified in such a way that $\{S_1, S_2\} = \mathcal{A}$. To determine the required reliability relation on the hypotheses, combinations of two scenarios are considered. To ensure that $S_1 \in \mathcal{A}$ and $S_3 \notin \mathcal{A}$, $\varphi \to \neg\beta$ has to be more reliable than $\psi \to \beta$. To ensure that $S_2 \in \mathcal{A}$ and $S_4 \notin \mathcal{A}$, $\psi \to \beta$ has to be more reliable than $\varphi \to \neg\beta$. Hence, the reliability 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 modeled, using the logic described in this paper.

Although Poole's framework without constraints can be expressed in the logic described in this paper, 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. The purpose of the logic described in this paper, however, is to derive useful conclusions from an inconsistent set of premisses.

9.3. Paraconsistent logics

Paraconsistent logics are a class of logics developed for reasoning with inconsistent knowledge [1]. Unlike classical logics, in paraconsistent logics there need not hold $\neg(\varphi \wedge \neg\varphi)$ for some proposition φ. Hence, an inconsistent set of premisses is not equivalent to the trivial theory; it does not imply the set of all propositions.

Unlike the logic described in this paper, a paraconsistent logic does not resolve an inconsistency. Instead it simply avoids that everything follows from an inconsistent theory. To illustrate this more clearly, consider the following reliability theory, without a reliability relation:

$$\Sigma = \{\alpha \wedge \beta, \neg\beta \wedge \gamma\}.$$

In the logic described in this paper, all maximal consistent subsets will be generated:

$$\{\alpha \wedge \beta\} \quad \text{and} \quad \{\neg\beta \wedge \gamma\}.$$

In a paraconsistent logic the proposition β will be contradictory but the propositions α and γ will consistently be entailed by the premisses.

The difference between the two approaches can be interpreted as the difference between a credulous and a skeptical view of knowledge sources. With a credulous view of a knowledge source, we try to derive as much as is consistently possible. According to Arruda [1], scientific theories for different domains, which conflict with each other on some overlapping aspect, are treated in this way. With a skeptical view of a knowledge source, we only believe one of the knowledge sources that support the conflicting information. So if part of someone's statement turns out to be wrong, we will not believe the rest of his/her statement. Although a credulous view of knowledge sources seems to be acceptable for scientific theories for different domains, a skeptical view seems to be better for knowledge based systems, which have to act on the information available.

9.4. Truth maintenance systems

In the logic justifications are introduced. Unlike the justifications used in Doyle's JTMS [4] or de Kleer's ATMS [3], the justifications in the logic are part of the deduction process. They follow directly from the requirement for the deduction process (Section 2). Therefore, the justifications are different from the ones introduced by Doyle and de Kleer. In an (A)TMS the justifications describe dependencies between propositions, while in the logic the in-justifications describe dependencies between propositions and premisses and the out-justifications describe dependencies among premisses. The in-justifications of the logic, however, can be compared with the labels in the ATMS [3]. Like a label, an in-justification describes from which premisses a proposition is derived. The out-justifications have more or less the same function as the set **nogood** in the ATMS. As with an element from the set **nogood**, the consequent and the antecedents of an out-justification may not be assumed to be true simultaneously. Unlike an element of the set **nogood**, an out-justification describes which element has to be removed from the set of premisses (assumptions).

Because in-justifications and labels are closely related, it is possible to describe an ATMS using a propositional 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 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:

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

Finally, let every justification be more reliable than any assumption. Then the set $\mathcal{A}$ is equal to the set of maximal (under the inclusion relation) environments of an ATMS. Furthermore, for any linear extension of the reliability relation the label for a node $n \in N$ is equal to the set:

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

The set of nogoods is equal to the set:

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

10. Applications

In the previous sections a logic for reasoning with inconsistent knowledge was described. In this section two applications will be discussed.

10.1. Unreliable knowledge sources

In situations where we must deal with sensor data the logic described in the previous sections can be applied. To be able to reason with sensor data, the data has to be translated into statements about the world. Because of measurement errors and misinterpretation of the data, these statements can be incorrect. This may result in inconsistencies. These inconsistencies may be resolved by considering the reliability of the knowledge sources used. To illustrate this consider the following example.

Example 10.1. Suppose that we want to determine the type of an airplane by using the characteristic of its radar reflection. The radar reflection of an airplane depends on the size and the shape of plane. Suppose that we have some pattern recognition system that outputs a proposition stating the type of plane, or a disjunction of possible types in case of uncertainty. Furthermore, suppose that we have an additional system that determines the speed and the course of the plane. The output of this system will also be stated as a proposition. Given the output of the two systems, we can verify whether they are compatible. If a plane is recognized as a Dakota and its speed is Mach 1.5, then, knowing that a Dakota cannot go through the sound barrier, we can derive a conflict. Since the speed measuring system is more reliable than the type identifying system, we must remove the proposition stating that the plane is a Dakota.

