Investigations in Belnap's Logic of Inconsistent and Unknown Information

Dissertation zur Erlangung des akademischen Grades Dr. rer. nat.

der Fakultät für Mathematik und Informatik der Universität Leipzig

eingereicht von Dipl. Inform. Stefan Weber, Freudenberg angefertigt an der Universität Dortmund, FB Informatik

betreut von

Prof. Dr. Gerhard Brewka, Universität Leipzig und Prof. Dr. Heinrich Herre, Universität Leipzig

Beschluß über die Verleihung des Doktorgrades vom 19. Oktober 1998

Die Annahme der Dissertation haben empfohlen:

- 1. Prof. Dr. Siegfried Gottwald, Universität Leipzig
- 2. Prof. Dr. Heinrich Herre, Universität Leipzig
- 3. Prof. Dr. John-Jules Meyer, Universität Utrecht

Diese Arbeit wurde auf Apple Macintosh mit ${\rm IAT}_{\rm E\!X}$ unter Verwendung der Computer Modern Roman Zeichensätze gesetzt.

Bibliographische Beschreibung

Weber, Stefan

<u>INVESTIGATIONS IN BELNAP'S LOGIC OF INCONSISTENT AND UNKNOWN</u> <u>INFORMATION</u> Universität Leipzig, Diss.,

172 S., 73 Lit., 9 Abb.

Referat

Nuel Belnap schlug 1977 eine vierwertige Logik vor, die – im Gegensatz zur klassischen Logik – die Fähigkeit haben sollte, sowohl mit widersprüchlicher als auch mit fehlender Information umzugehen. Diese Logik hat jedoch den Nachteil, daß sie Sätze der Form *wenn ..., dann ...* nicht ausdrücken kann. Ausgehend von dieser Beobachtung analysieren wir die beiden nichtklassischen Aspekte, Widersprüchlichkeit und fehlende Information, indem wir eine dreiwertige Logik entwickeln, die mit widersprüchlicher Information umgehen kann und eine Modallogik, die mit fehlender Information umgehen kann. Beide Logiken sind nicht monoton. Wir untersuchen Eigenschaften, wie z.B. Kompaktheit, Entscheidbarkeit, Deduktionstheoreme und Berechnungkomplexität dieser Logiken.

Es stellt sich heraus, daß die dreiwertige Logik, nicht kompakt und ihre Folgerungsmenge im Allgemeinen nicht rekursiv aufzählbar ist. Beschränkt man sich hingegen auf endliche Formelmengen, so ist die Folgerungsmenge rekursiv entscheidbar, liegt in der Klasse Σ_2^P der polynomiellen Zeithierarchie und ist DP-schwer. Wir geben ein auf semantischen Tableaux basierendes, korrektes und vollständiges Berechnungsverfahren für endliche Prämissenmengen an. Darüberhinaus untersuchen wir Abschwächungen der Kompaktheitseigenschaft.

Die nichtmonotone auf S5-Modellen basierende Modallogik stellt sich als nicht minder komplex heraus. Auch hier untersuchen wir eine sinnvolle Abschwächung der Kompaktheitseigenschaft. Desweiteren studieren wir den Zusammenhang zu anderen nichtmonotonen Modallogiken wie Moores autoepistemischer Logik (AEL) und McDermotts NML-2. Wir zeigen, daß unsere Logik zwischen AEL und NML-2 liegt.

Schließlich koppeln wir die entworfene Modallogik mit der dreiwertigen Logik. Die dabei enstehende Logik MK_3 ist eine Erweiterung des nichtmonotonen Fragments von Belnaps Logik. Wir schließen unsere Betrachtungen mit einem Vergleich von MK_3 und verschiedenen informationstheoretischen Logiken, wie z.B. Nelsons **N** und Heytings intuitionistischer Logik ab.

Table of Contents

1 Introduction	2
1.1 Belnap's approach	2
1.2 Organisation	4
1.2.1 Goals	4
1.2.2 Summary	5
1.3 Notation	7
2 Paraconsistency: The Propositional Logic K ₃	8
2.1 Semantical Analysis	10
2.1.1 Three-Valued Interpretations	11
2.1.2 Cautious Entailment	14
2.1.3 Examples	17
2.2 Properties of Cn_3	18
2.2.1 Basic Closure Properties	19
2.2.2 Deduction Theorems	22
2.2.3 Existence of a Preferred Model	24
2.2.4 Systems of Cumulative Reasoning	25
2.3 Compactness	34
2.3.1 Compactness of Monotonic K_3	34
2.3.2 Paraconsistent Compactness	35
2.3.3 Weak Compactness	36
2.4 Computational Complexity	39
2.5 Extracting Consistent Information	43
2.6 Philosophical Aspects	47
2.6.1 Rescher-Based Approaches	48
2.6.2 Belief Revision	51
2.6.3 Relationship to Relevance logics	54
2.7 Discussion	56
2.7.1 Historical Remarks	56
2.7.2 Open Problems	57
2.7.3 Conclusion	57
3 Paraconsistency: The First-Order Logic FOK,	60
3.1 Terminology and Notations	60
3.2 Properties of FOK ₃	65
3.2.1 Herbrand-structures	65
3.2.2 Preferred Model Existence	67
3.2.3 Deduction Theorems	70
3.3 Extracting Consistent Information	72
3.4 Discussion	73

4 Proof Theory	76
4.1 A Hilbert System for Monotonic K ₃	78
4.2 A Hilbert Style Approximation for K_3	81
4.2.1 Context-dependent reasoning	81
4.2.2 Restricted Modus Ponens	83
4.2.3 Relatively Consistent Modus Ponens	83
4.2.4 Relativised Normalisation	88
4.2.5 Paraconsistent Extensions	91
4.3 A Hilbert-Style Approximation for FOK_3	93
4.4 Tableaux Systems	97 101
4.4.1 Bloesch's Tableau System	101
4.4.2 Extending Dioescn's Tableau Procedure	103 106
4.5 A Sequent Calculus	108
4.6 Conclusion	113
	110
5 Reasoning About Unknown Information	114
5.1 Semantical Investigations	115
5.1.1 Terminology	115
5.1.2 Unknown Means Satisfiability 1	116
5.2 Properties 1	119
5.2.1 Basic Properties 1	119
5.2.2 Other Properties 1	120
5.3 Adequacy of the Entailment Relation	122
5.3.1 Sublanguages which allow for unique preferred models	122
5.3.2 Treating Implicational Formulas as Rules	128
5.4 Relationship to Nonmonotonic Modal Logics	130
5.4.2 Autoepistemic Logics	131
5.4.3 Nonmonotonic Logic II	136
5.5 Historical Remarks and Related Work	137
5.6 Conclusion	137
6 Combining K ₃ and MK	138
6.1 Semantical Entailment	138
6.2 Properties 1	142
6.2.1 Basic Closure Properties 1	142
$6.2.2 \text{ MK}_3$ -Tautologies1	144
6.2.3 Normal forms	145
6.2.4 P-Stable Sets	147
6.3 Relationship to Belnap's Approach	148
6.4 Related Work	151
6.4.2 Logical Omnissionco	151 155
6.5 Conclusion	157 157
	101
7 Closure	158

iii

Preface

No matter what we get of this – I know, I know, we'll never forget.

BLACKMORE, GILLAN, GLOVER, LORD, PAICE.

I started the work on Belnap's Useful Four-Valued Logic as a consequence of trying to grasp how existing AI-systems deal with unknown and contradicting information. Supplying a slightly modified version of Belnap's original logic with a preferential entailment relation yielded a formal system which is very close to how the AI-programs behaved (cf. [Weber and Bell, 1994]).

There were, however, some problems with my formal system: it didn't have any tautologies and as a consequence I did not know how to provide a syntactical characterisation for my entailment relation. This was not very satisfying since I had the impression of having defined another entailment relation whose abstract properties and proof-theory have to remain underdeveloped¹.

As a consequence I started to reformulate Belnap's modified logic by separating the part which deals with contradicting information from the one which deals with unknown information. The first one is very close to Kleene's strong three-valued logic and I therefore called it K_3 . I did not have a name for the second one but when presenting some results on this logic of Unknown at the FAPR'96 conference, I learned from J.J. Meyer that the logic of Unknown is nothing other than a generalisation of Halpern's and Moses' logic *M*K; this solved at least the naming-problem².

While investigating the properties of MK and K_3 it turned out that both logics are indeed very close to classical propositional and classical modal logic. Many important properties like compactness and deduction theorems have weakened counterparts which do also hold for these logics.

As might be easily seen from looking at the table of contents, K_3 appeared very attractive to me. In contrast to many other competing approaches which loose properties like Reflexivity, AND-Property, K_3 retains all these aspects from

¹ I suspect that this is a common problem in the field of Knowledge Representation; new logics are defined but their abstract properties are very often disregarded. As a consequence, competing formal systems can only be compared by 'logic-benchmarks' which is from a theoretical point not always satisfying. Thus, if the goal is to learn more about the structure of a certain logic then it is certainly a plus, if the logic under consideration is very close to classical logic (or any other logic whose properties have been extensively studied). This yields what Quine called *Minimal Mutilation*.

² Joeri Engelfriet gave me very useful references of the usage of Halpern's and Moses' logic in AI.

classical logic. Of course, in general we had to give up monotonicity, but monotonic inference behaviour can be retained as long as we add information to Xwhich does not contradict X.

Finally! I could combine K_3 and MK to MK_3 – my logic for reasoning with unknown and contradicting information which is an alternative to Belnap's logic; maybe not the only alternative, but a very useful one. While by no means revolutionary, I hope the reader finds these investigations welcome and reassuring.

I tried to make this text self-contained (which means that if you take the literature on logic into account then you won't need any additional explanation ;-).

Part of the material presented in this text has been published elsewhere.

- A three-valued logic for reasoning with unknown information appeared in the Proceeding of the Workshop on Inductive Logic Programming, 1993, [Bell and Weber, 1993a], [Bell and Weber, 1993b]. A four-valued variant of the logic K₃ has been published as a technical report, [Weber and Bell, 1994].
- 2. Some of the complexity results and a proof-theory restricted to Horn-clauses for K_3 has been presented at the Extensions of Logic Programming Workshop at the International Joint Conference and Symposium on Logic Programming, [Weber, 1996b].
- 3. The Mathematical Properties of K_3 and the sequent-style calculus for K_3 has been presented at the Congress on Paraconsistency, [Weber, 1997b], [Weber, 1997a].
- 4. The logic MK, has been presented at the FAPR '96 Practical Reasoning Conference, [Weber, 1996a].

Acknowledgments

I am very thankful to teachers, supervisors, colleagues, friends, and my family.

Joachim Hertzberg and Gerd Brewka encouraged me again and again over the past years. I benefit a lot from their comments on earlier papers as well as on this piece of text.

My colleagues Siggi Bell, Andreas Hallmann, Hagen Huwig and Anke Rieger helped me a lot to clarify my thoughts and to put my mind at rest during a coffee break. Mark Pattison made many comments on my use of the English language in this text, and in general.

John-Jules Meyer helped me a lot by discussing the logical aspects of paraconsistent and unknown information. I am thankful to him for his comments on an earlier version of the thesis.

I am especially indebted to my supervisors Gerd Brewka and Heinrich Herre – they were both wonderful. The discussions with Gerd Brewka had a strong impact on the overall structure of this thesis. I owe him many thanks for his constructive support. From Heinrich Herre I learned much about compactness

vi

and inference-frames. I am grateful to him for urging me to explore the mathematical aspects of paraconsistent logic. His comments greatly improved many of the results.

CHAPTER 1

Introduction

Ever since the achievement of a formal definition of the concepts of knowledge and belief, much effort has been made in the field of Philosophical Logic and Artificial Intelligence. One of the seminal works in this field is Belnap's *How a computer should think* and its successor *A useful four-valued logic*. Belnap tried to account for two major problems: First, the information in a database is never complete, and, second it is very likely that this information is contradictory. As a consequence, the use of classical logic is inappropriate because of the so-called ex-falso principle, according to which a contradiction sanctions the entailment of any formula.

Belnap's way out of this disaster was to come up with a four-valued logic. By choosing an additional designated truth-value, b, Belnap manages $\{A, \neg A\}$ to have a model and hence, the entailment relation is prevented from trivialisation¹. Logics which cannot be trivialised are called *paraconsistent logics*. Since nontrivialisability or paraconsistency is nothing other than the failure of the classical ex-falso principle (EFQ) this concept does not imply a unique logic. There is a wide variety of paraconsistent logics. This yields different opinions about how a computer should think in the presence of contradictory information. The aim of the Introduction is to formulate a wishlist for a paraconsistent entailment relation which might be useful for Computing Science.

1.1 Belnap's approach

[Belnap, 1977] starts with setting up a scenario of a computer which finds itself in a situation where it has been told, for example, that A holds, B does not hold etc. Belnap identifies two points which represent a major difference between how the computer should think and how classical logic operates.

The first point is the inferential behaviour in the presence of inconsistent information. As an example, Belnap pictures the situation where Elizabeth tells the computer that the Pirates won the Series in 1971, while Sam tells it otherwise. If the computer is a classical logician then the contradicting information on the 1971 Series justifies the inference of anything.

The second point is that the computer should also serve as a questionanswering device. It must therefore be able to identify those sentences which

¹ Adding a third designated truth-value is a very simple but effective method to invalidate the ex-falso rule of classical logic. It should be mentioned that a wide class of many-valued logics developed since the beginning of this century use this trick.

it does not know. For example, if it does not have any information on the 1966 Series then any question concerning this event should be answered by 'I do not know'.

Belnap stresses that the computer should merely answer questions by what it has been told. This yields that, basically, we have four possibilities of how a question can be answered: just told True, just told False, told neither True nor False, told both True and False. Each possible answer can be associated with a truth-value: t means just True, f means just False, u unknown, and b both True and False.

In order to get a formal system out of these ideas, Belnap defines a set- up^2 as a mapping from the set of atomic formulas to the set of truth-value $\{t, f, b, u\}$. The truth-value of complex formulas can be recursively obtained by means of the following truth-tables:

$t \mid f \mid u \mid b$	\vee	t	f	b	u	Λ	t	f	b	u
$\sim f t b u$	t f b	t t t	t f b	t b b	t u t	t f b	t f b	f f f f	b f b f	$\begin{array}{c} u\\ f\\ f\\ u \end{array}$

Please note that Belnap has no table for an implicational connective. The entailment relation is now defined as: a sentence Φ entails Ψ if and only if $s(\Phi) \leq s(\Psi)$ for every set-up s. The relation \leq is defined by the following lattice:



Let us call the above logic, i.e. the truth-tables and the entailment relation, **L4**. This logic is known to be monotonic and to have a nice and easy proof theory (see [Wagner, 1994] for a Sequent System for **L4**). However, it also has a big disadvantage: one cannot express rules like 'if A then B'. The only reasonable

² A set-up is a synonym for a truth-assignment function.

way to define an implicational connective from the above truth-tables is by means of material implication. Unfortunately, material implication does not work for L4. Here is what Belnap says:

Your first thought might be that you could get the effect of 'given A and B, infer C,' or 'if A and B, then C,' by feeding the computer ' $\sim A \lor \sim B \lor C$.' But that won't work: the latter formula will tend to split the set-up you've got into three, one in which A is marked told False etc.

Thus, if we tell the computer $A, B, \sim A \lor \sim B \lor C$ we get at least three set-ups: one in which A is marked both, true and false, B is just marked true and C marked false. Another one in which A is just marked True and B is both, True and False, C is just marked False etc.

To account for this problem, Belnap considers an implication as a mapping from epistemic states to epistemic states. Thus it reads as: if A is true in some state s then change s minimally in order to make B true. Hence, an implicational statement adds extensional information.

Belnap points out several drawbacks of his solution to the problem of implicational input. First, he says that he did not succeed in giving a *logic* for adding rules $A \to B$ to the database. Second, the computer cannot answer questions about the truth-status of implications like $A \to B$. In addition to that, handling implications in Belnap's logic is a metalogical enterprise. That is, they are read as inference rules: if A is true, then update the knowledge-base by B. Clearly, this differs from adding the sentence $A \to B$ to a set X of formulas a letting the entailment relation do all the work. To me it seems, that an all-in-one logic could be much more appealing.

1.2 Organisation

1.2.1 Goals

The general goal is to develop a logic which is able to handle implicational input, and which allows proper reasoning about contradicting and unknown information. Of course, the term 'proper reasoning' needs a careful analysis of (1) what should be entailed by contradicting information and (2) what does it mean if computer answers 'I do not know A'?