In this example, the reliability relation can be interpreted as denoting that if two premisses are involved in a conflict the least reliable premiss has the highest probability of being wrong. Since the relative probability is conditional on inconsistencies, information from one reliable source cannot be overruled by information from many unreliable knowledge sources. For example, the position of an object determined by seeing it is normally more reliable than the position determined by hearing it, independent of the number persons that heard it at some position. Notice that fault probabilities have no meaning because faults are context dependent. The positions where you hear an object can be incorrect because of reflections and the limited speed of sound. Usually, these factors cannot be predicted in advance.

10.2. Planning

Another possible application for the logic can be found in the area of planning. In [7] Ginsberg and Smith describe a possible worlds approach for reasoning about actions. What they propose is an alternative way of determining the consequences of an action. Instead of using frame axioms, default rules, or add and delete lists. They propose to determine the nearest *world* that is consistent with the consequences of an action. The advantage of this approach is that we do not have to know all possible consequences of an action in advance. For example, in general, we cannot know whether putting a plant on a table will obscure a picture on the wall. Hence, if we know that a picture is not obscured before an action, we may assume that it is still not obscured after the action when this fact is consistent with the consequences of the action.

Example 10.2. Figure 1 can be described a set of premisses. This set of premisses is divided in to three subsets, viz. the domain constraints, the

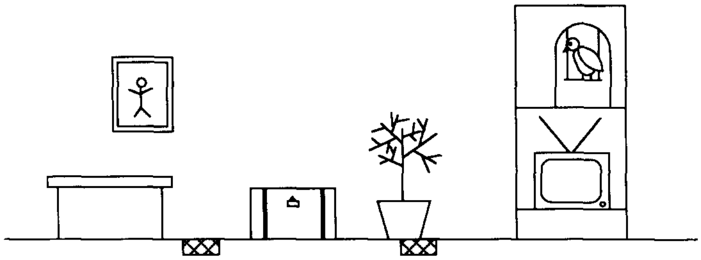

Fig. 1. Living-room.

structural facts, and the remaining facts. The domain constraints are:

(1) $on(x,y) \wedge y \neq z \rightarrow \neg on(x,z)$,

(2) $on(z,y) \wedge z \neq x \wedge y \neq floor \rightarrow \neg on(z,y)$,

(3) $rounded(x) \rightarrow \neg on(x,y)$,

(4) $duct(d) \wedge \exists x.on(x,d) \rightarrow blocked(d)$,

(5) $\exists x.on(x, table) \leftrightarrow obscured(picture)$,

(6) $blocked(duct1) \wedge blocked(duct2) \leftrightarrow stuffy(room)$.

The structural facts are:

(7) $rounded(bird)$,

(8) $rounded(plant)$,

(9) $duct(duct1)$,

(10) $duct(duct2)$,

(11) $in(bottom\text{-}shelf, bookcase)$,

(12) $in(top\text{-}shelf, bookcase)$.

The situational facts are:

(13) $on(bird, top\text{-}shelf)$,

(14) $on(tv, bottom\text{-}shelf)$,

(15) $on(chest, floor)$,

(16) $on(plant, duct2)$,

(17) $on(bookcase, floor)$,

(18) $blocked(duct2)$,

(19) $\neg obscured(picture)$,

(20) $\neg stuffy(room)$.

Clearly, the situational facts are less reliable than the structural facts and the domain constraints. Furthermore, facts added by recent actions are on average more reliable than facts added by less recent actions.

Now suppose that we move the *plant* from *duct*2 to the *table*. This can be described by adding the situational fact $on(plant, table)$. From the new set of premisses we can derive two inconsistencies:

$$\{\exists x.on(x, table) \leftrightarrow obscured(picture),$$
$$\neg obscured(picture), on(plant, table)\}$$

and

$$\{[on(z,y) \wedge z \neq x \wedge y \neq floor] \rightarrow \neg on(z,y),$$
$$on(plant, duct2), on(plant, table)\}.$$

The least reliable premisses in these sets of premisses are respectively the facts $\neg obscured(picture)$ and $on(plant, duct2)$. Hence, they have to be removed from the set of premisses.

11. Conclusions

In this paper a logic for reasoning with inconsistent knowledge has been described. One of the original motivations for developing this logic was based on the view that default reasoning is a special case of reasoning with inconsistent knowledge. To describe defaults in this logic, such as Poole's framework for default reasoning, formulas containing free variables can be used. These formulas denote a set of ground instances. If we do not generate these ground instances, but—by using unification of terms containing free variables—we reason with formulas containing free variables, we can derive conclusions representing sets of instances. This would seem to be a very useful property.

Since, in the logic described here a default rule can only be described by using material implication, a default rule has a contraposition. It is possible, however, that the contraposition may not hold for default rules. For example, the contraposition of the default rule, "someone who owns a driving licence, can drive a car", is not valid. A better candidate for default reasoning would be Reiter's default logic [11] or Brewka's approach [2].

Although it is likely that the logic is not suited for default reasoning, it is suited for reasoning with knowledge coming from different and not fully reliable knowledge sources. For this use of the logic, it seems plausible that the logic satisfies the properties of system ***P***. As was shown in the examples described in Section 10, the reliability relation can be given plausible probabilistic and ontological interpretations. Furthermore, the current belief set with respect to the inferences made can be determined efficiently. One important disadvantage is that, given a set of premisses containing many inconsistencies and insufficient knowledge about the relative reliability, the number of possible belief sets can become exponential in the number minimal inconsistencies detected.

Appendix A. Proofs

Theorem 4.7 (Soundness). *For each $i \geqslant 0$:*

$$\text{if } P \Rightarrow \varphi \in J_i, \text{ then}$$
$$P \subseteq \Sigma \quad \text{and} \quad P \models \varphi.$$

Proof. By the soundness of propositional logic,

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

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

$$\text{if } P \Rightarrow \varphi \in J_i, \text{ then } P \subseteq \Sigma \text{ and } P \vdash \varphi.$$

We can prove this by induction on the index i of J_i.

Basis: For $i = 0$:

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

Therefore, $\{\varphi\} \vdash \varphi$.

Induction step: 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 4.3 or 4.4.

- If $P \Rightarrow \varphi \in J_k$, then, by the induction hypothesis, $P \subseteq \Sigma$ and $P \vdash \varphi$.
- If $P \Rightarrow \varphi$ is introduced by Rule 4.3, then it is an axiom. Therefore, $P = \emptyset$ and $\vdash \varphi$.
- If $P \Rightarrow \varphi$ is introduced by Rule 4.4, then there exist $Q \Rightarrow \psi \in J_k$ and $R \Rightarrow (\psi \rightarrow \varphi) \in J_k$. Therefore, $P = (Q \cup R)$. According to the induction hypothesis, we have:

$$Q, R \subseteq \Sigma,$$
$$Q \vdash \psi,$$
$$R \vdash \psi \rightarrow \varphi.$$

Hence,

$$P \subseteq \Sigma \text{ and } P \vdash \varphi. \qquad \square$$

Theorem 4.8 (Completeness). *For each $P \subseteq \Sigma$:*

if $P \models \varphi$, then

there exists a $Q \subseteq P$ such that for some $i \geqslant 0$: $Q \Rightarrow \varphi \in J_i$.

Proof. Let $P \subseteq \Sigma$ and $P \models \varphi$. By the completeness of first-order logic, we have:

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

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

- $\varphi_j \in P$, or
- φ_j is an axiom, or
- there exist φ_k and 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.

Basis: For $n = 1$, $\langle \varphi_1 \rangle$ is the deduction sequence for $P \vdash \varphi$.

- If $\varphi_1 \in P$, then $\{\varphi_1\} \Rightarrow \varphi_1 \in J_0$.
- If φ_1 is an axiom, then there exists some $i_0 \geqslant 0$ such that:

$$J_{i_0} = J_{i_0 - 1} \cup \{\emptyset \Rightarrow \varphi_0\} \text{ and } \emptyset \Rightarrow \varphi_0 \text{ is added by Rule 4.3.}$$

Hence, the theorem holds for deduction sequences of length 1.

Induction step: Let $\langle \varphi_0, \varphi_1, \dots, \varphi_{m+1} \rangle$ be a deduction sequence for $P \vdash \varphi_{m+1}$.

- If $\varphi_{m+1} \in P$, then $\{\varphi_{m+1}\} \Rightarrow \varphi_{m+1} \in J_0$.
- If φ_{m+1} is an axiom, then there exists an i_{m+1} such that:

$$J_{i_{m+1}} = J_{i_{m+1} - 1} \cup \{\emptyset \Rightarrow \varphi_{m+1}\} \text{ and } \emptyset \Rightarrow \varphi_{m+1} \text{ is added by Rule 4.3.}$$

- If there exist φ_k and φ_l with $k, l < m + 1$ 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:

$$R \Rightarrow \varphi_k \in J_{i_k},$$

$$S \Rightarrow (\varphi_k \rightarrow \varphi_{m+1}) \in J_{i_l},$$

$$R, S \subseteq P.$$

Because of the fairness Assumption 4.6, there must exist an i_{m+1} with $i_k, i_l < i_{m+1}$ such that:

$$(R \cup S \Rightarrow \varphi_{m+1}) \in J_{i_{m+1}}.$$

Hence there exists some i_{m+1} such that $Q \Rightarrow \varphi_{m+1} \in J_{i_{m+1}}$ and $Q \subseteq P$. $\square$

Theorem 4.9 (Soundness). *For each $i \geqslant 0$:*

$$\text{if } P \not\Rightarrow \varphi \in J_i, \text{ then:}$$

$$(P \cup \{\varphi\}) \subseteq \Sigma \quad \text{and} \quad (P \cup \{\varphi\}) \text{ is not satisfiable.}$$

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

Basis: For $i = 0$ the theorem holds vacuously, because there is no $P \not\Rightarrow \varphi \in J_0$.