Another important point is to stick to what Quine calls *Minimal Mutilation*: we want to stay as close as possible to classical logic. This demand has a lot of consequences:

- 1. If X is consistent, then we wish the paraconsistent consequences of X to be identical to the classical consequences. This is a property which is required for so-called Adaptive Logics.
- 2. We wish to retain as many properties of classical logic as possible. Especially,
 - The paraconsistent consequence relation should be reflexive. In my point of view reflexivity is *the* basic property of a deductive system: you get at least out what you have put in. Any irreflexive relation cannot be a consequence relation (it's a sort of transformation system which, like a program, performs state transitions).
 - The connectives \lor and \land should behave normally, i.e. $A \land B$ is entailed if and only if A and B are both entailed (AND property). Moreover, if we can conclude A then we should also be able to conclude $A \lor B$ (OR property).
 - Contraposition should be valid, i.e. from $A \to B$ we wish to conclude $\sim B \to \sim A$.
 - The consequence relation should (at least for the propositional case) be recursive, i.e.decidable.

The above characterises our main demands on the behaviour in the presence of contradicting information. As for the missing information, we wish that ∇A , to be read as A is unknown, is entailed by X if and only if neither A nor $\sim A$ is entailed by X. If ∇A is provable, then A and $\sim A$ are both not provable. Thus, ∇ encodes some concept of provability. In order to make the whole thing become meaningful, we have to require that the consequence relation is decidable. Only in this case we can compute whether some formula is entailed or not. The computer can therefore verify if ∇A holds, or not. Hence, we wish that for formulas of the type ∇A the tertium non-datur principle holds, that is each corresponding semantical structure satisfies either ∇A or $\sim \nabla A$. In other words: we do not wish to take the impact of paraconsistency so far that $\nabla A \wedge \sim \nabla A$ has a model.

1.2.2 Summary

Let me give a summary of the main results and an overview of the structure of this text. The basic plot is to decouple aspects of contradicting and unknown information. That is, we shall first introduce a logic which is able to reason properly in the presence of contradicting information. Next, we shall develop a modal system for reasoning about unknown information. The last step is to combine both systems in order to be able to reason about contradicting *and* unknown information.

Chapter 2 We shall investigate how a computer should answer questions in the presence of contradicting information. Starting from a re-definition of the truth-value semantics for the basic connectives $\neg, \lor, \land, \rightarrow$ we obtain a logic whose entailment relation, which is based on Shoham's preferred model, is identical to the one given by [Priest, 1991].

We analyse in-depth the properties of this logic. We show that the corresponding consequence operator Cn_3 is a pre-closure operator which satisfies various variants of the deduction theorem. However, Cn_3 is not compact even though several weaker versions of the compactness theorem hold. Moreover, Cn_3 satisfies all important properties of Kraus, Lehman and Magidor's System **C** and **P**.

We show that the problem of computing a preferred model for X is in LINTIME^{NP[O(log n)]}; whether it is also LINTIME^{NP[O(log n)]}-hard is an open problem. Deciding $Cn_3(X)$ can be done by a polynomial time bounded TM which has an NP-oracle. Only a linear amount of NP-oracle calls need to be made. Thus, retaining inconsistent information is not more expensive than revising the knowledge-base by means of Belief Revision, which is Π_2^P -complete for most operators.

- Chapter 3 In this chapter we shall discuss the first-order version FOK_3 of K_3 . The logic FOK_3 is much more troublesome than K_3 . For example, we cannot guarantee that every set which has a model does also have a preferred model. Moreover, Herbrand's theorem does not hold for K_3 . In order to guarantee basic properties we have to restrict ourselves to universal theories.
- Chapter 4 This chapter is devoted to the proof theory for K_3 and FOK_3 . We show that $Cn_3(X)$ is Σ_2^0 -hard, i.e. not even recursively enumerable. We develop the concept of recursively enumerable approximation to give a proof-theory for K_3 . We shall present three different proof-systems: a Hilbert-style calculus, a tableau-based calculus and a sequent-style calculus. We show soundness for each of them.
- Chapter 5 Here we account for the problem of unknown information. We define a modal operator \diamond such that we can conclude $\diamond \Phi \land \diamond \sim \Phi$ from X, whenever Φ cannot be inferred from X, provided that X has a unique preferred model. We further show by providing a syntactical characterisation that this logic MK perfectly fits in the gap between Moore's autoepistemic logic and McDermott's NML-2.
- Chapter 6 We can now combine K_3 and MK by supplying MK's possible world semantics by three-valued interpretation functions. We give a fixpoint characterisation of a Hilbert-style proof-system for MK_3 . It turns out that MK_3 and Belnap's logic agree on the main aspects. That is, if Belnap's logic judges

a formula to be unknown, then MK_3 judges this formula to be unknown, even though they are in general incomparable. This is mainly because MK_3 's language is richer. If, however, we restrict ourselves to nonmodal Φ , we can show that every Φ entailed by Belnap's logic is also entailed by MK_3 .

1.3 Notation

We shall mostly deal with a propositional language \mathcal{L} closed under the usual connectives $\neg, \rightarrow, \lor, \land$; note, that Belnap's negation operator will be replaced by \neg . In Chapter 5 we shall extend \mathcal{L} by some modal operators \Box and \Diamond . Precisely, let $\Sigma = \{A, B, C, \ldots\}$ be a countable set of propositional variables; we call Σ a propositional signature. Every member of Σ is also called an *atomic formula*. The set of *well-formed formulas* w.r.t a given signature Φ is the smallest set Form (Σ) such that

2. if Φ, Ψ in Form (Σ) then $\neg \Phi, \Phi \rightarrow \Psi, \Phi \land \Psi, \Phi \lor \Psi \in Form (\Sigma)$.

If it is clear from the context or not important, I shall omit the reference to a special signature Σ and then just talk about a language \mathcal{L} . If X is a set of propositional formulas then $\operatorname{ATOM}(X) =_{\operatorname{def}} \{A \mid A \in \Sigma \text{ and } A \text{ appears as}$ subformula in some $\Phi \in X\}$.

Capital letters like A, B, C etc will normally be used to denote atomic formulas, whereas Φ, Ψ, Ξ etc will be used to denote arbitrary formulas. Letters like X, Y, Z will be used to denote sets of formulas.

A *literal* is an atomic formula preceded by an odd number of negation signs (*negative literal*) or an even number of negation signs (*positive literal*).

The degree $d(\Phi)$ of a propositional formula Φ is defined as follows

 $d(\Phi) =_{\scriptscriptstyle def} \begin{cases} 0 & \text{if } \Phi \text{ is atomic} \\ d(\Phi') + 1 & \text{if } \Phi \text{ has the form } \neg \Phi' \\ max(\Phi', \Phi'') + 1 & \text{if } \Phi \text{ has the form } \Phi' \land \Phi'' \text{ or } \Phi' \lor \Phi'' \text{ or } \Phi' \to \Phi'' \end{cases}$

The concept of the degree of a formula will mainly be used in inductive proofs.

Any other conventions shall be introduced when we need them.

^{1.} $\Sigma \subseteq$ Form (Σ) and

CHAPTER 2

Paraconsistency: The Propositional Logic K₃

This chapter discusses how a computer should deal with contradicting information. We assume that the computer has been given a set of sentences, also called the database. We assume further that the user who puts queries to the database is familiar with classical logic and that he expects the computer to answer questions according the principles of classical logic. There is, however, another important presupposition the user makes: he assumes the database to be consistent. I think any user of a question-answering system assumes that the information given by system is consistent. For instance, if you consult your lawyer and ask for an information then you always assume that the information he gives you is correct; otherwise it would not make sense to ask him.

The same holds for a database system. You assume that the answers are correct and thus, that the database is consistent. But, what if not? Suppose the database has been built up by different experts. One expert might have asserted that *High taxes will help to lower the unemployment rate* while another expert is quite sure that *High taxes will NOT help to lower the unemployment rate*. If the latter wants to update the database by *High taxes will NOT help to lower the unemployment rate*, then the computer could reject this input because it is inconsistent with what he already has been told.

In order to accept the input *High taxes will NOT help to lower the unemployment rate*, we must prevent the computer from believing the contrary, i.e. we must remove the information *High taxes will help to lower the unemployment rate.*

We are confronted with two problems here: first, checking consistency is NPcomplete. Checking consistency each time an input occurred could result in a quite useless system. Second, withdrawing information from the database is nontrivial, because e.g. *High taxes will help to lower the unemployment rate* must not be given explicitly but can be implicitly inferred from the database. We also might have several possibilities to prevent the inference of *High taxes* Computing all those possibilities is quite complex as well. Even if we finally manage to compute all the pieces of information, which sanction the belief in *High taxes will help to lower the unemployment rate*, which piece should be removed?

We thus can conclude that maintaining consistency is extremely difficult (it is in fact Π_2^P -complete for many update operators):

- Checking for consistency is NP-complete

- Finding the possible culprits is a nontrivial task.

- Deciding which possible culprit should be deleted from the database might require expert-knowledge and additional user-interaction.

With respect to the above problems, we decide not to remove information from the database but to remove the classical consequence relation and replace it by some paraconsistent consequence relation. Since there is great variety of paraconsistent consequence relations, the question is: How should a useful, paraconsistent consequence relation behave? Especially, how should a paraconsistent consequence relation handle implicational input?

The very principle is that the truth status of contradicting sentences is quite doubtful. We therefore do not want to believe any sentence whose justification relies on contradicting information. For example, assume that we have the following information:

- 1. If social standards are on the decline, then the government will not be reelected.
- 2. Social standards are on the decline.
- 3. Social standards stagnate.

We have contradicting information concerning the development of social standards. A safe way of dealing with contradictions would be to put the contradiction aside¹ and see what follows from the rest. In the above case only 1) would remain and we cannot conclude that 'the government will not be re-elected', which is reasonable, because it would rely on the information that social standards decrease (which is quite doubtful since we have also information contradicting this).

Therefore, we have that 1)-3) should ideally not entail that the government will be re-elected. This means that modus ponens (MP) is not a valid rule of inference.

It is quite bizarre to have a logic in which modus ponens is not a valid rule of inference. Clearly, this point requires further explanation. We want to block the entailment of a sentence B from $A, A \to B$ only if the truth-status of A is doubtful, i.e. only if we have information contradicting A. Thus, if there is no contradiction to A, then MP can be applied.

Let us put together a little wishlist for the desired entailment relation:

Paraconsistency: There is a B which is not a consequence of $\{A, \neg A\}$.

¹ Of course, the putting-aside-operation should not be accomplished by means of belief revision because we have just argued that the attempt to re-establish consistency is fairly complex. The putting-aside-operation should only serve as a metaphor.

Conservativity: Whenever X is consistent then the paraconsistent consequences should coincide with the classical consequences. A special case is the empty set. Hence, the tautologies of our new logic should be identical to the classical ones.

Cautiousness: $\{A, \neg A, A \rightarrow B\}$ should not entail B.

Preclosure: The new entailment operator should be at least inclusive and idempotent.

This chapter is structured as follows: we start by defining a three-valued paraconsistent entailment relation. The corresponding entailment relation has been independently defined in [Priest, 1991]. Much effort is spent on a mathematical (Section 2.2) and computational (Section 2.4) analysis of the entailment relation. The main results are:

- The consequence operator is a preclosure operator for which various versions of the deduction theorem hold. Moreover, it is conservative, cautious and enjoys all properties of Kraus, Lehmann and Magidor's abstract systems C and P except for Right Weakening.
- Our paraconsistent logic is not compact. There are, however, some interesting weakened versions of the compactness theorem which hold for our logic. For example, if $A \wedge \neg A$ is entailed by X, then it is already entailed by a finite subset of X.
- Computing a preferred model for a finite set X is from the standpoint of Turing-reducibility as difficult as the corresponding problem for classical logic. However, deciding whether Φ is a paraconsistent consequence of X is in Π_2^P , i.e. on the second level of the polynomial time hierarchy. Thus, reasoning in our paraconsistent logic is not more complex than revising the beliefs via a belief revision operation (which is Π_2^P -complete).
- We isolate a class of tautologies which have the so-called variable sharing property. This property plays an important role for Relevance Logics.

2.1 Semantical Analysis

From a semantical point of view, the reason for the validity of the ex-falso principle (EFQ) in classical logic is that the set $\{A, \neg A\}$ does not have a model. The definition of the entailment relation as 'X entails Φ if and only if every model of X is a model for Φ ' yields that an inconsistent set (i.e. a set having no model at all) entails every formula of a given language. The aim is to define the concept of a model such that every set (even those which are not satisfiable by a classical, two-valued interpretation) will have a model.

2.1.1 Three-Valued Interpretations

The easiest way to guarantee that the set $\{A, \neg A\}$ has a model is to introduce an additional designated² truth-value. Denote this value by \top . Now let *I* be an assignment from the set of atomic formulas to the set of truth-values $\{t, f, \top\}$ (true, false and paraconsistent). Consider the following truth-table for the negation operator \neg (note that this table differs from Belnap's negation \sim):

$$\frac{|t| \top |f|}{\neg |f| \top |t|}$$

Since \top is designated, we have that $I(A) = \top$ satisfies $\{A, \neg A\}$. Whenever some formula A is assigned the value \top , this should represent something like 'there is information indicating that both, A and $\neg A$ have been told'.

The above semantics for \neg is not merely a technical trick to invalidate EFQ; it is also reasonable in our case to choose this semantics. We said that the truthstatus of any contradicting sentence is doubtful; hence any sentence which has been assigned \top has a doubtful truth-status. But if the truth-status of A is doubtful, so is the truth-status of $\neg A$. Therefore, $\neg A$ should also receive the value \top . This justifies the above truth-table.

Let us motivate the semantics for the connectives \vee and \wedge . Assume that our database contains some contradicting sentences, say Φ and $\neg \Phi$. We assume that, even if both Φ and $\neg \Phi$ have been told to the database, only one of them holds in the real world. Thus, after some knowledge revision process, the database either knows Φ or it knows $\neg \Phi$. In other words, even if the truth-status of some sentence is doubtful at some point of time, this sentence will turn out to be true or turn out to be false sometime in the future.

Consider the sentence $A \wedge B$. If both, A and B receive a truth-value from $\{t, f\}$, then the truth-value of the conjunction is identical to its truth-value in classical logic. This guarantees that the truth-table for \wedge will be a conservative extension of the classical semantics. Thus,

² Recall that in a multi-valued setting the set of truth-values is divided into two subsets: the set of 'truth-like' and the set of 'false-like' values. The first set is called set of designated truth-values whereas the latter is called set of non-designated truth-values. Designated truth-values generalise the concept of tautology: any sentence which takes a designated truth-value under every assignment is called a tautology. In other words, an assignment I satisfies a formula Φ if and only if $I(\Phi)$ is designated. In the case of classical logic t is the only designated truth-value.

\wedge	t	Т	f
t	t	?	f
Т	?	?	?
f	f	?	f

If we want to fill the gaps we have to keep in mind that any sentence which receives the value \top will in the long run turn out to be either t or f. Now, if I(A) = f and $I(B) = \top$, then no matter what B will turn out to be, the conjunction $A \wedge B$ will always be false. On the other hand if I(A) = t and $I(B) = \top$, then the truth-value of the compound statement depends on what B turns out to be; thus $A \wedge B$ receives in this case the value \top . A similar argument holds, if both A and B are assigned the value \top . This yields the following truth-table:

Analogous arguments hold for the connective \lor . If we define implication as material implication, then we get the following truth-tables:

t	$\top \int f$	V	t	Т	f	_	\rightarrow	t	Т	f
$\neg f$	\top t	t	t	t	t		t	t	Т	f
	·	Т	t	Т	Т		Т	t	Т	Т
		f	t	Т	f		f	t	t	t

The above truth-tables³ serve as a basis for a satisfiability relation \models . Let Σ be a propositional signature, i.e an enumerable set of propositional variables;

³ These tables are identical to the ones Kleene gave for his strong three-valued logic. [Kleene, 1952] did, however, consider the third truth-value to mean something like 'undefined' to account for the non-recursiveness of functions. As a consequence, Kleene's third truth-value is non-designated which yields a logic where $A \rightarrow A$ is not a tautology. Moreover Kleene's logic without a second designated truth-value does not have any tautologies at all

a function $I : \Sigma \to \{t, f, \top\}$ is called a *three-valued interpretation function*. The function I will be extended in the usual way to determine recursively the truth-value of any $\Phi \in \mathcal{L}$ by means of the above truth-tables. We say that I is a K₃-model of a formula Φ , denoted by $I \models \Phi$ if and only if $I(\Phi) \in \{t, \top\}$. The relation \models is naturally extended to sets. A class \Im of interpretations satisfies a set X of formulas if and only if $I \models X$, for every $I \in \Im$. The class of all models of a given set X is denoted by $\mathsf{MOD}(X)$.