Induction step: Suppose that $P \not\Rightarrow \varphi \in J_{k+1}$. We have that $P \not\Rightarrow \varphi \in J_{k+1}$ if and only if $P \not\Rightarrow \varphi \in J_k$ or $P \not\Rightarrow \varphi$ has been added by Rule 4.5.

- If $P \not\Rightarrow \varphi \in J_k$, then, by the induction hypothesis, $(P \cup \{\varphi\}) \subseteq \Sigma$ and $(P \cup \{\varphi\})$ is not satisfiable.
- If $P \not\Rightarrow \varphi$ is introduced by Rule 4.5, then there exist $R \Rightarrow \psi \in J_k$ and $Q \Rightarrow \neg\psi \in J_k$ such that:

  $$\varphi = min(Q \cup R) \quad \text{and} \quad P = (R \cup Q)/\varphi.$$

  By Theorem 4.7:

  $$R, Q \subseteq \Sigma,$$

  $$R \vdash \psi \quad \text{and} \quad Q \vdash \neg\psi.$$

  Hence $(P \cup \{\varphi\}) \subseteq \Sigma$, and $(P \cup \{\varphi\})$ is inconsistent. Since inconsistency implies unsatisfiability:

  $$(P \cup \{\varphi\}) \subseteq \Sigma \quad \text{and} \quad (P \cup \{\varphi\}) \text{ is not satisfiable.} \qquad \square$$

Theorem 4.10 (Completeness). *For each $P \subseteq \Sigma$:*

$$\text{if } P \text{ is a minimal unsatisfiable set of premisses and } \varphi = min(P),$$

$$\text{then for some } i \geqslant 0\text{: } P/\varphi \not\Rightarrow \varphi \in J_i.$$

Proof. Let P be a minimal unsatisfiable subset of Σ with $\varphi = min(P)$. Since P is a minimal unsatisfiable set, P is a minimal inconsistent set. Therefore, there exists a proposition ψ such that:

$$P \vdash \psi \quad \text{and} \quad P \vdash \neg\psi.$$

By Theorem 4.8 there exists a $j, k \geqslant 0$ such that:

$$S \Rightarrow \psi \in J_j, \quad S \subseteq P,$$

$$T \Rightarrow \neg\psi \in J_k, \quad T \subseteq P.$$

Hence, $(S \cup T) \subseteq P$.

Since P is minimal inconsistent:

$$(S \cup T) = P.$$

Because of the fairness Assumption 4.6 there exists an $l > j, k$ such that:

$$(P/\varphi) \not\Rightarrow \varphi \in J_l. \qquad \square$$

Property 4.13. *For every i, the set Δ_i exists and is unique.*

Proof.

Existence: Let $\delta_0 \subset \delta_1 \subset \cdots \subset \delta_k$ be a sequence of sets of premisses such that:

- $\Sigma = \delta_0$,
- $\delta_{j+1} = \delta_j - \{\varphi\}$ where φ is the most reliable premiss in δ_j such that $P \not\Rightarrow \varphi$ and $P \subseteq \delta_j$.

Then, by induction on the index of the sequence, we can prove that:

$$\Sigma - Out_i(\delta_j) \subseteq \delta_j.$$

Basis: For $j = 0$, clearly, there holds $\Sigma - Out_i(\delta_0) \subseteq \delta_0$.

Induction step: Let the induction hypothesis hold for $l \leqslant j$.

If $\Sigma - Out_i(\delta_j) \subset \delta_j$, then there exists a most reliable $\varphi \in \delta_j$ such that $P \not\Rightarrow \varphi$ and $P \subseteq \delta_j$.

Now suppose that $\Sigma - Out_i(\delta_{j+1}) \not\subseteq \delta_{j+1}$. Then there exists a $\psi \notin Out_i(\delta_{j+1})$ such that $\psi \notin \delta_{j+1}$.

Suppose that $\psi \in \delta_j$. Then $\psi = \varphi$. Since φ is the most reliable premiss such that $P \not\Rightarrow \varphi$ and $P \subseteq \delta_j$, $P \subseteq \delta_{j+1}$. Hence, $\psi \in Out_i(\delta_{j+1})$. Contradiction.

Hence, $\psi \notin \delta_j$ and, by the construction of δ_j, $\varphi \prec' \psi$. Since $\psi \notin \delta_j$, by the induction hypothesis, $\psi \in Out(\delta_j)$. Therefore, there exists a $Q \not\Rightarrow \psi \in J_i$ such that $Q \subseteq \delta_j$. Since $\varphi \prec' \psi$, $Q \subseteq \delta_{j+1}$. Hence, $\psi \in Out_i(\delta_{j+1})$. Contradiction.

Hence, $\Sigma - Out_i(\delta_{j+1}) \subseteq \delta_{j+1}$.

Let k be the highest index in the sequence. Then there does not exist a $\varphi \in \delta_k$ such that $P \not\Rightarrow \varphi \in J_i$ and $P \subseteq \delta_k$. Hence, $\Sigma - Out_i(\delta_k) = \delta_k$, otherwise there would exist a $\varphi \in \delta_k$ such that $P \not\Rightarrow \varphi \in J_i$ and $P \subseteq \delta_k$. Hence, there exists at least one Δ_i such that:

$$\Delta_i = \Sigma - Out_i(\Delta_i).$$

Uniqueness: Suppose Δ_i is not unique. Then there exist at least two different subsets $\Delta_i, \Delta_i' \subset \Sigma$ satisfying Definition 4.12. Let φ be the most reliable proposition in Σ such that:

$$\varphi \notin \Delta_i \quad \text{and} \quad \varphi \in \Delta_i'.$$

Hence, there exists a $P \not\Rightarrow \varphi \in J_i$. By Theorem 4.10 we have:

$$P \cup \{\varphi\} \text{ is unsatisfiable.}$$

Therefore, there exists a minimal inconsistent set of premisses Q with $\varphi = min(Q)$. Since $\varphi \notin \Delta_i$ and $\varphi \in \Delta_i'$, there exists a $\psi \in Q$ such that:

$$\psi \in \Delta_i, \qquad \psi \notin \Delta_i', \qquad \varphi \prec \psi.$$

Hence, φ is not the most reliable proposition in Σ such that $\varphi \notin \Delta_i$ and $\varphi \in \Delta_i'$. Contradiction.

Hence Δ_i is unique. $\square$

Property 4.15. *For each* $\varphi \in B_i$:

$$\Delta_i \vdash \varphi.$$

Proof. Suppose $\varphi \in B_i$. Then there exists a $P \Rightarrow \varphi \in J_i$ such that $P \subseteq \Delta_i$. Therefore, by Theorem 4.7, $P \vdash \varphi$ and $P \subseteq \Delta_i$.

Hence, $\Delta \vdash \varphi$. $\square$

Property 4.17. Δ_∞ *is maximal consistent.*

Proof. Suppose that Δ_∞ is inconsistent. Then there exists a minimal inconsistent subset M of Δ_∞. Let $\varphi = min(M)$. Then by Theorem 4.10 there exists an i with

$$P \not\Rightarrow \varphi \in J_i.$$

Hence $P \not\Rightarrow \varphi \in J_\infty$. Because $P \subseteq \Delta_\infty$, $\varphi \notin \Delta_\infty$. Contradiction.

Suppose that some Δ_∞ is not maximal consistent. Then there exists a $\varphi \in (\Sigma - \Delta_\infty)$ such that $\{\varphi\} \cup \Delta_\infty$ is consistent. Since $\varphi \in (\Sigma - \Delta_\infty)$, $\varphi \in Out_\infty(\Delta_\infty)$. Therefore, there exists a $P \not\Rightarrow \varphi \in J_\infty$ such that $P \subseteq \Delta_\infty$. Since $P \not\Rightarrow \varphi \in J_\infty$, $P \cup \{\varphi\}$ is inconsistent. Hence $\Delta_\infty \cup \{\varphi\}$ is inconsistent. Contradiction. $\square$

Property 4.18.

$$B_\infty = Th(\Delta_\infty),$$

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

Proof. According to Property 4.15:

$$\text{if } \varphi \in B_\infty, \text{ then } \Delta_\infty \vdash \varphi.$$

Suppose there exists a φ such that:

$$\varphi \notin B_\infty \quad \text{and} \quad \varphi \in Th(\Delta_\infty).$$

Since $\varphi \in Th(\Delta_\infty)$, $\Delta_\infty \vdash \varphi$. By Theorem 4.8 there exist some i and some $P \Rightarrow \varphi \in J_i$ such that:

$$P \subseteq \Delta_\infty.$$

Therefore, there exists a $P \Rightarrow \varphi \in J_\infty$ such that:

$$P \subseteq \Delta_\infty.$$

Hence, by Definition 4.14: $\varphi \in B_\infty$. Contradiction.
Hence $B_\infty = Th(\Delta_\infty)$. □

Theorem 4.19. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Then there holds:*

$$\mathcal{A} = \{\Delta_\infty \mid \textit{for some linear extension of } \prec, \Delta_\infty \textit{ can be derived}\}.$$

Proof. Let Δ_∞ be a set of believed premisses given a linear extension $\prec'$ of $\prec$. Furthermore, let $\sigma_1, \ldots, \sigma_m$ be an enumeration of Σ such that for every $\sigma_j \prec' \sigma_k$: $k < j$. Clearly, given this enumeration of Σ, Δ_∞ will satisfy Definition 3.1.