The truth-values of a compound formula can be characterised by associating with each member of $\{t, f, \top\}$ a value from $\{0, 1, 2\}$. We assign t the value 2, \top the value 1 and f the value 0. By considering the standard relation \leq among natural numbers we obtain a linear ordering which can be visualised by the following Hasse-diagrams:



Please note that this ordering coincides with Belnap's ordering \leq mentioned in Chapter 1. The only difference is that Belnap uses \leq to (partially) order his set of four truth-values, while we excluded the value u from our considerations.

There is a close relationship between the connectives \lor , \land , \neg and lattice operations:

$$A \wedge B = min(A, B)$$
$$A \vee B = max(A, B)$$
$$\neg A = 2 - A$$

where $A, B \in \{0, 1, 2\}$.

Now that we have defined the semantics for our propositional language, we can start to define an entailment relation.

2.1.2 Cautious Entailment

We identified the classical entailment relation ('every two-valued model of X is also a two-valued model of Φ ') as the sole culprit for the proliferating set of consequences in the case of a contradiction (EFQ)⁴. Let us replace the classical two-valued entailment relation by its three-valued counterpart $||_{Bolz}$. We say that $X ||_{Bolz} \Phi$ if and only if every three-valued model of X is also a three-valued model of Φ . Since entailment relations of the form 'every model of X is also a model of Φ ' are due to Bolzano, we use the subscript 'Bolz' for the above three-valued entailment relation. Note that the only difference between these two relations is the type of model they talk about (two-valued or three-valued).

How far can we go with \parallel_{Bolz} in order to achieve our goals of the aforementioned wishlist (paraconsistency, conservativity etc.)? Not too far, unfortunately. For example, conservativity is violated as shown by the following example:

Example 2.1. Let $\Sigma = \{A, B\}, X = \{A, A \to B\}$. The following three-valued interpretations satisfy X:

$$\begin{array}{c|c} I_1(A) = t & I_2(A) = t \\ I_1(B) = t & I_2(B) = \top \end{array} \begin{array}{c|c} I_3(A) = \top & I_4(A) = \top & I_5(A) = \top \\ I_3(B) = t & I_4(B) = f \end{array} \begin{array}{c} I_5(A) = \top & I_5(B) = \top \end{array}$$

Because there is an interpretation, namely I_4 , which satisfies X but not B, we have that $X \not\Vdash_{Bolz} B$.

The problem is that the above I_4 interprets A to be paraconsistent but – when looking at X - A cannot be suspected to be paraconsistent at all. There is nothing which indicates that both A and $\neg A$ have been told. Interpreting A to be paraconsistent can be seen as an overdefinition of A's truth-value. This holds for any of the above models, except for I_1 .

In order to force an entailment relation not to care about overdefined models we shall replace \parallel_{Bolz} by an entailment relation $\parallel \vdash$ which bases on Shoham's idea of *preferred models* [Shoham, 1988]. A model is preferred if it does not overinterpret any sentence of X. The entailment relation is then defined as: X (preferentially) entails Φ if and only if every *preferred* model of X is a model of Φ '.

⁴ One might find the term EFQ inappropriate because the truth-value f does not play the same role as in classical two-valued logic. In [Wagner, 1994] this is replaced by ECSQ (ex contradictione sequitur quodlibet) which might be better. Since the EFQ principle is widely known, I would like to stick to it, though.

The relative degree of overinterpretation among models can be expressed by a relation \Box . In our case, the task of the ordering relation \Box is to rule out all models, in which the truth-value of some atomic formula is unnecessarily overdefined. Consider the following ordering relation to be read as 'less-informed' among truth-values. Note that this semilattice is similar to Belnap's lattice L4 and to the lattice which is induced by the ordering relation \leq of Section 2.1.1 if we put it on its side and replace b by \top^5 .



According to the above semi-lattice we have that t is less informed than \top and f is less informed than \top . In order to express that f and t are both equally informed we wish that $f \sqsubseteq t$ and $t \sqsubseteq f$ holds (note that this is not reflected by the above Hasse-diagram. We thus associate with each truth-value a degree of information i, i.e. a mapping $i : \{t, f, \top\} \rightarrow \{0, 1\}$ with i(t) = 0, i(f) = 0 and $i(\top) = 1$. Consider the linear ordering 0 < 1 and define the relation \Box to be the set of ordered pairs (a, b) such that i(a) < i(b). Moreover, define $a \equiv b$ if and only if i(a) = i(b). We write $a \sqsubseteq b$ if $a \sqsubset b$ or $a \equiv b^6$.

Proposition 2.1. The relation \sqsubseteq is a partial ordering on the set of truth-values $\{t, f, \top\}$.

Proof. Clearly we have, $a \sqsubseteq a$ for every $a \in \{t, f, \top\}$ (reflexivity), and $a \sqsubseteq b$ and $b \sqsubseteq c$ implies $a \sqsubseteq c$ (transitivity). Moreover, $a \sqsubseteq b$ and $b \sqsubseteq a$ implies $a \equiv b$ (anti-symmetry).

This partial order on truth-values can naturally be extended to three-valued interpretations I_1, I_2 of a signature Σ :

 $I_1 \sqsubseteq I_2$ if and only if $I_1(A) \sqsubseteq I_2(A)$ for all $A \in \Sigma$

⁵ In his paper [Belnap, 1977], Belnap discusses philosophical aspects of both lattices.

 $^{^6}$ The above definition of the relations \square and \square shows the relationship to Belnap's

ideas. We can also write them is a straight forward manner: $\sqsubseteq = \{(a, b) \mid a \neq \top \text{ or } b = \top\}$ and $\sqsubset = \{(a, b) \mid a \sqsubseteq b \text{ and NOT } b \sqsubseteq a\}$.

The relation \sqsubset is defined in an analogous way:

$$I_1 \sqsubset I_2$$
 if and only if $I_1 \sqsubseteq I_2$ and $I_1(A) \sqsubset I_2(A)$, for some $A \in \Sigma$.

Proposition 2.2. If $I \subseteq J$ then for all formulas Φ we have $I(\Phi) \subseteq J(\Phi)$.

Proof. By structural induction on the degree $d(\Phi)$ of Φ . For $d(\Phi) = 0$, i.e. for atomic Φ , the proposition follows immediately from the definition of \sqsubseteq . Assume that the proposition holds for all formulas with degree at most n. For the inductive step we have to consider several cases:

$$\begin{split} \Phi &= \neg \Psi \text{ Since } I(\Psi) \sqsubseteq J(\Psi) \text{ we have immediately that } I(\neg \Psi) \sqsubseteq J(\neg \Psi). \\ \Phi &= \Psi \lor \Psi' \text{ Assume to the contrary that we do not have } I(\Psi \lor \Psi') \sqsubseteq J(\Psi \lor \Psi'). \end{split}$$

That is we must have $I(\Psi \lor \Psi') = \top$ and $J(\Psi \lor \Psi') \neq \top$. Hence we have

$$I(\Psi) = \top$$
 or $I(\Psi') = \top$.

Now, if $J(\Psi \lor \Psi') = f$ we have that $J(\Psi) = J(\Psi') = f - a$ contradiction to the induction hypothesis. Thus, we must have $J(\Psi \lor \Psi') = t$. Thus,

$$J(\Psi) = t \text{ or } J(\Psi') = t$$

which contradicts the hypothesis that $I(\Psi) \sqsubseteq J(\Psi)$ and $I(\Psi') \sqsubseteq J(\Psi')$.

The other cases are similar.

The above proposition shows that the truth-value of a compound formula is limited by the truth-value of its subformulas: if I is less informed than, or equally informed as J w.r.t. atomic formulas, then I is less informed than, or equally informed as J w.r.t. to compound formulas.

Let us now turn back to our models. In order to prevent the overinterpretation of formulas by excessive usage of \top we wish to consider only those models from MOD(X) which are minimal according to \sqsubseteq . The following definition grasps exactly those models which do not overinterpret any sentence of X.

Definition 2.1 (Preferred three-valued model). An interpretation I is a *preferred three-valued model* for a set of sentence $X, I \models X$, if and only if $I \models X$ and there is no I' such that $I' \sqsubset I$ and I' is a K₃-model of X.

The following proposition states that the amount of information grows with the size of the database:

Proposition 2.3. Let $X \subset Y$ be a strict subset of Y and consider a preferred model I of X. There is a preferred model J of Y such that $I \sqsubseteq J$.

The set of all preferred models of X is denoted by $\mathsf{PMOD}(X)$. We can now give an entailment relation based on preferred three-valued models.

Definition 2.2 (\Vdash , Cn_3). Define the relation $\Vdash \subseteq 2^{\mathcal{L}} \times \mathcal{L}$, to be read as 'K₃entails', as follows: $X \Vdash \Phi$ if and only if every preferred K₃-model I of X is a K₃-model for Φ . The set $Cn_3(X)$ of three-valued consequences of X is defined as $Cn_3(X) =_{def} \{\Phi \mid X \Vdash \Phi\}.$

Consider again Example 2.1. I_1 is the only preferred K₃-model. Thus $X \parallel \vdash B$. The figure below visualises the relation \sqsubseteq among I_1, \ldots, I_5 .



Let us call K_3 the logic which results from taking the above truth-tables semantics and the preferential entailment relation.

Observation 2.1. The logic K_3 coincides with the logic LP(m) given in [Priest, 1991].

2.1.3 Examples

We shall now present some examples. The first example shows that Cn_3 is cautious.

Example 2.2. Let $X = \{A, \neg A, A \rightarrow B, C, C \rightarrow D\}$. This set of formula has two preferred K₃-models.

$$I_{1}(A) = \top | I_{2}(A) = \top I_{1}(B) = t | I_{2}(B) = f I_{1}(C) = t | I_{2}(C) = t I_{1}(D) = t | I_{2}(D) = t$$

We have $A, D, C \in Cn_3(X)$ but $B \notin Cn_3(X)$

The above example shows that due to the paraconsistency of A the formula B is not entailed by X and hence, Cn_3 is cautious. In this example it is easy to identify the formula which caused the inconsistency. The following example illustrates that a set of formulas can be inconsistent (i.e. it has no two-valued model) but that we cannot identify a unique paraconsistent formula.

Example 2.3. Let $X = \{A, A \to B, B \to C, C \to \neg A\}$. The following interpretations are preferred models for X.

$$\begin{array}{c|c} I_1(A) = \top & |I_2(A) = \top & |I_3(A) = \top & |I_4(A) = t \\ I_1(B) = t & |I_2(B) = f & |I_3(B) = f & |I_4(B) = \top & |I_5(A) = t \\ I_1(C) = t & |I_2(C) = t & |I_3(C) = f & |I_4(C) = f & |I_5(C) = \top \\ \end{array}$$

We have that $A \in Cn_3(X)$ but $B, C, \neg A, \neg B, \neg C \notin Cn_3(X)$.

2.2 Properties of Cn_3

We shall now prove that the consequence operator Cn_3 does indeed have the desired properties conservativity, paraconsistency and reflexivity. The fact that Cn_3 is cautious has been demonstrated in Example 2.3. Beside these rather specific properties, we shall discuss the following questions, some of which are related to classical, mathematical properties:

- 1. Which basic properties, like inclusion, idempotency, cumulativity etc. does Cn_3 have?
- 2. Does every set $X \subseteq \mathcal{L}$ have a preferred model?
- 3. Are there any deduction theorems valid for K_3 ?
- 4. Is K_3 's consequence operator Cn_3 compact?

Some of the properties related to the above questions play an important role in the area of mathematical logic (inclusion, idempotency, compactness) or in the field of nonmonotonic logic (cumulativity). Other properties pave the way for a deeper understanding of K_3 's treatment of contradicting information (preferred model existence, deduction theorems).

The plan is as follows: I shall first discuss some basic closure properties and deduction theorems (Sec 2.2.1). Next, I shall show that every set X has a preferred model (Sec. 2.2.3). I shall continue with relating K_3 to systems of nonmonotonic cumulative inference (Section 2.2.4). In a subsequent step I shall investigate the question whether Cn_3 is compact (Sec. 2.3).

2.2.1 Basic Closure Properties

There are two properties which I consider to be fundamental for every logical operator: inclusion and idempotency. Inclusion means that you are able to extract at least what you put in. The lack of Inclusion might indicate some sort of revision or transformation. The second property, idempotency, simply says that our consequence operator returns all consequences.

Theorem 2.1 (Preclosure). Let $X \subseteq \mathcal{L}$, then

1.
$$X \subseteq \operatorname{Cn}_3(X)$$

2. $\operatorname{Cn}_3(X) = \operatorname{Cn}_3(\operatorname{Cn}_3(X))$

Proof. Follows immediately from the definition of Cn_3 (Definition 2.2).

J

Any operator which is inclusive and idempotent is called, in algebraical terms, a *preclosure operator*. A closure operator is a preclosure operator which is monotone.

A fundamental theorem of classical propositional logic is the so-called *replacement theorem*. It says that by replacing equivalent parts we obtain equivalent propositions.

There are various techniques of proving the replacement theorem. See for example [de Swart, 1993] for a graphical one, [Hilbert and Bernays, 1934] for a cryptic one or [Genesereth and Nilsson, 1987] for a lazy one. I would like to adopt the one given in [van Dalen, 1980]. First, we have to define what a replacement exactly is; given Φ we denote the replacement of A by Ψ in Φ as $\Phi[A/\Psi]$. We define $\Phi[A/\Psi]$ recursively. Let A be a variable, $\circ \in \{\lor, \land, \rightarrow, \leftrightarrow\}$

$$\begin{split} \Phi[A/\Psi] &=_{\scriptscriptstyle def} & \begin{cases} \varPhi & \text{if } \varPhi \text{ is atomic and } A \neq \varPhi \\ \Psi & \text{if } \varPhi = A \end{cases} \\ (\varPhi_1 \circ \varPhi_2)[A/\Psi] &=_{\scriptscriptstyle def} & \varPhi_1[A/\Psi] \circ \varPhi_2[A/\Psi] \\ (\neg \varPhi)[A/\Psi] &=_{\scriptscriptstyle def} & \neg(\varPhi[A/\Psi]) \end{split}$$

Theorem 2.2 (Replacement Theorem). If $\Psi_1 \leftrightarrow \Psi_2$ is a K₃-tautology then $\Phi[A/\Psi_1] \leftrightarrow \Phi[A/\Psi_2]$ is a K₃-tautology.

Proof. By structural induction on Φ . Let $d(\Phi) = 0$. We have two cases. If $\Phi \neq A$ we have $\Phi = \Phi[A/\Psi_1] = \Phi[A/\Psi_2]$, hence $\models \Phi \leftrightarrow \Phi$ and we're done. In the case where $\Phi = A$ we have to show that $\models \Psi_1 \leftrightarrow \Psi_2$. But this is guaranteed by the prerequisite.

Assume that the proposition holds for all Φ with degree smaller than n. Let $d(\Phi) = n$. Again we have several cases:

 $\Phi = \Phi_1 \circ \Phi_2$ We have to show

$$\begin{split} &\models \varPhi[A/\Psi_1] \leftrightarrow \varPhi[A/\Psi_2] \\ \Leftrightarrow &\models (\varPhi_1 \circ \varPhi_2)[A/\Psi_1] \leftrightarrow (\varPhi_1 \circ \varPhi_2)[A/\Psi_2] \\ \Leftrightarrow &\models \varPhi_1[A/\Psi_1] \circ \varPhi_2[A/\Psi_1] \leftrightarrow \varPhi_1[A/\Psi_2] \circ \varPhi_2[A/\Psi_2] \end{split} \tag{0}$$

By the induction hypothesis we have

$$\begin{split} &\models \varPhi_1[A/\Psi_1] \leftrightarrow \varPhi_1[A/\Psi_2] \\ \Leftrightarrow \quad I(\varPhi_1[A/\Psi_1]) = \top \text{ or } I(\varPhi_1[A/\Psi_2]) = \top, \text{ or } I(\varPhi_1[A/\Psi_1]) = I(\varPhi_1[A/\Psi_2]), \\ \text{ for any interpretation } I \end{split}$$

The same holds for Φ_2 . It follows that for each subformula $\Phi_i[A/\Psi_j] = \top$, or $I(\Phi_1[A/\Psi_1]) = I(\Phi_1[A/\Psi_2])$ and $I(\Phi_2[A/\Psi_1]) = I(\Phi_2[A/\Psi_2])$. It is now easy to check that for every $\circ \in \{\lor, \land, \rightarrow, \leftrightarrow\}$ (**0**) holds. $\Phi = \neg \Phi'$ Similar to the above case.

$$= \neg \Psi$$
 Similar to the above case.