Let D be a most reliable consistent set of premisses given an enumeration $\sigma_1, \ldots, \sigma_n$ of Σ. Furthermore, let $\prec'$ be a linear extension of $\prec$ such that for each $k < j$: $\sigma_j \prec' \sigma_k$. Observe that for each $\varphi \notin D$ there exists a minimal inconsistent set $\{\sigma_{i_1}, \ldots, \sigma_{i_n}\}$ with $i_j < i_{j+1}$ and $\varphi = \sigma_{i_n}$. Hence, by Theorem 4.10 and by the definition of the extended reliability relation $\prec'$:

$$(\{\sigma_{i_1}, \ldots, \sigma_{i_n}\} \not\Rightarrow \sigma_{i_n}) \in J_\infty.$$

Let $D \subseteq \Sigma$ be a set satisfying Definition 3.1. Now suppose that

$$D \neq \Sigma - Out_\infty(D).$$

Hence, there exists a most reliable premiss $\varphi \in \Sigma$ such that either

$$\varphi \in D \quad \text{and} \quad \varphi \in Out_\infty$$

or

$$\varphi \notin D \quad \text{and} \quad \varphi \notin Out_\infty.$$

If $\varphi \in Out_\infty$, then for some $P \not\Rightarrow \varphi \in J_\infty$ there holds $P \subseteq D$. Since $P \subseteq D$ and since $\varphi \in D$, D is inconsistent. By Definition 3.1, however, D must be consistent. Contradiction.

If $\varphi \notin Out_\infty$, then for no $P \not\Rightarrow \varphi \in J_\infty$ there holds $P \subseteq D$. Hence, for each $P \not\Rightarrow \varphi \in J_\infty$ there exists a $\psi \in P$ such that $\psi \notin D$. Since $\psi \notin D$, according to Definition 3.1, $D \cup \{\psi\}$ is inconsistent. Furthermore, by Definition 3.1 there exists a minimal inconsistent set of premisses Q containing φ such that:

$$Q/\varphi \subseteq D, \qquad Q = \{\sigma_{i_1}, \ldots, \sigma_{i_n}\} \quad \text{with } i_j < i_{j+1}.$$

Therefore $\psi = \sigma_{i_n}$ and $Q/\psi \not\Rightarrow \psi \in J_\infty$. Hence, $\psi \notin Out_\infty(D)$. Contradiction. $\square$

Theorem 6.6. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Furthermore, let $\mathcal{A}$ be the corresponding set of all most reliable consistent sets of premisses. Then:*

$$Mod_{\sqsubset}(\langle \Sigma, \prec \rangle) = \bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty),$$

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

Proof. The proof of

$$Mod_{\sqsubset}(\langle \Sigma, \prec \rangle) = \bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty)$$

can be divided into the proof of the soundness:

$$\bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty) \subseteq Mod_{\sqsubset}(\langle \Sigma, \prec \rangle)$$

and the proof of the completeness:

$$Mod_{\sqsubset}(\langle \Sigma, \prec \rangle) \subseteq \bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty)$$

of the logic.

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

$$\mathcal{M} \notin Mod_{\sqsubset}(\langle \Sigma, \prec \rangle).$$

Then there exists an interpretation $\mathcal{N}$ such that $\mathcal{M} \sqsubset \mathcal{N}$. According to Proposition 4.17, since $Prem(\mathcal{M}) = \Delta_\infty$:

$$\Delta_\infty \not\subseteq Prem(\mathcal{N}).$$

Let $\varphi \in \Delta_\infty$ be the most reliable premiss according to the linear extension $\prec'$ of $\prec$, such that $\varphi \in (\Delta_\infty - Prem(\mathcal{N}))$. Now by Definition 6.4 there exists a $\psi \in (Prem(\mathcal{N}) - \Delta_\infty)$ such that $\varphi \prec' \psi$. Since $\psi \notin \Delta_\infty$, there exists a $P \not\Rightarrow \psi \in J_\infty$ such that $P \subseteq \Delta_\infty$. Now, $P \not\subseteq Prem(\mathcal{N})$, otherwise $Prem(\mathcal{N})$ would be inconsistent. Hence, there exists a $\mu \in P$ such that:

$$\mu \in (\Delta_\infty - Prem(\mathcal{N})).$$

Since $P \not\Rightarrow \psi \in J_\infty$, $\psi \prec' \mu$. Hence, $\varphi \prec' \psi \prec' \mu$. Contradiction.

Hence,

$$\bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty) \subseteq Mod_{\sqsubset}(\langle \Sigma, \prec \rangle).$$

Soundness: Suppose there exists a structure $\mathcal{M} \in Mod_{\sqsubset}(\langle \Sigma, \prec \rangle)$ such that for each linear extension $\prec'$ of $\prec$ there holds:

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

Then we have the following two cases.