The following proposition shows that Cn_3 is a useful operator, which allows reasoning as in classical logic in the case where X is consistent. However, unlike classical logic Cn_3 does not collapse to triviality in the presence of contradicting information.

Proposition 2.4 (Conservativity). Let $X \subseteq \mathcal{L}$, Cn_{cl} the classical two-valued consequence operator. Then the following holds:

1. If X has a two-valued model, then $\operatorname{Cn}_{cl}(X) = \operatorname{Cn}_3(X)$.

2.
$$\operatorname{Cn}_3(\{\Phi, \neg \Phi\}) \neq \mathcal{L}.$$

3. Φ is a two-valued tautology if and only if Φ is a K₃-tautology.

Proof. The proof is easy.

- Ad 1: The set X has a two-valued model if and only if ever $I(A) \neq \top$, for every preferred three-valued model I of X and every variable A. Note that this means, that basically every two-valued model of X is a preferred (three-valued) model of X.
- Ad 2: By example. $B \notin Cn_3(\{\Phi, \neg \Phi\})$, for some variable $B \in \Sigma$.
- Ad 3: The set of K₃-tautologies coincides with the set of tautologies of Kleene's strong three-valued logic with \top as a designated truth-value, because both systems have the same truth-tables. By [Rescher, 1969], p. 341., we know that the tautologies of Kleene's strong three-valued logic with \top a designated value, coincides with the set of classical tautologies.

The above proposition shows that K_3 is indeed very close to classical logic. They share the same tautologies and in some cases they even have the same set of consequences. As a consequence, we shall now see that they also share several normal forms. Every formula Φ is semantically equivalent to some $\Phi_{\rm CNF}$ in conjunctive normal form. This will be very useful in carrying out proofs.

Proposition 2.5 (Conjunctive Normal Form). Let Φ be a formula. There is a semantically equivalent formula Φ_{CNF} such that

$$\Phi_{\rm CNF} = D_1 \wedge \ldots \wedge D_n$$

where each D_i is a disjunction of literals, i.e. $D_i = L_{i,1} \vee \ldots \vee L_{i,m_i}$.

Proof. By Proposition 2.4 Ψ is a classical tautology if and only if Ψ holds in all three-valued interpretations, i.e. Ψ is a K₃-tautology. Thus, $\Phi \leftrightarrow \Phi_{\rm CNF}$ is a K₃-tautology. Therefore, $\Phi_{\rm CNF}$ is valid in a three-valued interpretation I if and only if Φ is.

Sometimes we need sets of formulas to be given in a certain normal form which guarantees that each member of the set is consistent; the conjunction of all formulas of the set, however, must not necessarily be consistent.

Definition 2.3. A formula Φ is in *implicational normal form* (INF) if and only if it has the form

$$L_1 \wedge L_2 \wedge \ldots \wedge L_{n-1} \to L_n$$

where each L_i is a literal, i.e. an atomic or negated atomic formula. We call $L_1 \wedge \ldots \wedge L_{n-1}$ the *body* of an INF-formula and L_n the *head* of an INF-formula.

As the following proposition shows, INF is not very restrictive.

Proposition 2.6. Let Φ be a formula. Then there is a finite set X_{Φ} of INF-formulas such that for every interpretation I we have: $I \models \Phi$ if and only if $I \models X_{\Phi}$.

Proof. We know that Φ can be transformed into a semantically equivalent formula Φ_{CNF} in conjunctive normal form. Let $\Phi_{\text{CNF}} = D_1 \wedge \ldots \wedge D_n$ where each D_i is a disjunction of literals. Then Φ_{CNF} is semantically equivalent to the set $X = \{D_1, \ldots, D_n\}$. Clearly each D_i can be transformed into an INF-formula. \Box

The above normal forms imply that for every set X there is a semantically equivalent set X' such that X' is in *clausal normal form*. That is every formula of X' is a clause, i.e. a disjunction of literals. Since every clause has a classical two-valued model, we can split up any set of clauses into maximal consistent subsets.

Proposition 2.7. Let X be a set of clauses. There are consistent subsets Y_1, Y_2, \ldots of X such that

$$\int Y_i = X$$

In other words: every set of clauses can be cut up into maximal consistent subsets.

2.2.2 Deduction Theorems

The above properties are quite useful for carrying out proofs. Another useful tool which is very often used is the deduction theorem (i.e. $X \cup \{\Phi\} \Vdash \Psi$ iff $X \Vdash \Phi \to \Psi$). The deduction theorem shows that the notion of entailment is fully reflected in the object language.

Unfortunately the full version of the deduction theorem does not hold for K_3 . Despite this fact there are several approximations of the full version which do hold. To see why the full version fails, substitute Φ by $A \wedge (A \to B)$ and Ψ by B. Then, $\Phi \to \Psi$ is a tautology. Now, let $X = \{\neg A\}$. Then,

 $X \Vdash \Phi \to \Psi$ does not imply $X \cup \{\Phi\} \Vdash \Psi$.

The following weak version of the deduction theorem does hold for K_3 .

Proposition 2.8 (Weak Deduction Theorem). If $X \cup \{\Phi\} \Vdash \Psi$ then $X \Vdash \Phi \to \Psi$.

Proof. Let I be a preferred K_3 -model of X. We have to show that I is also a K_3 -model for $\Phi \to \Psi$. We shall distinguish two cases: 1) Φ is valid in I. It is easy to see that I is a preferred K_3 -model of $X \cup \{\Phi\}$. Since $X \cup \{\Phi\} \parallel \vdash \Psi$ we have that Ψ is valid in I; hence, $\Phi \to \Psi$ is valid in I. For the second case 2) assume that Φ is invalid in I. Thus, $\Phi \to \Psi$ is valid in I.

The full version of the deduction theorem fails, because the implication does not fulfil what [Rosser and Turquette, 1952] call 'the normalisation condition' (i.e. $\Phi \to \Psi$ takes a non-designated truth-value iff Φ takes a designated one and Ψ takes a non-designated one):

 $I \not\models \Phi \to \Psi \Leftrightarrow I \models \Phi, I \not\models \Psi$, for every I (Normalisation Condition, NC)

The set of all K₃-tautologies which satisfy the normalisation condition are denoted by Taut_{NC}. To see that K₃'s implication does not satisfy the Normalisation Condition, consider the following counter-example: $I(\Phi) = \top$ and $I(\Psi) = f$. Since $I \models \Phi$ and $I \not\models \Psi$ we should according to the Normalisation Condition have $I \not\models \Phi \rightarrow \Psi$. But, $I(\Phi \rightarrow \Psi) = \top$ and hence \rightarrow does not satisfy NC.

Remark 2.1. There are non-tautological formulas which satisfy NC.

As an example of such a formula, consider $\neg A \rightarrow A$.

The invalidity of the normalisation condition is the sole culprit for the invalidity of the deduction theorem:

Theorem 2.3 (Normalised Deduction Theorem). If $\Phi \to \Psi$ is a tautology which satisfies the normalisation condition then we have

$$X \Vdash \Phi \to \Psi \quad iff \quad X \cup \Phi \Vdash \Psi$$

Proof. By assumption, $\Phi \to \Psi$ is a tautology. Hence, for any X we have $X \Vdash \Phi \to \Psi$. Thus, we have only to show that $X \Vdash \Phi \to \Psi$ implies $X \cup \Phi \Vdash \Psi$. But since $\Phi \to \Psi$ satisfies the normalisation condition, we have that under any assignment I where Φ takes a designated value, Ψ takes a designated one. Hence, every preferred model of $X \cup \{\Phi\}$ is also a model for Ψ .

Another consequence of having an implication which does not satisfy the normalisation condition is that the rule of detachment (modus ponens) is not valid for K_3 . As we did for the deduction theorem, we can state under which circumstances we can conclude Ψ from $\Phi, \Phi \to \Psi$. We shall discuss this issue in Section 4.2 of Chapter 4, when we shall give a syntactical characterisation of the set $Cn_3(X)$.

Let me close this section with another version of the deduction theorem which will be useful.

Proposition 2.9 (Paraconsistent Deduction Property). Let $\{\varphi_1, \ldots, \varphi_n\}$ $\parallel \vdash \Phi \land \neg \Phi$. Then $\varphi_1 \land \ldots \land \varphi_n \rightarrow \Phi$ is a tautology which satisfies NC. *Proof.* By Proposition 2.8 $\varphi_1 \wedge \ldots \wedge \varphi_n \to \Phi$ is a tautology. It remains to show that it satisfies NC. Obviously, $\varphi_1 \wedge \ldots \wedge \varphi_n$ is inconsistent. In other words any K_3 -model for $\varphi_1 \wedge \ldots \wedge \varphi_n$ assigns the value \top to this formula. We show that Φ does also take the value \top in any of these models.

We assume the contrary. Let I be a model for $\varphi_1 \wedge \ldots \wedge \varphi_n$ and $I(\Phi) \neq \top$. Further, let $J \sqsubseteq I$ be a \sqsubseteq -minimal model for $\varphi_1 \wedge \ldots \wedge \varphi_n$. By Proposition 2.2 we have $J(\Phi) \sqsubseteq I(\Phi)$. Since J is a preferred model for $\varphi_1 \wedge \ldots \wedge \varphi_n$ we have that $\varphi_1 \wedge \ldots \wedge \varphi_n \not\models \Phi \wedge \neg \Phi$ – a contradiction.

Thus Φ takes the value \top in every model of $\varphi_1 \land \ldots \land \varphi_n$. Hence $\varphi_1 \land \ldots \land \varphi_n \rightarrow \Phi$ satisfies NC.

The Paraconsistent Deduction Property assumes that the set of formulas by which $\Phi \wedge \neg \Phi$ is entailed to be finite. Compared to the Normalised or Weak Deduction Theorem this might be a disadvantage. We shall see that in the background of the paraconsistent compactness property (Section 2.3) the limitation to finite antecedents in the above theorem is not crucial.

2.2.3 Existence of a Preferred Model

The main difference between our logic and Kleene's strong three-valued logic is the way the consequence relation is defined. Instead of taking all models into account we consider only those models which are preferred. The theory of preferred models makes only sense if the preference relation \square is well-founded, that is, only if we can guarantee that if X has a model, then it has a preferred model.

Theorem 2.4 (Preferred-Model Existence Theorem, [Priest 1991]). Every $X \subseteq \mathcal{L}$ has a preferred model.

Proof. Let $I_0 \models X$. We have to show that there is some $J \sqsubseteq I$ such that $J \models X$. Consider the chain

$$I_0 \sqsupseteq I_1 \sqsupset I_2 \sqsupset \dots$$

We shall show that this chain has a lower bound. It then follows by Zorn's Lemma that there is a minimal element J such that $J \sqsubset I_i$.

By Proposition 2.5 we can assume that X is in clausal normal form. Define

Inc $=_{def} \{A \mid I_i(A) = \top, \text{ for every } I_i \text{ of the above chain.}\}.$

In subsequent step, we extract those formulas from X which do not contain a literal which takes the value \top in every model of the above chain, i.e.

$$X' =_{def} \{ \Phi \in X \mid \operatorname{ATOM}(\Phi) \cap \operatorname{Inc} = \emptyset \}.$$

It follows that there is a model I_n of the above chain such that $I_n \models \Phi$ for every $\Phi \in X'$. Hence, every finite subset X_{fin} of X' has a classical two-valued model and by the classical compactness theorem, X' has a two-valued model H.

We can now define a K_3 -interpretation J with

$$J(A) =_{{}_{def}} \begin{cases} \top & \text{if } A \in \text{Inc} \\ H(A) & \text{otherwise} \end{cases}$$

J is a lower bound of the above chain. It remains to show that $J \models X$. Let $\Phi \in X$; if $\Phi \in X'$ then $H(\Phi) = t$. Moreover, for $A \in \operatorname{ATOM}(\Phi)$ we have $H(A) \neq \top$ and therefore, $J(A) \neq \top$. Hence, $J(\Phi) = t$ and $J \models \Phi$. If $\Phi \in X \setminus X'$ there is $A \in \operatorname{ATOM}(\Phi)$ such that $A \in \operatorname{Inc.}$ Thus, $J(A) = \top$. It follows that $J(\Phi) \in \{t, \top\}$. Hence, $J \models \Phi$.

Theorem 2.4 shows that the relation \sqsubset is indeed well-founded. The following lemmas will be very useful:

Lemma 2.1. Let $I \models X$ and $I \models \Phi$ then $I \models X \cup \{\Phi\}$.

Proof. Assume to the contrary that I is not a preferred model of $X \cup \{\Phi\}$. Hence, by Theorem 2.4 there is some $J \sqsubset I$ such that $J \models X \cup \{\Phi\}$. But then $J \models X$ and since $J \sqsubset I$, I cannot be preferred – a contradiction.

Lemma 2.2. Let $X \subseteq Y$ and $I \not \cong X$. Then, for every $J \not \boxtimes Y$ we have $J \not \sqsubset I$.

Proof. Assume that there is $J \models Y$ and $J \sqsubset I$. Since $X \subseteq Y$ we have $J \models X$ and I cannot be a preferred model of X.

Let me close this section with the remark that even though every set has a preferred model, there could be infinitely descending chains $I_0 \square I_1 \square I_2 \square \ldots$. For example, let Σ be infinite and $X = \emptyset$. Consider I_0 such that $I_0(A) = \top$, for every $A \in \Sigma$. Then there is an infinite chain of models of X such that $I_0 \square I_1 \square I_2 \square \ldots$.

2.2.4 Systems of Cumulative Reasoning

Nonmonotonic logics are, by name, characterised by a property which they do not have, namely monotonicity. Even though the lack of monotonicity might be the most discriminating attribute when being compared to many other logics, it is quite pessimistic to categorise logics solely by means of missing properties. It turned out this class of logics can in fact be described by a number of positive properties. The seminal paper in this field is [Kraus et al., 1990] who compiled several classes of properties which are called 'systems'. **The System C and P.** We shall relate K_3 to the system **C** of cumulative inference as well as to system **P** of preferential reasoning. Both systems are characterised by a set of Gentzen-style inference rules and axioms. **C** is the basic system. Let $|\sim$ be the meta-symbol separating the antecedent and the succedent of a sequent. In the case of nonmonotonic logic we wish to read $|\sim$ as *if* ..., *then normally* The system **C** is defined by the following axioms and rules:

System \mathbf{C}^7					
$\Phi \sim \Phi$	Reflexivity				
$\frac{\left \sim \Phi \leftrightarrow \Upsilon \Phi \mid \sim \Psi\right }{\Upsilon \mid \sim \Psi}$	Left logical equivalence				
$\frac{\left \sim \Phi \to \Upsilon \Psi \mid \sim \Phi}{\Psi \mid \sim \Upsilon}$	Right weakening				
$\left \frac{\Phi, \Upsilon \succ \Psi \Phi \mid \sim \Upsilon}{\Phi \mid \sim \Psi} \right $	Cutty				
$\boxed{\begin{array}{c c} \underline{\Phi} & & & & \\ \hline \underline{\Phi}, \Upsilon & & & \\ \hline \Psi \end{array}}$	Cumulativity				

Reflexivity, which is formulated as an axiom, is equivalent to the inclusion property. That is, an entailment relation is reflexive if and only if the corresponding consequence operator is inclusive. Theorem 2.1 states that Cn_3 is inclusive, hence $\parallel \vdash$ is reflexive.

$$\frac{= \Phi \leftrightarrow \Upsilon \quad \Phi \triangleright \Psi}{\Upsilon \triangleright \Psi}$$

⁷ There are some notational differences to Kraus, Lehman and Magidor's original version. For example, Left Logical Equivalence is noted as:

where \models denotes the classical satisfiability relation. However, by Proposition 2.4 we know that $\models \varPhi \leftrightarrow \Upsilon$ is equivalent to $\varnothing \Vdash \varPhi \leftrightarrow \Upsilon$ and thus, if we take \succ to mean \Vdash , we see that both versions are identical. The same argument applies to Right Weakening. Cumulativity is originally called 'Cautious Monotonicity'. The 'Cutty' is originally called Cut in [Kraus et al., 1990]. This rule is, however, much weaker than Gentzen's Cut rule, which will be extensively discussed in Chapter 4. To avoid confusion I use Cut for Gentzen's Cut and Cutty for Kraus, Lehmann and Magidor's Cut.

Theorem 2.5. The entailment relation $\parallel \vdash$ has the following properties:

- 1. Cumulativity
- 2. Left logical equivalence
- 3. Cutty

Proof.

Ad (1) Cumulativity: We have to show that

 $\{\Phi_1,\ldots,\Phi_n\} \Vdash \Upsilon \text{ and } \{\Phi_1,\ldots,\Phi_n\} \Vdash \Psi \text{ implies } \{\Phi_1,\ldots,\Phi_n,\Upsilon\} \Vdash \Psi.$

Let $X = \{\Phi_1, \ldots, \Phi_n\}$. Since $X \Vdash \Upsilon$ we know that every preferred model I of X is a model of Υ . Denote the set of all preferred models of $X \cup \{\Upsilon\}$ by \mathfrak{I} . By $\mathfrak{I} \subseteq \{I \mid I \models X\}$ we have that Ψ is valid in every $I \in \mathfrak{I}$. Hence, $X \cup \{\Upsilon\} \Vdash \Psi$.

Ad (2) Left logical equivalence: Let $\Phi \leftrightarrow \Upsilon$ be a K₃-tautology. It follows that Φ and Υ are semantically equivalent, i.e. $I \models \Phi$ if and only if $I \models \Upsilon$, for every I. We first show that this implies

$$I \triangleq \Upsilon \Leftrightarrow I \triangleq \Phi, \text{ for every } I \tag{(*)}$$

Assume to the contrary that (*) does not hold. Without loss of generality we restrict ourselves to the case that there is an I such that $I \models \Upsilon$ but $I \not\models \Phi$. Since $I \not\models \Phi$ we know that there must be an $I' \sqsubset I$ such that $I' \models \Phi$. Further, from $I \models \Upsilon$ we can conclude that $I' \not\models \Upsilon$. This contradicts the condition that Φ and Υ are semantically equivalent.

By (*) we can conclude that

$$I \models X \cup \{\Upsilon\} \Leftrightarrow I \models X \cup \{\Phi\}, \text{ for every } I.$$

That is, the sets $X \cup \{\Upsilon\}$ and $X \cup \{\Phi\}$ have the same preferred models. Thus, they have the same consequences Ψ .

Ad (3) Cutty: Let $X = \{\Phi_1, \ldots, \Phi_n\}$. We have to show that if $X \cup \{\Phi, \Upsilon\} \parallel \vdash \Psi$ and $X \cup \{\Phi\} \parallel \vdash \Upsilon$, then $X \cup \{\Phi\} \parallel \vdash \Psi$. From $X \cup \{\Phi\} \parallel \vdash \Upsilon$ we can conclude that,

$$I \models X \cup \{\Phi\} \Leftrightarrow I \models X \cup \{\Phi, \Upsilon\}, \text{ for every } I$$

That is, $X \cup \{\Phi\}$ and $X \cup \{\Phi, \Upsilon\}$ have the same preferred models. Thus, $X \cup \{\Phi\} \parallel \vdash \Psi$.
NB The above proof mentions only finite sets X. This is because Gentzen considered only finite sets. The proof, however, does not make use of the assumption that X is finite. It works without modification for infinite sets X.

Theorem 2.5 and Theorem 2.1 show that four of the five properties by which \mathbf{C} is characterised do also hold for K_3 . The last one, Right Weakening⁸, does not hold and thus K_3 and system \mathbf{C} are different.

To see that Right Weakening is invalid consider $\Phi = A \land (A \to B)$. Clearly, $\Phi \to B$ is a K₃-tautology, hence

$$\varnothing \Vdash \Phi \to B$$

Moreover, with $\Psi = A \land \neg A \land (A \to B)$ we have

 $\Psi \Vdash \Phi$

But we do not have $\Psi \parallel \vdash B$.

The failure of Right Weakening is an immediate consequence of K_3 's nonmonotonic behaviour and is strongly related to the issue of the normalisation condition which lead to the invalidity of the deduction theorem.

Remark 2.2. The logic K_3 is different from the system C.

How crucial is the failure of Right Weakening? According to Kraus, Lehmann and Magidor (henceforth denoted by KLM), System **C** is the rockbottom without which a system should not be considered a logical system. I do not share this point of view for two reasons. First, KLM consider their system to be the weakest possible. Gabbay (cf.[Gabbay, 1985]) on the other hand proposes only three conditions to be essential: reflexivity, cutty and cumulativity. Hence, there is a reasonable, weaker system than **C**. The second reason is that there is a version of Right Weakening, called Normalised Right Weakening, which is valid for K_3 :

The \dagger means: provided that $\Phi \to \Upsilon$ satisfies the Normalisation Condition.

⁸ Please do not confuse Right Weakening with Right Thinning (to be discussed in Chapter 4) which is a valid rule of inference for K_3 .

Proposition 2.10. Normalised Right Weakening is a valid rule of inference for K_3

Proof. From $\Psi \triangleright \Phi$ we can conclude that Φ takes designated truth-value in every preferred model of Ψ . Moreover, since $\Phi \to \Upsilon$ is a tautology which satisfies NC we know that Υ takes a designated value whenever Φ does. Hence, Υ takes a designated value in every preferred model of Ψ . Therefore, $\Psi \models \Upsilon$.

Replacing Right Weakening by Normalised Right Weakening yields a system which can be regarded as a normalised version of System C. Apparently, K_3 has all properties of this normalised version. As said before this system of cumulative inference is considered to be a very basic one. Beside the system C Kraus, Lehmann and Magidor defined system P which is nothing other than system C plus the following rule

$$\frac{\Phi \succ \Psi \quad \Upsilon \succ \Psi}{\Phi \lor \Upsilon \succ \Psi} \qquad (CA)$$

System \mathbf{P} of preferential reasoning is strictly stronger than \mathbf{C} and it assumes obviously the existence of disjunction in the language.

The following is easy to verify:

Observation 2.2. CA is a valid rule of inference in K_3 .

There are several important rules which are derivable in \mathbf{P} and which rely on the validity of Right Weakening. These derived rules are also valid in K_3 even though only Normalised Right Weakening holds in K_3 . This shows that Normalised Right Weakening is not a critical restriction at all.

Proposition 2.11 (Kraus, Lehmann and Magidor). Let $\Phi, \Psi, \Upsilon \in \mathcal{L}$.

1. In the presence of Reflexivity, Right Weakening and Left Logical Equivalence, the rule CA implies

$$\frac{\Phi \land \Psi \models \Upsilon}{\Phi \models \Psi \to \Upsilon} \qquad (S)$$

2. In the presence of Right Weakening, $\frac{\Phi \vdash \Psi \quad \Phi \vdash \Upsilon}{\Phi \vdash \Psi \land \Upsilon}$ (AND) and S imply Cutty.

3. In the presence of Reflexivity, Right Weakening and Left Logical Equivalence,

(a) CA implies

$$\frac{\Phi \wedge \neg \Psi \models \Upsilon \quad \Phi \wedge \Psi \models \Upsilon}{\Phi \models \Upsilon} \qquad (\mathsf{D})$$

(b) D implies CA in the presence of AND.

Proof. See [Kraus et al., 1990].

Proposition 2.12. The rules S and D are valid for K_3 .

Proof. The validity of S follows immediately from the weak deduction theorem (Proposition 2.8).

To see that D is valid:

From

$$\Phi \land \neg \Psi \Vdash \Upsilon \text{ and } \Phi \land \Psi \Vdash \Upsilon$$

we can get by CA to

$$(\varPhi \land \neg \varPsi) \lor (\varPhi \land \varPsi) \Vdash \varUpsilon$$

and thus

$$\varPhi \land (\varPsi \lor \neg \varPsi) \Vdash \varUpsilon$$

 and

 $\Phi \Vdash \Upsilon$.

However, not all derived rules of System C or P are valid in K_3 . For example, Modus Ponens in the consequent is invalid in K_3 :

$$\frac{\Upsilon \mid \sim \Phi \to \Psi \quad \Upsilon \mid \sim \Phi}{\Upsilon \mid \sim \Psi} \qquad \mathsf{MPC}$$

The proof of MPC requires AND and Right Weakening.

Anyway, the above shows that many properties considered by KLM are valid for K_3 , even though it should be emphasised that K_3 is certainly not the type of nonmonotonic logic which KLM had in mind. They draw their attention to the field of Nonmonotonic Reasoning (cf.[Brewka, 1991], [Marek and Truszczynski,1993]. Anyway, logics obtained from dropping monotonicity can indeed be treated systematically, whatever their original motivation might be (commonsense or paraconsistency).

Congruence. Another interesting and important property of cumulative systems is *congruence* with respect to a monotonic operator. Roughly speaking, the congruence property says that if the monotonic consequences of two sets are identical, so are the nonmonotonic consequences. For the nonmonotonic logic K_3 it is natural to choose the following monotonic operator:

 $Cn_{Bolz}(X) =_{def} \{ \Phi \mid I \models X \text{ implies } I \models \Phi \}$

It follows immediately that Cn_{Bolz} is a closure operator, i.e. embedding, monotone and idempotent. Note that $Cn_{\text{Bolz}}(X) = \{ \Phi \mid X \Vdash_{\text{Bolz}} \Phi \}$, where \Vdash_{Bolz} is the monotonic consequence based on Bolzano's notion of entailment. The following proposition is immediate:

Proposition 2.13. $\operatorname{Cn}_{\operatorname{Bolz}}(X) \subseteq \operatorname{Cn}_3(X)$.

Proof. Let $\Phi \in Cn_{Bolz}(X)$. Hence $MOD(X) \subseteq MOD(\{\Phi\})$. To see that $\Phi \in Cn_3(X)$ note that $PMOD(X) \subseteq MOD(X) \subseteq MOD(\{\Phi\})$.

The congruence property can be stated as:

 $Cn_{Bolz}(X) = Cn_{Bolz}(Y)$ implies $Cn_3(X) = Cn_3(Y)$.

This property plays an important role in theory of inference-frames (cf. [Dietrich, 1995]). I shall explain later why.

Proposition 2.14 (Congruence). Let $\operatorname{Cn}_{\operatorname{Bolz}}$ be the monotonic, three-valued consequence operator (as defined above). Then for any set X, Y we have $\operatorname{Cn}_{\operatorname{Bolz}}(X) = \operatorname{Cn}_{\operatorname{Bolz}}(Y)$ implies $\operatorname{Cn}_3(X) = \operatorname{Cn}_3(Y)$.

Proof. By Proposition 2.13 and Theorem 2.1 we know that for any set Z we have $Z \subseteq Cn_{Bolz}(Z) \subseteq Cn_3(Z)$. Since Cn_3 is cumulative we have $Cn_3(Z) \subseteq Cn_3(Cn_{Bolz}(Z) \cup Z)$ and since Cn_{Bolz} is embedding we have

$$Cn_3(Z) \subseteq Cn_3(Cn_{Bolz}(Z)).$$

We show: $Cn_3(Cn_{\text{Bolz}}(Z)) \subseteq Cn_3(Z)$. Assume to the contrary that there is $\Phi \in Cn_3(Cn_{\text{Bolz}}(Z))$ such that $\Phi \notin Cn_3(Z)$. Hence, there must be an I such that $I \models Z$ and $I \not\models \Phi$. However, $I \models Cn_{\text{Bolz}}(Z)$ and by Lemma 2.1 we know that $I \models Z \cup Cn_{\text{Bolz}}(Z)$, hence $I \models Cn_{\text{Bolz}}(Z)$. But then we have $\Phi \notin Cn_3(Cn_{\text{Bolz}}(Z))$ – a contradiction.

It follows that for all sets Z we have

$$Cn_3(Z) = Cn_3(Cn_{\text{Bolz}}(Z)). \tag{(\bigstar)}$$

Since $Cn_{Bolz}(X) = Cn_{Bolz}(Y)$ we have $Cn_3(Cn_{Bolz}(X)) = Cn_3(Cn_{Bolz}(Y))$. It follows from \star that $Cn_3(X) = Cn_3(Y)$.

Remark 2.3. The classical consequence operator Cn_{cl} and Cn_3 are not congruent.

Proposition 2.15. Cn_{Bolz} is a deductive basis of Cn_3 , i.e.

- 1. for all X we have $\operatorname{Cn}_{\operatorname{Bolz}}(X) \subseteq \operatorname{Cn}_3(X)$
- 2. for all X we have $\operatorname{Cn}_{\operatorname{Bolz}}(\mathcal{R}(X)) = \operatorname{Cn}_3(X)$
- 3. for all X, Y we have $\operatorname{Cn}_{\operatorname{Bolz}}(X) = \operatorname{Cn}_{\operatorname{Bolz}}(Y)$ implies $\operatorname{Cn}_3(X) = \operatorname{Cn}_3(Y)$.

Proof. Part 1 is by Proposition 2.13. For Part 2, note that since Cn_{Bolz} is embedding, we have $Cn_3(X) \subseteq Cn_{\text{Bolz}}(\mathcal{R}(X))$. To see that $Cn_{\text{Bolz}}(\mathcal{R}(X)) \subseteq Cn_3(X)$, assume to the contrary that there is some $\Phi \in Cn_{\text{Bolz}}(\mathcal{R}(X))$ such that $\Phi \notin Cn_3(X)$. By Proposition 2.13 we have

$$Cn_{Bolz}(Y) \subseteq Cn_{\ell}Y), \text{ for all } Y.$$

Hence, $\Phi \in Cn_3(Cn_3(X))$ and $\Phi \notin Cn_3(X)$ which contradicts the fact that Cn_3 is idempotent.

Part 3 is by Proposition 2.14.

Let me now explain why congruence is so important. Consider a monotonic operator C such that there is no monotonic operator C' such that $C(X) \subset C'(X) \subseteq Cn_3(X)$. That is, C can be seen as the largest monotonic operator beyond Cn_3 . It is known that such a largest C must be congruent w.r.t. Cn_3 . An interesting question is: which is the largest monotonic operator beyond Cn_3 ? The fact that Cn_3 is cumulative guarantees the existence of such a largest operator. Moreover, the fact that Cn_{Bolz} is congruent w.r.t. Cn_3 makes Cn_{Bolz} a possible candidate. However, it is an open question whether Cn_{Bolz} is in fact the largest monotonic operator beyond Cn_3 . **Monotonic Behaviour.** The discussion of the systems **C** and **P** shows that – no matter for what problem our nonmonotonic logic might have been designed for – there are core properties. Among these core properties cumulativity plays a special role because it tells us something about the monotonic behaviour of a nonmonotonic logic: if we add to X a formula Φ which is (nonmonotonically) entailed by X, then $Cn(X) \subseteq Cn(X \cup \{\Phi\})$, for some (nonmonotonic) operator Cn.

We shall now give an additional condition to characterise Cn_3 's monotonic aspects. It basically says that adding Φ to X does not cause any nonmonotonic effects as long as Φ does not cause any new inconsistency.

Proposition 2.16. Let X be a set of formulas, Φ a formula such that Φ has a two-valued model. If Φ does not take the value \top in any preferred model of $\operatorname{Cn}_3(X \cup \{\Phi\})$, then $\operatorname{Cn}_3(X) \subseteq \operatorname{Cn}_3(X \cup \{\Phi\})$.

Proof. It suffices to show that

$$\{I \mid I \models X\} \supseteq \{I \mid I \models X \cup \{\Phi\}\}$$
^(*)

We show that there is no $A \in \Sigma$ such that $J(A) \in \{t, f\}$ and $I(A) = \top$ for any $J \models X, I \models X \cup \{\Phi\}$. Since every arbitrary model of $X \cup \{\Phi\}$ is also a model of X we get then (*).

Suppose such an A exists. Without loss of generality assume that A is unique and that Φ is in CNF, i.e. $\Phi = D_1 \wedge \ldots \wedge D_n$. Since Φ takes the value t in every preferred model I of $X \cup \{\Phi\}$ we have that for every I there is $L_{i,0}$ such that $I(L_{i,0}) = t, L_{i,0} \in D_i$.

We can change A's truth-value from \top to either t or f and the resulting interpretation is still a model of $X \cup \{\Phi\}$: Let $J \in MOD(X)$ such that

$$J(B) = \begin{cases} I(B) & \text{, if } B \neq A \text{ for every } B \in \Sigma \\ \in \{t, f\} & \text{, if } B = A \end{cases}$$

The existence from such a J follows from $MOD(X) \supseteq MOD(X \cup \{\Phi\})$. Clearly, $J \sqsubset I$ and $J(L_{i,0}) = I(L_{i,0})$, hence $J(\Phi) = t$ and thus $J \models X \cup \{\Phi\}$. Therefore I cannot be a preferred model of $X \cup \{\Phi\}$ – a contradiction.

Proposition 2.16 says that the set of theorems grows, when we add information which is consistent w.r.t. X (i.e. which does not force us to assign additional sentences the value \top). Or, the other way round: the set of theorems might decrease if we add information which produces a new inconsistency.