Case 1. There exists a φ such that

$$\varphi \in Prem(\mathcal{M}) \quad \text{and} \quad \varphi \notin \Sigma - Out_\infty(Prem(\mathcal{M})).$$

Hence, there exists a $P \not\Rightarrow \varphi \in J_\infty$ such that $P \subseteq Prem(\mathcal{M})$. Because $P \subseteq Prem(\mathcal{M})$, $Prem(\mathcal{M})$ is inconsistent. Contradiction.

Case 2. There exists a φ such that:

$$\varphi \notin Prem(\mathcal{M}) \quad \text{and} \quad \varphi \in \Sigma - Out_\infty(Prem(\mathcal{M})).$$

Then $Prem(\mathcal{M}) \cup \{\varphi\}$ is either consistent or inconsistent. If it is consistent, then we have for each structure $\mathcal{N} \in Mod(Prem(\mathcal{M}) \cup \{\varphi\})$ that $\mathcal{M} \sqsubset \mathcal{N}$. 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. Suppose that $\varphi = min(P)$. Then by Theorem 4.10 there exists a $P/\varphi \not\Rightarrow \varphi$. Since $P/\varphi \subseteq Prem(\mathcal{M})$, $\varphi \notin \Sigma - Out_\infty(Prem(\mathcal{M}))$. Contradiction. Hence $\varphi \neq min(P)$.

Let *MIN* be the set of all the premisses $min(P)$ for each minimal inconsistent subset P of $Prem(\mathcal{M}) \cup \{\varphi\}$. Since φ is in each minimal inconsistent set P and $\varphi \neq min(P)$, 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 $\eta \in (Prem(\mathcal{M}) - Prem(\mathcal{N}))$ there holds $\eta \prec' \varphi$, and because $\varphi \in (Prem(\mathcal{N}) - Prem(\mathcal{M}))$ we have: $\mathcal{M} \sqsubset \mathcal{N}$. Contradiction.

Hence,

$$Mod_{\sqsubset}(\langle \Sigma, \prec \rangle) \subseteq \bigcup_{\Delta_\infty \in \mathcal{A}} Mod(\Delta_\infty). \qquad \Box$$

Lemma 7.1. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. Furthermore, let*

$$\widehat{\alpha} = \{\mathcal{M} \mid \mathcal{M} \models \alpha\},$$

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

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

Then $\mathcal{M} \in Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$ if and only if $\mathcal{M} \in \widehat{\alpha}$ and for no $\mathcal{N} \in \widehat{\alpha} : \mathcal{M} \sqsubset \mathcal{N}$.

Proof.
Case 1. 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 6.3,

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

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

$$\varphi \prec' \alpha.$$

Hence, by Definition 6.4, for each $\mathcal{M} \in \widehat{\alpha}$ and $\mathcal{N} \notin \widehat{\alpha}$:

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

Case 2. Suppose that $\mathcal{M}, \mathcal{N} \in \widehat{\alpha}$. Since $\mathcal{M}, \mathcal{N} \models \alpha$, we have that 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
- $\psi \prec \varphi$ if and only if $\psi \prec' \varphi$.

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

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

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

Theorem 7.2. *Let $\langle \Sigma, \prec \rangle$ be a reliability theory. $\langle S, l, < \rangle$ is a preferential model for $\langle \Sigma, \prec \rangle$ if and only if S is the set of all possible interpretations for the language L, $l : S \to S$ is the identity function and for each $\mathcal{M}, \mathcal{N} \in S$:*

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

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

Suppose that $<$ is not smooth. Then by Lemma 7.1 for some proposition α and some $\mathcal{M} \in \widehat{\alpha}$ there holds neither that $\mathcal{M} \in Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$, nor does there exist an $\mathcal{N} \in Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$ such that $\mathcal{M} \sqsubset \mathcal{N}$.

Since $\mathcal{M} \notin Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$, there must exists an $\mathcal{L}_1$ such that $\mathcal{M} \sqsubset \mathcal{L}_1$. Suppose that for some $\mathcal{L}_i$ with $i \geqslant 1$ there does not exist an $\mathcal{L}_{i+1}$ such that $\mathcal{L}_i \sqsubset \mathcal{L}_{i+1}$. Then $\mathcal{L}_i \in Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$. Contradiction. Hence, there exists an infinite sequence $\mathcal{M} \sqsubset \mathcal{L}_1 \sqsubset \mathcal{L}_2 \sqsubset \cdots$.

For each $\mathcal{L}_i$ there exists a $\Gamma_i \subseteq \Sigma$: $\Gamma_i = Prem(\mathcal{L}_i)$. Suppose that for some $j < i$: $\Gamma_i = \Gamma_j$. Then $\mathcal{L}_j \not\sqsubset \mathcal{L}_i$. Contradiction. Hence, for each $\mathcal{L}_i$ and $\mathcal{L}_j$ with $i \neq j$ we have $\Gamma_i \neq \Gamma_j$.