2.3 Compactness

The next stop on our tour of algebraical properties is compactness. The compactness theorem for classical logic is, beside the Löwenheim-Skolem theorem, one of the first fundamental theorems in model theory. It allows us to reduce questions about infinite sets of formulas to questions about finite sets of formulas. Here is what it says (the term 'classical' means 'classical propositional logic' or 'classical first-order logic'):

- 1. If Φ is classically entailed by X then there is a finite subset of X which entails Φ .
- 2. A possibly infinite set X has a classical model if and only if every finite subset of X has a classical model.

The second item is also called 'finiteness' or 'compactness of satisfiability'.

In classical logic, the compactness of Cn_{cl} is proved by showing that if every finite subset X_{fin} of X has a model, then X has a model. Thus, compactness is proved by reduction to *finiteness*, or *compactness of satisfiability*. However, every set X has a (preferred) K₃-model (at least one which assigns every variable the value \top). Now the question is: is every formula entailed by a set X also entailed by a finite subset of X?

2.3.1 Compactness of Monotonic K₃

Let us first consider the monotonic version of K_3 which is built upon Bolzano's notion of entailment. Recall the definition of Cn_{Bolz} ,

$$Cn_{Bolz}(X) =_{def} \{ \Phi \mid I \models X \text{ implies } I \models \Phi \}.$$

Proposition 2.17 (Falsity Finiteness). Let X be a set of formulas, Φ a formula. If every finite subset X_i of X has a model in which Φ takes the value f then X has a model in which Φ takes the value f.

Proof. Let M(n) be the proposition 'every finite subset of X has a model in which the variables A_1, \ldots, A_n take the value $I(A_1), \ldots, I(A_n)$ '.

Let A_1, \ldots, A_k be the set of all variables occurring in Φ and let $I(A_1), \ldots, I(A_k)$ be an assignment such that $I(\Phi) = f$. By the prerequisite we have M(k). Now suppose, M(n), n > k holds. Consider A_{n+1} and let $I(A_{n+1}) = \top$. We have to show M(n+1).

Suppose there is a finite set X_{fin} which has no model in which the variables A_1, \ldots, A_{n+1} take the value $I(A_1), \ldots, I(A_n), \top$. Because of M(n) and the fact

that every set X has a K₃-model, we know that there must be a model of X_{fin} in which A_1, \ldots, A_{n+1} take the value $I(A_1), \ldots, I(A_n), t$ or $I(A_1), \ldots, I(A_n), f$. But then there must also a model in which A_1, \ldots, A_{n+1} take the values $I(A_1), \ldots, I(A_n), \top$.

Theorem 2.6 (Compactness of Monotonic K₃). Cn_{Bolz} is compact.

Proof. Suppose there is some Φ such that $\Phi \in Cn_{\text{Bolz}}(X)$ but for every finite $X_{\text{fin}} \subseteq X$ we have $\Phi \notin Cn_{\text{Bolz}}(X)$. Then every X_{fin} has a model I such that $I(\Phi) = f$. By Proposition 2.17 we have that X has a model in which Φ takes the value f. Hence, $\Phi \notin Cn_{\text{Bolz}}(X)$ – a contradiction.

Remark 2.4. Note, that it is possible to strengthen Falsity Finiteness as follows: If every finite $X_{\text{fin}} \subseteq X$ has a *preferred model* in which Φ takes the value f, then X has a model in which Φ takes the value f (*Restricted Preferred Falsity Finiteness*).

We shall see in the next section that it is not possible to guarantee: if every finite $X_{\text{fin}} \subseteq X$ has a preferred model in which Φ takes the value f, then X has a preferred model in which X takes the value f.

2.3.2 Paraconsistent Compactness

Even though every set X has a K₃-model, it is not the case that if $X \parallel \vdash \Phi$ then $X_{\text{fin}} \parallel \vdash \Phi$, for some finite $X_{\text{fin}} \subseteq X$. As a counter-example consider

$$\Phi_n = \bigwedge_{i=0}^n (A_i \wedge \neg A_{i+1}) \wedge (\neg A_0 \vee A_{n+1}) \wedge (\neg A_0 \vee B)$$

Now let

$$X = \{ \Phi_n \mid n \le \omega \}$$

The infinite set X has a unique preferred model I with

$$I(A_0) = t$$

$$I(A_i) = \top \text{ for all } 0 < i \le \omega$$

$$I(B) = t$$

Hence, we have $X \parallel \vdash B$. However, we have that $X_{\text{fin}} \not \Vdash B$ for every finite $X_{\text{fin}} \subset X$. To see this, consider an arbitrary X_{fin} and let $N = max\{i \mid \Phi_i \in X_{\text{fin}}\}$. Then X_{fin} has the following preferred models:

$$\begin{array}{c} I_1(A_0) = t \\ I_1(A_i) = \top & 0 < i \le N \\ I_1(A_{N+1} = \top & I_2(A_i) = \top & 0 < i \le N \\ I_2(A_i) = \top & 0 < i \le N \\ I_2(A_{N+1}) = t \\ I_2(B) = f & I_3(A_i) = \top & 0 < i \le N \\ I_3(A_{N+1}) = t \\ I_3(B) = t \end{array}$$

Since $I_2(B) = f$ we have $X_{\text{fin}} \not\Vdash B$ for every finite $X_{\text{fin}} \subset X$.

Theorem 2.7. Cn_3 is not compact.

There is, however, a weak version of the compactness theorem: for any formula Φ we have that if $X \parallel \Phi \land \neg \Phi$ then there is a finite $X_{\text{fin}} \subseteq X$ such that $X_{\text{fin}} \parallel \Phi \land \neg \Phi$. In order to prove this we show a variant of finiteness called consistency-finiteness: if for every finite $X_{\text{fin}} \subseteq X$ there is a preferred model $I_{X_{\text{fin}}}$ such that $I_{X_{\text{fin}}}(\Phi) \in \{t, f\}$, then there is a preferred model I of X such that $I(\Phi) \in \{t, f\}$.

Lemma 2.3. Let X be a set of formulas, Φ , a formula. If for every finite $X_{\text{fin}} \subseteq X$ there is a preferred model I such that $I(\Phi) \in \{t, f\}$, then there is a preferred model J of X such that $J(\Phi) \in \{t, f\}$.

Proof. By König's Lemma there is a model H of X such that $H(\Phi) \in \{t, f\}$. By the preferred model existence theorem (Theorem 2.4) we know that there is a preferred model J of X such that $J \sqsubseteq H$. It follows from Proposition 2.2 that $J(\Phi) \in \{t, f\}$.

Theorem 2.8 (Paraconsistent Compactness). If $X \Vdash \Phi \land \neg \Phi$ then there is a finite $X_{\text{fin}} \subseteq X$ such that $X_{\text{fin}} \Vdash \Phi \land \neg \Phi$.

Proof. Suppose there is no such X_{fin} , i.e. for every X_{fin} there is a preferred model I such that $I(\Phi) \neq \top$. Hence $I(\Phi) \in \{t, f\}$. It follows from Lemma 2.3 that Φ takes a value from $\{t, f\}$ in some preferred model of X and thus $X \not\Vdash \Phi$ – a contradiction.

2.3.3 Weak Compactness

[Dietrich, 1995] introduced another version of compactness called weak compactness. Weak compactness requires the existence of an idempotent, monotonic operator Cn such that for every X we have $Cn(X) \subseteq Cn_3(X)$ (supradeductivity). Again let,

$$Cn_{Bolz}(X) =_{def} \{ \Phi \mid I \models X \text{ implies } I \models \Phi \}$$

We mentioned earlier that Cn_{Bolz} is a closure operator (i.e. embedding, idempotent and monotone). Moreover we have $Cn_{Bolz}(X) \subseteq Cn_3(X)$.

The operator Cn_3 is said to be *weak compact* if and only if

$$\Phi \in Cn_3(X) \Leftrightarrow$$
 ex. finite $X_{\text{fin}} \subset Cn_{\text{Bolz}}(X)$ such that $\Phi \in Cn_3(X_{\text{fin}})$.

It is known that there are logical systems (e.g. minimal reasoning in twovalued propositional logic) whose consequence operators satisfy weak compactness. On the other side, minimal reasoning in two-valued first-order logic does not satisfy weak compactness (cf. [Herre, 1995]). See [Dietrich, 1995] for a detailed discussion on weak compactness of nonmonotonic logics.

The following equivalence relation is taken from [Dietrich, 1995]: Let Σ_n be a finite subset of Σ . Define the following relation $\simeq_{\Sigma_n} \subseteq \Im \times \Im$ between three-valued interpretation functions:

$$I_1 \simeq_{\Sigma_n} I_2$$
 if and only if $I_1 / \Sigma_n = I_2 / \Sigma_n$

where I_i / Σ_n denotes the restriction of I_i to Σ_n .

The relation \simeq_{Σ_n} is an equivalence relation. Denote the corresponding equivalence classes by $[I]_{\Sigma_n}$. For every $\mathfrak{I}' \subseteq \mathfrak{I}$ define $\mathfrak{I}'_{\Sigma_n} =_{def} \{[I]_{\Sigma_n} \mid I \in \mathfrak{I}'\}.$

Lemma 2.4. Let X be a set of formulas, $\Sigma_n \subset \Sigma$ a finite set of variables. There is a formula $\Phi_{\Sigma_n}^X \in Cn_{Bolz}(X)$ such that

$$\bigcup \mathsf{MOD}(X)_{\Sigma_n} = \mathsf{MOD}(\{\varPhi_{\Sigma_n}^X\})$$

Proof. Since Σ_n is finite, there are only finitely many interpretation functions $I_1, \ldots, I_m : \Sigma_n \to \{t, f, \top\}$ such that $I_i \in [J]_{\Sigma_n}$, for some K₃-model J of X. Define

$$\varPhi_{\Sigma_n}^X =_{{}_{def}} \bigvee_{j=1}^m \bigwedge_{i=1}^n \alpha_i \quad \text{ where } \alpha_i =_{{}_{def}} \begin{cases} A_i & \text{ if } I_j(A_i) = t \\ \neg A_i & \text{ if } I_j(A_i) = f \\ A_i \land \neg A_i & \text{ if } I_j(A_i) = \top \end{cases}$$

Obviously, if $H \in \bigcup MOD(X)_{\Sigma_n}$ then $H \models \Phi^X_{\Sigma_n}$. For the converse, suppose that $H \models \Phi^X_{\Sigma_n}$. By the above construction $\Phi^X_{\Sigma_n}$ has the form $\Phi_1 \lor \ldots \lor \Phi_m$ where

each $\Phi_j \equiv \bigwedge_{i=1}^n \varphi_i$. Because of $H \models \Phi_{\Sigma_n}^X$, there must be Φ_i such that $H \models \Phi_i$. Hence,

$$H \models \varphi_1 \land \ldots \land \varphi_n.$$

Each φ_i has one of the following forms:

$$\varphi_{i} \equiv A_{i} \text{ where } A_{i} \in \Sigma_{n}$$

$$\varphi_{i} \equiv \neg A_{i} \text{ where } A_{i} \in \Sigma_{n}$$

$$\varphi_{i} \equiv A_{i} \land \neg A_{i} \text{ where } A_{i} \in \Sigma_{n}$$

If $\varphi_i \equiv A_i \land \neg A_i$ we know that the model I_j by which φ_i is obtained in the above construction assigns $I_j(A_i) = \top$. Since $H \models \varphi_i$ we conclude $H(A_i) = \top$. Now, if φ has either the form A_i or $\neg A_i$ we know that $I_j(A_i) = t$ or $I_j(A_i) = f$ respectively. There are several possibilities how H can behave: $H(A_i) = t$ $(H(A_i) = f$, resp.) or $H(A_i) = \top$. In the first case we are done because H coincides with I on A_i . In the second case it is easy to see that if there is some I_j such that $I_j(A_i) = t$ $(I_j(A_i) = f$, resp.) then there must be a K₃-model $J_k \in \bigcup \text{MOD}(X)_{\Sigma_n}$ such that

$$J_k(B) = \begin{cases} I_j(B) & \text{if } B \neq A_i, \quad B \in \Sigma_n \\ \top & \text{if } B = A_i, \quad B \in \Sigma_n \\ \top & \text{otherwise} \end{cases}$$

For every Ψ we have that if $I_j \models \Psi$ then $J_k \models \Psi$ (this can shown by a simple structural induction on Ψ).

It follows that there must be a model J in $\{I_1, \ldots, I_m\}$ such that $H \in [J]_{\Sigma_n}$ and hence, $H \in \bigcup MOD(X)_{\Sigma_n}$.

Theorem 2.9 (Restricted Weak Compactness). Let MOD(X) be finite. If $\Phi \in Cn_3(X)$ then there is some $X_{fin} \subseteq Cn_{Bolz}(X)$ such that $\Phi \in Cn_3(X_{fin})$.

Proof. Let Σ be a signature. We first show that there is some n such that for every $J \models \Phi_{\Sigma_n}^X$ there is some $I \models X$ with $J \simeq_{\Sigma_n} I$, where Σ_n is a finite subset of Σ with cardinality n.

Let $\Sigma_k \subset \Sigma$ be of cardinality $k, k \in I\!N$ such that Σ_k contains all atomic formulas appearing in Φ . By Lemma 2.4 we have for any preferred model I of X:

$$I/\Sigma_k \not \sqsubset J/\Sigma_k$$
, for all $J \models \Phi_{\Sigma_k}^X$.

Now suppose that there is some preferred model of $\Phi_{\Sigma_k}^X$ which cannot be extended to a preferred model of X. Then there must be a variable B, $\Sigma_{k+1} =_{def} \Sigma_k \cup \{B\}$ such that some preferred model K of $\Phi_{\Sigma_k}^X$ cannot be extended to a preferred model of $\Phi_{\Sigma_{k+1}}^X$, i.e.

$$K/\Sigma_k \neq H/\Sigma_k$$
, for every $H \models \Phi^X_{\Sigma_{k+1}}$.

In this case we exclude all models of X which coincide with K on Σ_k .

Next, we have to check whether there is some $H \models \Phi_{\Sigma_{k+1}}^X$ which cannot be extended to a preferred model of X. By extending Σ_{k+1} as in the above case we can rule out those preferred models H of $\Phi_{\Sigma_{k+1}}^X$ with

$$H/\Sigma_{k+1} \neq I/\Sigma_{k+1}.$$

Since MOD(X) is finite, we end up with some set Σ_n such that for every preferred model J of $\Phi_{\Sigma_n}^X$ there is some $I \models X$ such that

$$J \simeq_{\Sigma_n} I$$

Then, $\Phi \in Cn_3(\{\Phi_{\Sigma_n}^X\})$: for assume that there is a preferred model of J of $\Phi_{\Sigma_n}^X$ which does not satisfy Φ . Since Σ_n contains all variables appearing in Φ , there must a preferred model of X which does not satisfy Φ – a contradiction. \Box

This completes our discussion on Cn_3 's mathematical properties for a while. We have seen that the paraconsistent entailment operator Cn_3 retains many classical algebraical properties: it is a cumulative, preclosure operator, for which several weaker versions of compactness hold. This is a plus. But not every classical positive hallmark can be preserved: the deduction theorem is not valid for K_3 indicating that the concept of entailment has no proper equivalent in the object language. This can be seen as an immediate consequence of choosing the implication to be material; an intensional connective might fix this problem but, on the other hand, would take us too far away from classical logic and might render other difficulties.

2.4 Computational Complexity

How complex is reasoning in K_3 ? For classical propositional logic, the entailment problem is known to be coNP-complete (cf.[Garey and Johnson, 1979]). We shall see that for the question of computing a preferred model of some set X, it

depends on the type of reducibility notion whether K_3 is more difficult than classical logic. As for deciding the set of consequences, we can determine with help of a non-deterministic, polynomial time bounded Turing machine which has an NP oracle, whether $\Phi \notin Cn_3(X)$ holds.

The satisfiability problem (SAT) for classical propositional logic was the first problem which was shown to be NP-complete. We shall now investigate the computational complexity of deciding $\Phi \in Cn_3(X)$. Like in the classical case we try to find a preferred model of X in which Φ takes the value f. A preferred model minimises the set of literals taking the value \top . If we compare SAT with a problem like Hamiltonian Path, then finding a preferred model is comparable to the problem of finding a shortest Hamiltonian Path. Typically, these so-called optimisation problems are more complex.

Consider the following problem:

PROBLEM: *k*-Preferred SAT

Instance A set $X \subseteq \mathcal{L}, k \in \mathbb{N}$. **Question** Is there a K_3 -model for X in which at most k atomic formulas take the value \top ?

Proposition 2.18. k-Preferred SAT is NP-complete.