Let $k = |\mathcal{P}(\Sigma)|$. Then $\{\Gamma_1, \ldots, \Gamma_k\} = \mathcal{P}(\Sigma)$. But there also holds that $\Gamma_{k+1} \in \mathcal{P}(\Sigma)$. Contradiction. Hence, $<$ is smooth.

Hence, $\langle S, l, < \rangle$ is a preferential model according to the definition of Kraus et al. [9].

Theorem 7.3. *Let $W = \langle S, l, < \rangle$ be a preferential model for $\langle \Sigma, \prec \rangle$. Then the following equivalence holds:*

$$\alpha \mid\!\sim_W \beta \quad \textit{if and only if}$$

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

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

$$\beta \in Th(\langle \Sigma', \prec' \rangle).$$

Proof. According to Theorem 6.6 we have that $\beta \in Th(\langle \Sigma', \prec' \rangle)$ if and only if for each $\mathcal{M} \in Mod_{\sqsubset'}(\langle \Sigma', \prec' \rangle)$: $\mathcal{M} \models \beta$.

Therefore, by Lemma 7.1: $\beta \in Th(\langle \Sigma', \prec' \rangle)$ if and only if for each $\mathcal{M} \in min(\widehat{\alpha})$: $\mathcal{M} \models \beta$.

Hence, by the definition of the nonmonotonic entailment relation $\mid\!\sim$ we have:

$$\beta \in Th(\langle \Sigma', \prec' \rangle) \quad \text{if and only if} \quad \alpha \mid\!\sim_{W} \beta. \qquad \square$$

Theorem 7.5. *Let belief set $K = Th(\langle \Sigma, \prec \rangle)$ be the set of theorems of the reliability theory $\langle \Sigma, \prec \rangle$. Suppose that $K^*[\alpha]$ is the belief set of the premisses $\Sigma \cup \{\alpha\}$ with reliability relation:*

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

i.e. $K^*[\alpha] = \{\beta \mid \alpha \mid\!\sim_W \beta\}$, where W is a preferential model for $\langle \Sigma, \prec \rangle$. Then the following postulates are satisfied:

(1) $K^*[\alpha]$ *is a belief set,*
(2) $\alpha \in K^*[\alpha]$,
(6) *If* $\vdash \alpha \leftrightarrow \beta$, *then* $K^*[\alpha] = K^*[\beta]$.

Proof. Postulate (1) follows from Property 4.18
Postulate (2) follows from $\alpha \mid\!\sim_W \alpha$ (reflexivity).
Postulate (6) is a result of

$$\frac{\models \alpha \leftrightarrow \beta, \alpha \mid\!\sim_W \gamma}{\beta \mid\!\sim_W \gamma} \quad \text{(left logical equivalence).} \qquad \square$$

Acknowledgement

I thank C. Witteveen for the discussion we had on the contents of this paper. Furthermore, I thank the reviewers. Due to their comments the paper has improved significantly. Finally, I thank the Delft University of Technology (TU-Delft) and and the National Aerospace Laboratory (NLR) for supporting the research reported on.

References

[1] A.I. Arruda, A survey of paraconsistent logic, in: A.I. Arruda, R. Chuaqui and N.C.A. da Costa, eds., *Mathematical Logic in Latin America* (North-Holland, Amsterdam, 1980) 1–41.
[2] G. Brewka, Cumulative Default Logic: in defense of nonmonotonic inference rules, *Artif. Intell.* **50** (1991) 183–205.
[3] J. de Kleer, An assumption-based TMS, *Artif. Intell.* **28** (1986) 127–162.
[4] J. Doyle, A truth maintenance system, *Artif. Intell.* **12** (1979) 231–272.
[5] H.B. Enderton, *A Mathematical Introduction to Logic* (Academic Press, New York, 1972).
[6] P. Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (Bradford Books/MIT Press, Cambridge, MA, 1988).
[7] M.L. Ginsberg and D.E. Smith, Reasoning about action I: a possible worlds approach, *Artif. Intell.* **35** (1988) 165–195.
[8] J.W. Goodwin, A theory and system for non-monotonic reasoning, Department of Computer and Information Science, Linköping University, Linköping, Sweden (1987).
[9] S. Kraus, D. Lehmann and M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, *Artif. Intell.* **44** (1990) 167–207.
[10] D. Poole, A logical framework for default reasoning, *Artif. Intell.* **36** (1988) 27–47.
[11] R. Reiter, A logic for default reasoning, *Artif. Intell.* **13** (1980) 81–132.
[12] N. Rescher, *Hypothetical Reasoning* (North-Holland, Amsterdam, 1964).
[13] N. Roos, A preference logic for non-monotonic reasoning, Report 88-94, Department of Computer Science, Delft University of Technology, Delft, Netherlands (1988).
[14] Y. Shoham, Reasoning about change, Dissertation Yale University, Department of Computer Science, New Haven, CT (1986) 85–109.
[15] Y. Shoham, Non-monotonic logic: meaning and utility, in: *Proceedings IJCAI-87*, Milan, Italy (1987) 388–393.