Proof. For membership in NP note that we can nondeterministically guess a K_3 interpretation I and then verify in polynomial time whether I is a K_3 -model of X and whether at most k variables take the value \top . To show hardness, it is easy to see that it has SAT (which is NP-complete) as an instance by setting k = 0.

We can now show how difficult it is to compute a preferred model of X.

Proposition 2.19. Computing a preferred model I of X can be done in LINTIME^{NP[O(log n)]}, where n is the number of atomic subformulas occurring in the formulas of X.

Proof. We can use k-Preferred SAT as an NP-oracle for testing whether X has a preferred model with at most k variables set to \top . By using binary search we can find a minimal k_{min} . This requires at most $\lceil \log n \rceil$ queries to the oracle. \Box

The following results give a lower bound for the complexity of the decision problem $\Phi \in Cn_3(X)$.

Proposition 2.20. Deciding $Cn_3(X)$ is coNP-hard, for finite X.

Proof. By reduction to 2-Valued-Tautology (2-VT). The problem of 2-VT is: does Φ belong to $Cn_{c1}(\emptyset)$. Clearly, this holds, if $\neg \Phi \notin SAT$. Since SAT is NPcomplete, 2-VT is coNP-complete. Obviously, $\Phi \in 2$ -VT $\Leftrightarrow \Phi \in Cn_3(\emptyset)$.

Proposition 2.21. Deciding $Cn_3(X)$ is NP-hard.

Proof. By reduction to SAT. Suppose we are given a SAT instance, i.e. a set $X = \{C_1, \ldots, C_n\}$ of clauses⁹. For each clause $C_i = L_1 \lor \ldots \lor L_n \in X$ we construct two clauses:

$$C_i^1 = L_1 \lor \ldots \lor L_n \lor A$$
$$C_i^2 = L_1 \lor \ldots \lor L_n \lor \neg A$$

where A is a variable not appearing in any clause of X. Define

$$Y = \{C_1^1, C_1^2, \dots, C_n^1, C_n^2\}.$$

Clearly, A takes the value \top in some preferred model of Y if and only if X has no two-valued model. Consider the set $Y' = Y \cup \{A \lor B, \neg A \lor B\}$, where B is again a fresh variable. We have $B \in Cn_3(Y')$ if and only if A does not take the value \top in any preferred model of Y. Hence, $B \in Cn_3(Y')$ if and only if X has a two-valued model. Moreover, Y' can be generated from X in polynomial time. \Box

We can tighten the above result a little bit. The complexity classes NP and coNP are contained in a class called DP which was introduced in [Papadimitriou and Yannakakis,1984] and is defined as follows:

Definition 2.4 (DP). A language L is in the class DP if and only if there are two languages $L_1 \in NP$ and $L_2 \in coNP$ such that $L = L_1 \cap L_2$.

Languages in DP can be thought of as being accepted by an oracle TM M which puts one query to an NP-oracle and another query to a coNP-oracle (cf. [Papadimitriou, 1994]).

The following problem is known to be DP-complete:

PROBLEM: Critical SAT

Instance A set $X \subseteq \mathcal{L}$ of clauses

⁹ A clause C_i is a set of literals. A clause is satisfied by an interpretation I if at least one literal occurring in C_i is satisfied.

Question Is X inconsistent and does the deletion of any clause of X yield a consistent set?

Proposition 2.22. Cn₃ is DP-hard.

Proof. By reduction to Critical SAT. Suppose we are given X. Let $\Phi = \bigwedge \varphi_i$, $\varphi_i \in X$ be the conjunction of all clauses appearing in X. We have that X is inconsistent if and only if $\Phi \land \neg \Phi \in Cn_3(X)$.

Now suppose that $\{A_1, \ldots, A_n\}$ is the set of all atomic formulas appearing in Φ . Construct in polynomial time a new set Z which is identical to $X \setminus \{C\}$, for some $C \in X$. Replace every occurrence of A_i in the formulas of X by C_i . In a subsequent step construct analogous to the proof of Proposition 2.21 a set Y' such that

$$B \in Cn_3(Y') \Leftrightarrow Z$$
 is consistent.

It follows that we have

$$(\varPhi \land \neg \varPhi) \land B \in Cn_3(X \cup Y')$$

if and only if X is a member of Critical SAT.

The next proposition gives an upper bound for the complexity of Cn_3 .

Proposition 2.23. We can use a polynomial time bounded nondeterministic TM with an NP oracle to decide $\Phi \in Cn_3(X)$. Moreover, at most |X| NP oracle calls must be made.

Proof. In a first step we guess an assignment I to all the variables appearing in $X \cup \{\Phi\}$. We can verify in polynomial time whether I satisfies X and whether $I(\Phi) = f$. We shall now describe how to verify with an NP oracle whether I is a preferred model of X.

Let $\operatorname{Inc}(I) =_{def} \{A \mid I(A) = \top \text{ and } A \text{ is atomic } \}$. If we can assure that there is no model J of X such that $\operatorname{Inc}(J) \subset \operatorname{Inc}(I)$, we can conclude that I is a preferred model. Consider the following

PROBLEM: Preferred Subset

Instance A set $X \subseteq \mathcal{L}$, a set $\Sigma' \subseteq \Sigma$. **Question** Is there a K₃-model *I* for *X* such that $Inc(I) = \Sigma'$? It is easy to see that Preferred Subset is NP-complete.

Suppose we have guessed I. Let $k = |\operatorname{Inc}(I)|$. There are k subsets of cardinality k - 1; denote them by $\Sigma_1, \ldots, \Sigma_k$. We call the Preferred Subset oracle k times with parameters X and Σ_i . If any of these oracle calls succeeds, we know that there is a model J of X such that $\operatorname{Inc}(J) \subset \operatorname{Inc}(I)$. Therefore, I cannot be preferred.

Hence, k oracle calls sufficient to verify that I is a preferred model of X. Clearly, our machine accepts if and only if $\Phi \notin Cn_3(X)$.

Note that if we use Turing-reducibility instead of m-reducibility then computing a preferred K₃-model is not more difficult than finding a two-valued model of X.

Summarising, we can say that reasoning in K_3 is very difficult: we need a nondeterministic machine M with an NP oracle to decide $\Phi \in Cn_3(X)$. At most a polynomial number of NP oracle calls are necessary to decide $Cn_3(X)$. We shall see later that this class is located within the second level of the polynomial time hierarchy (cf. Section 2.6.2). There are many other natural problems which are of similar complexity, i.e. which are also located in the second level of the polynomial time hierarchy(PH). The second level contains several Travelling Salesman optimisation problems like, for example, the Master Tour property. But also the following problem from boolean logic is contained in the second level: given a propositional formula Φ in Disjunctive Normal Form and an integer n. Is there a semantically equivalent formula Φ_{CNF} in CNF which has at most n clauses? As a last example consider the problem of determining whether a given Default Theory has an extension. This problem, Default SAT, is also located within the second level of PH (cf.[Papadimitriou and Yannakakis, 1992]).

2.5 Extracting Consistent Information

This section is a small excursion investigating an interesting property of Cn_3 which appears when we consider sets of clauses. In the field of Logic Programming one restricts the general question 'Is Φ entailed by X?' to 'Is the literal L entailed by P?' where P is a set of clauses, which is also very often called a program.

We shall show that if the literal L is entailed by the program P, then L is either entailed by a consistent subset P' of P (where 'consistent' means that no element of P' takes the value \top in a preferred model of P^{10}), or $\{L\} \in P$ (i.e. L is explicitly given).

Before going into details we show an important property of sets of clauses (or, programs) P. Here is what it basically says: let P be a *non-redundant* set of clauses (non-redundant means that no clause C of P is entailed by $P \setminus C$); further, let $C = \{L_1, \ldots, L_n\}$ be a clause of P. Now, if C takes the value \top in some preferred model of P, then there must be a preferred model I_j of P in which L_j is \top , for every $L_j \in C$.

In order to picture this, note that every clause C has a two-valued model. This is because C is nothing but a disjunction of literals. Since only conjunctions might cause contradictions, P must contain at least two clauses in order to become contradictory. Consider the following program:

$$C_1: L_1$$

$$C_2: L_2$$

$$\vdots$$

$$C_n: L_n$$

$$C_{n+1}: \neg L_1 \lor \neg L_2 \lor \ldots \lor \neg L_n$$

Clearly, C_{n+1} is inconsistent with $\{C_1, \ldots, C_n\}$. Since we are unable to fix a single literal which causes the inconsistency, we have several alternatives resulting in n different preferred models for P. Hence there is a preferred model I_1 such that $I_1(L_1) = \top$, another one for which we have $I_2(L_2) = \top$ and so on.

Let us put this observation in more formal style:

Lemma 2.5. Let P be a non-redundant set of clauses, $C \in P$ a clause with $I(L_n) = \top$ for some $L_n \in C$, $I \models P$. Then, for every $L_i \in C$ there is some preferred model I' of P such that $I'(L_i) = \top$.

Proof. Let $C = \{L_1, \ldots, L_n\}$ and $I(L_n) = \top$ for some preferred model I of P. Since I is a preferred model and P is nonredundant, we know that there is a minimal inconsistent subset $P' \subseteq P$ such that $C \in P'$.

Assume to the contrary that there is $L_1 \in C$ which does not receive the value \top in any preferred model of P. Hence, the set

$$P'' := P' \setminus \{C\} \cup L_1$$

¹⁰ In Chapter 4 we shall use the notion of strongly relatively consistency to denote that no element of P' takes the value \top in any preferred model of P. Since no confusion can arise at the moment, we wish to use the simpler term 'consistent', though.

has a two-valued model. But $L_1 \parallel \vdash C$ and since P'' is consistent there must be a two-valued model of $P'' \cup C$. From $P' \subseteq P'' \cup C$ we conclude that P' must have a two-valued model – a contradiction.

Now let us come back to our original goal to show that if $L \in Cn_3(P)$ then L is entailed by a consistent subset (in the above sense) of P, or $\{L\} \in P$.

The basic observation is that if we are given a set P of clauses then we can transform P to a set P^T such that

$$L \in Cn_3(P)$$
 implies $L \in Cn_3(P^T)$ or $L \in P$. (*)

 P^T is obtained from P by deleting each clause C_i which contains a literal L_j such that $I(L_j) = \top$, for some preferred model I of P. For example, if P consists of the following clauses

$$\begin{array}{rcl} C_0: & D \\ C_1: & A \\ C_2: & \neg A \\ C_3: & A \lor B \lor C \end{array}$$

then we delete C_1, C_2, C_3 . The following theorem states that deletion does not affect the entailment of literals, i.e. it proves that (*) holds.

Theorem 2.10. Let P be a set of clauses, $P^T \subseteq P$ obtained from P as noted above. Then, $L \in Cn_3(P)$ implies $L \in Cn_3(P^T)$ or $L \in P$.

Proof. Assume to the contrary $L \notin P$ and that for every $P' \subseteq P$ with $P' \Vdash L$ there is some clause which contains a literal taking the value \top in some preferred model of P. Without loss of generality suppose that P' is nonredundant, i.e. for every $C \in P'$ we have $P \setminus \{C\} \not\Vdash C$. From the set of all nonredundant P' select a minimal set P'.

Let $C = \{L_1, \ldots, L_n, L\} \in P'$; since $L \notin P$ we have $n \ge 1$. From Lemma 2.5 and the minimality of P' we can conclude that for every $L_i \in C$ there is some preferred model of P in which L_i takes the value \top . We shall show that there is a preferred model of P in which L takes the value f which contradicts the assumption that $P \parallel \vdash L$.

To see that there is a preferred model of P which satisfies $\neg L$ but not L we prove the following

Lemma 2.6. If $C = \{L_1, \ldots, L_n, L\}$ is nonredundant w.r.t. P and there is some preferred model I_1 of P such that $I_1 \models L_1 \lor \ldots \lor L_n$ then there is some $I_2 \models P$ with $I_2 \not\models L$.

Proof. For sake of clarity let n = 1. Assume to the contrary that there is no such I_2 , i.e. for every preferred model I of P we have

$$I \models L_1 \Rightarrow I \models L \tag{0}$$

Since every preferred model of P satisfies C we have additionally:

$$I \not\models L_1 \Rightarrow I \mid\equiv L \tag{2}$$

Since C is nonredundant we have that $P \not\Vdash L_1$ and hence, there a preferred model of P which satisfies L_1 and another one which does not satisfy L_1 . From this and **0** and **0** it follows that L_2 is satisfied in every preferred model of $P \setminus \{C\}$. Hence, $P \setminus \{C\} \not\Vdash L_1 \lor L$ – a contradiction. \Box

By the above lemma there must be a preferred model of P in which L takes the value f. Hence L is not entailed by P and thus, L cannot be entailed by a set P' which contains a clause having a literal L which becomes \top under a preferred assignment.

Corollary 2.1. Let P be a set of clauses, $L \in Cn_3(P)$ a literal. There is some $P' \subseteq P$ such that

1. P' has a two-valued model. 2. $L \in Cn_3(P')$ (and hence, $L \in Cn_{cl}(P')$)

Proof. Note that the set P^T has a two-valued model. Thus, let $P' = P^T$ or $P' = \{L\}$.

Theorem 2.10 tells us that the transformation is complete. The following theorem ensures the soundness of the transformation.

Theorem 2.11. Let P be a set of clauses, P^T the transformation as described above. If $L \in Cn_3(P^T)$ then $L \in Cn_3(P)$.

Proof. Since no literal of any clause in P^T takes the value \top in some preferred model of P^T we have that no clause of P^T takes the value \top in such a model. It

follows easily from the construction of P^T that no formula of P^T takes the value \top in some preferred model of P. Hence, by Proposition 2.16 $Cn_3(P^T) \subseteq Cn_3(P)$.

The above theorem does not only tell us that the transformation is sound. It also shows that any classical consequence of a subset P' of P which is not involved in any contradiction (i.e. no literal in any clause of P' takes the value \top in some preferred model of P) is also a K₃-consequence of P'.

2.6 Philosophical Aspects

In the sequel I shall discuss properties which might be of interest from the perspective of Philosophical Logic. The problem of contradicting information is well-known in this field and there is a number of different approaches:

- Relevance Logics These logics are maybe the most famous paraconsistent logics. The original motivation was to develop and study an implicational connective to capture the notion of logical causal relationship. These logics have a property called *variable sharing property* which expresses the fact that if $\Phi \rightarrow \Psi$ is a tautology of a relevance logic, then Φ and Ψ share some sential variable, thus Φ and Ψ are 'connected' by this variable. We shall show that a similar result holds for K₃ provided that Ψ is *not* a tautology.
- Rescher-Based Approaches Rescher defined a notion of paraconsistent consequence which is based on reasoning about maximal consistent subsets. This idea has been taken up by Brewka to account for problems in the field of Default Logic.
- Belief Revision Even though not a bona fide paraconsistent approach, many revision operators base on some notion of consistent subsets. Belief revision is a competitive approach to paraconsistency because it tries to re-establish consistency.

This section is organised as follows: I shall first relate Rescher's approach to K_3 . To my impression, Rescher's approach was the most influential one for AI. Subsequently, I shall argue why retaining inconsistencies should be favoured to the approach of maintaining consistency, as done by belief revision. I shall close this section with some remarks on Relevance Logic.

2.6.1 Rescher-Based Approaches

In order to deal with an inconsistent set X, [Rescher, 1964] considers the set of all maximal consistent subsets of X. He defines two consequence operators based on Cn_{cl} : a sceptical one and a credulous one (also referred to as weak and strong consequence).

 $\operatorname{Cred}(X) =_{\operatorname{def}} \bigcup Cn_{\operatorname{cl}}(X'_i) \quad X'_i \text{ is a maximal consistent subset of } X$ $\operatorname{Scep}(X) =_{\operatorname{def}} \bigcap Cn_{\operatorname{cl}}(X'_i) \quad X'_i \text{ is a maximal consistent subset of } X$

 $\mathsf{Scep}(X)$ is always consistent, whereas $\mathsf{Cred}(X)$ in general is not.

Both consequence operators have a big disadvantage: they are not inclusive. Moreover Cred is not idempotent. To see that they are not inclusive, note that $A \notin \text{Scep}(\{A, \neg A\})$ and $A \land \neg A \notin \text{Cred}(\{A \land \neg A\})$. As for idempotency we have that $A \land B \in \text{Cred}(\{A, B\})$ but not in $\text{Cred}(\{C \land (C \rightarrow A), \neg C \land (\neg C \rightarrow B)\})$. Contrary to this we have

Proposition 2.24. Scep is idempotent.

Proof. Follows easily from the fact that the intersection of all sets closed under a given relation R is also closed under R.

There is a close relationship between the intersection of maximal consistent subsets of a set P of clauses and the set P^T as constructed in Section 2.5.

Proposition 2.25. Let P be a set of clauses. Then,

 $P^T \subseteq \bigcap P'_i \quad P'_i \text{ is a maximal consistent subset of } P.$

Proof. Assume that there is a clause $C \in P^T$ such that $C \notin \bigcap P'_i$. Hence, there is a maximal consistent subset P'_i such that $C \notin P'_i$. Thus, $P'_i \cup \{C\}$ has no two-valued model, but has a preferred three-valued model I such that $I(L) = \top$, for some $L \in C$. Since $P'_i \cup \{C\} \subseteq P$ we have by Proposition 2.3 that there is a preferred model J of X such that $J(L) = \top$. Hence, $C \notin P^T$ – a contradiction. \Box

The following proposition relates Scep to Cn_3 provided that the set X contains only literals.

Proposition 2.26. Let P be a set of clauses. Then $Scep(P) \subseteq Cn_3(P)$.

Proof. For any consistent X'_i we have by Proposition 2.4 $Cn_3(P'_i) = Cn_{cl}(P'_i)$. Since P'_i contains only literals this yields $\bigcup Cn_{cl}(P'_i) \subseteq Cn_3(P)$ and thus, by definition of Scep we have $Scep(P) \subseteq Cn_3(P)$.

The above proposition does not hold for arbitrary X as the following counterexample shows. Let $X = \{A, \neg A, A \rightarrow B, \neg A \rightarrow B\}$. We have $B \in \text{Scep}(X)$ but $B \notin Cn_3(X)$. The example shows a behaviour of Cn_3 which might not always be desired: X is in K₃ semantically equivalent to $\{A, \neg A, A \lor \neg A \rightarrow B\}$. Since $A \lor \neg A$ is a tautology one might want to conclude B, because no matter what A turns out to be in the end, true or false, B will be entailed.

This looks like a price we have to pay for having a straightforward entailment relation. In my opinion the benefits of our entailment relation make up for this little drawback. If, however, one is interested in removing this tiny deficiency, here is a solution: take a given set X and construct $Cn_3(X)$. If $\Phi \to \Psi$ is a member of $Cn_3(X)$ and Φ is a tautology, then add Ψ to $Cn_3(X)$; call the resulting set Y. Finally, apply Cn_3 to Y. This construction yields a new operator:

$$Cn_3^+ =_{{}_{def}} Cn_3(Cn_3(X) \cup \{\Psi \mid \Phi \to \Psi \in Cn_3(X) \text{ and } \Phi \text{ is a tautology. }\})$$

The operator Cn_3^+ and Scep are in fact comparable:

Proposition 2.27. Let X be a set of formulas in implicational normal form (INF). Then, $Scep(X) \subseteq Cn_3^+(X)$.

Proof. Let $\Phi \in \text{Scep}(X)$. We have to show that $\Phi \in Cn_3^+(X)$. Let $X'_1, X'_2...$ be maximal consistent subsets of X such that $X'_i \vdash \Phi$, for every $i \models i$ is the provability relation of classical logic. By compactness and the deduction theorem for classical logic we have, that for every X'_i there is some finite subset $\{\varphi_1, \ldots, \varphi_m\}$ such that $\{\varphi_1^i, \ldots, \varphi_m^i\} \vdash \Phi$. Let $\Psi_i = \varphi_1^i \land \ldots \land \varphi_m^i$. Hence, for every X'_i there is some Ψ_i such that $\Psi_i \to \Phi$ is a tautology.

We have for every n that

$$\Psi_1 \lor \Psi_2 \lor \ldots \lor \Psi_n \to \Phi$$

is a tautology. We consider now two cases.

Case 1 $\Psi_1 \equiv \Psi_2 \equiv \ldots \equiv \Psi_n$ for all n. Thus, for every maximal consistent subset X'_i it holds that Φ is entailed by the same set $Y \subseteq X'_i$. In other words, there is some $Y \subset \bigcap_{i=1}^n X'_i$ such that Y K₃-entails Φ . Clearly, no element from Y takes the value \top in some preferred model of X, hence $\Phi \in Cn_3(X \cup Y) = Cn_3(X)$.

Case 2 $\Phi_j \not\equiv \Phi_n$ for some n, j. In this case we have that there are two max. consistent sets X'_j , X'_n such that there is no $Y \subseteq X'_j \cap X'_n$ with Y K₃entails Φ . Hence Φ_j and Φ_n must be contradictory. But then the disjunction $\Phi_j \lor \Phi_n$ must be a tautology. Since $(\Phi_j \lor \Phi_n) \to \Phi \in Cn_3(X)$ we have that $\Phi \in Cn_3^+(X)$.

Let me stress again that although Cn_3^+ allows us to relate an extension of K_3 to Rescher's Scep, I do not consider Cn_3^+ as an alternative to Cn_3 because it is a strange mixture of a semantical and a syntactical consequence relation (remember that we added formulas to $Cn_3(X)$ only because of their syntactical structure). Anyway, I think that Cn_3^+ is a good device to compare our approach with Rescher's.

As for a comparison with Cred, Cn_3 could be more credulous than Cred. Consider for example the set $X = \{A, \neg A\}$. We have $A \land \neg A \in Cn_3(X)$ but $A \land \neg A \notin Cred(X)$ (which is admittedly somewhat obscure and shows a limit of Rescher's approach). Thus, we have in general that

$$Cn_3(X) \not\subseteq Cred(X).$$

To see that both operators are incomparable, it suffices to note that $B \notin Cn_3(\{A, \neg A, A \rightarrow B\})$ but $B \in Cred(\{A, \neg A, A \rightarrow B\})$, hence

$$Cred(X) \not\subseteq Cn_3(X).$$

There is, however, a way to relate Cn_3 and Cred on the level of clauses:

Proposition 2.28. Let P be a set of clauses, L a literal. Then the following holds, $L \in Cn_3(P)$ implies $L \in Cred(P)$.

Proof. Since L is a literal, it is consistent. Because P is a set of clauses we have by Corollary 2.1 that L is already K₃-entailed by some consistent subset P' of P (i.e. P' has a two-valued model). Hence, $L \in Cn_{cl}(P')$ and thus, $L \in Cred(P)$. \Box

The above discussion shows that paraconsistent reasoning by applying Cn_{cl} to some consistent subsets is severely limited. We do not only lose important aspects of any inference operation, namely inclusion and idempotency, but also inference rules like 'if X entails A and B, then X entails $A \wedge B$ ' are orphaned.

2.6.2 Belief Revision

In the beginning of this chapter we stated that, in general we have two possibilities when an inconsistency has been detected: we can live with the inconsistency or we can try to get rid of it. We argued in an informal manner that getting rid of the inconsistent information is a very complex process. This is because whenever it occurs that we have $\Phi, \neg \Phi \in Cn(X)$, for some Φ and some Cn, then it is quite difficult to find out what has caused the inconsistency.

The problem of belief revision (or knowledge-base revision) is to change X to X' in a minimal way such that X' no longer entails Φ . The minimality condition means that there is no set X'' such that $X' \subset X'' \subseteq X$ where X'' does not entail Φ .

Such revision operations can be used to perform a naive form of conditional reasoning by applying the so-called Ramsey test for conditionals:

Ramsey test: Accept a conditional sentence of the form 'If Φ , then Ψ ' in a belief state X if and only if the minimal change of X required to accept Φ also requires accepting Ψ .

The most interesting conditionals are those whose premise Φ is known to be false; these are called counterfactuals. Thus, counterfactuals are sentences of the form 'If Bizet was Italian, then he and Verdi were compatriots'. According to Ramsey, we change our knowledge-base minimally so that it does no longer entail that Bizet was not Italian. Then we can add the hypotheses that Bizet was Italian and see whether we can conclude that Bizet and Verdi were compatriots.

Let \circ be any revision operator. We say that a counterfactual 'If Φ , then Ψ ' (denoted by $\Phi > \Psi$) is true w.r.t. to a given theory X if and only if $X \circ \Phi$ (classically) entails Ψ . Probably the easiest revision operator is the following: Let $\hat{X} = Cn_{\rm cl}(\hat{X})$

$$\widehat{X} \oplus_{\mathsf{FMR}} \Phi =_{def} \left(\bigcap_{Y \in (\widehat{X} \downarrow \neg \Phi)} Y\right) \cup \Phi$$

where $X \downarrow \neg \Phi$ is the set of all maximal consistent subsets of X which do not classically entail Φ . Note that \widehat{X} is deductively closed. The relationship to Rescher's Scep is immediate. Thus, we can have that $\widehat{X} \not\in \widehat{X} \oplus_{\mathsf{FMR}} \Phi$, which corresponds to the failure of inclusion. Of course, failure of inclusion, or in other the loss of information is intended by revision operators to restore consistency.

The computational complexity of the corresponding operator FMR – full meet revision – has been analysed in [Nebel, 1992]. Recall the definition of the

polynomial time hierarchy (cf. [Stockmeyer, 1976]):

$$\Sigma_0^P = \Pi_0^P = \Delta_0^P = \mathsf{P}$$

and for all $k \ge 0$

$$\begin{split} \Delta_{k+1}^{P} &= \mathsf{P}^{\Sigma_{k}^{P}} \\ \Sigma_{k+1}^{P} &= \mathsf{N}\mathsf{P}^{\Sigma_{k}^{P}} \\ \Pi_{k+1}^{P} &= \mathrm{co}\text{-}\Sigma_{k+1}^{P} \end{split}$$

where $\mathsf{P}^{\Sigma_k^P}$ denotes the class of all languages accepted by a deterministic, polynomial time bounded TM M with a Σ_k^P -oracle.

Proposition 2.29 (Nebel, 1992). FMR $\in \Delta_2^P - (\Sigma_1^P \cup \Pi_1^P)$, provided that $\operatorname{coNP} \neq \operatorname{NP}$.

From the point of Turing reducibility¹¹ FMR and the problem of two-valued satisfiability (SAT) are equal, i.e. FMR \leq_T SAT and SAT \leq_T FMR. This has been observed by Nebel.

Instead of considering only deductively closed set (theories), Nebel proposes a so-called *base revision*, i.e. revisions on sets of formulas which are not necessarily deductively closed. The corresponding revision operator is called *simple base revision* (SBR):

$$X \oplus_{\mathsf{SBR}} \Phi =_{def} \left(\bigcap_{Y \in (X \downarrow \neg \Phi)} Cn_{\mathsf{cl}}(Y) \right) \cup \Phi$$

where $X \downarrow \neg \Phi$ is the set of all consistent subsets of X which do not entail Φ . Again, the basic idea is very similar to Scep. The complexity of the revision operation changes dramatically:

Proposition 2.30 (Nebel, 1992). SBR is Π_2^P -complete.

In other words, SBR is not Turing-reducible to SAT. A nondeterministic polynomial time bounded machine with NP oracle would, however, accept SBR. That is, the difference between SBR and FMR is similar to the difference between

¹¹ Recall that a set S_1 is Turing-reducible to S_2 , denoted by $S_1 \leq_T S_2$, if and only if there is a deterministic polynomial time oracle machine M such that $S_1 = L(M, S_2)$ (cf.[Balcazar et al., 1988], Chapter 4).

 P and $\mathsf{NP}.$ This shows that a relative weak revision operation like SBR is fairly complex.

If we want to know what follows from a knowledge-base revised by Φ , we have to compute $Cn_{cl}(X \oplus_{\mathsf{SBR}} \Phi)$ (or, $Cn_{cl}(X \oplus_{\mathsf{FMR}} \Phi)$). Surprisingly this does not add complexity as shown in [Eiter and Gottlob, 1992]. Eiter and Gottlob proved that deciding $Cn_{cl}(X \oplus_{\mathsf{SBR}} \Phi)$, i.e. deciding whether the counterfactual $\Phi >_{\mathsf{SBR}} \Phi$ holds is Π_2^{P} -complete¹². Eiter and Gottlob discussed many other revision operators. For most of them, the corresponding counterfactual problem is Π_2^{P} -complete.

Theorem 2.12 (Eiter and Gottlob). Deciding if a counterfactual $\Phi > \Psi$ is true w.r.t. to a set X is Π_2^P complete for $> \in \{>_G, >_{Win}, >_B, >_S, >_F\}$, and Π_2^P hardness holds even if X is atomic and $\Psi \in X$, where $>_G$ is the counterfactual based on the revision operation introduced in [Ginsberg, 1986], $>_{Win}$ the one by [Winslett, 1990], $>_B$ the one by [Borgida, 1985], $>_S$ the one by [Satho, 1988] and $>_F$ the one by [Forbus, 1989].

The above results show indeed that maintaining consistency is more complex than reasoning in the paraconsistent logic K₃. More precisely, reasoning in K₃ is coNP hard whereas most operators discussed in [Eiter and Gottlob, 1992] are Π_2^P -hard, i.e. one level higher in the polynomial time hierarchy. As for membership in a complexity class, most belief revision operators discussed by Eiter and Gottlob are also in Π_2^P , thus – taking the above into account – are Π_2^P -complete. But how difficult is reasoning in K₃ when being compared to belief revision?

By Proposition 2.23 we have immediately:

Proposition 2.31. Deciding $Cn_3(X)$ is in Π_2^P .

Thus, reasoning in our paraconsistent logic is not more difficult than belief revision. As for practical impact, many revision systems require user-interaction. Consider for example the following database: $X = \{A, A \rightarrow B, B \rightarrow C, C \rightarrow \neg A\}$. The set X does not have a two-valued model. Changing X minimally in order to obtain consistency leads to several alternatives. We can eliminate $A \rightarrow B$ or $B \rightarrow C$ or $C \rightarrow A$ or A to make X consistent. A minimal change would require that somebody tells us which formula should be removed.

All in all, we have just shown that living with the inconsistencies is from a computational point of view not more expensive than revising an inconsistent knowledge-base each time contradicting information has been entered. This is what we have suspected. Of course, it would be a little unfair to evaluate the

¹² The revision operator \oplus_{SBR} is called \oplus_G in [Eiter and Gottlob, 1992].

usefulness of belief revision only by means of computational complexity. Both approaches attack the same problem but yield different results. Even Belnap is about to flirt a little with the idea of supplying his reasoner with 'some strategy of giving up part of what it believes'. Belnap drops this idea immediately since '(he has) never heard of a practicable, reasonable, mechanizable strategy for revision of belief in the presence of contradiction', [Belnap, 1977]. He concludes that a *part* of the complete reasoner should be able to cope with contradicting information. Another part might be able to revise beliefs.

2.6.3 Relationship to Relevance logics

Relevance logics (cf. [Dunn, 1986]) were developed to avoid certain peculiarities of material implication. The basic concept is that of relevant deduction. A sentence Ψ is relevantly deducible from a set $\{\Phi_1, \ldots, \Phi_n\}$ of premises just in case all the Φ_i are actually used. This idea is very close to Ackermann's *Begründung einer strengen Implikation* as well as to C.I. Lewis' concept of a strict implication, where A strictly implies B if and only if it is not possible (in a modal sense) that A is true and B is false (note that this is identical to $\Box(A \to B)$, where \to is material implication). Thus, the concept of strict implication is an intensional (not truth-functional) concept.

The treatment of the implication as an Intensional Relation has also been discussed by L. Nelson. His ideas have then been further developed by Anderson and Belnap. In their seminal work Entailment, [Anderson and Belnap, 1975], they developed the propositional logics \mathbf{E} of entailment and \mathbf{R} of relevant implication. Unfortunately these two main systems are motivated from a syntactical perspective: Anderson and Belnap present \mathbf{E} and \mathbf{R} as a system of axioms, thus leaving the question of a semantics open. In fact, the semantics for these logics are quite complicated and there have been recent attempts by Arnon Avron (cf. [Avron, 1989], [Avron, 1990a], [Avron, 1990b]) to give a more elegant and intuitive semantics for those systems in order to make relevance logic more attractive for applications in Computing Science.

Despite these promising developments in the field of relevance logics, there is a drawback which knocks out all attempts to incorporate these logics faithfully into practical application: They are even in the propositional case undecidable. Precisely, Urquhart showed that the systems of relevant implication \mathbf{R} , entailment \mathbf{E} and of ticket entailment \mathbf{T} are undecidable, [Urquhart, 1984]. This makes them useless for an application in deductive databases.

I shall briefly explain the basic relationship between K_3 's aspects of relevance and the relevantist's aspects of relevance. In [Norman and Sylvan, 1989] we can find a 'working classification of logics'. According to this classification, all logics which reject Disjunctive Syllogism (see below) are called *pseudo relevant*.

$$\frac{\neg \Phi, \, \Phi \lor \Psi}{\Psi} \quad \text{Disjunctive Syllogism (DJ)}$$

Thus, from the relevance enterprise's point of view K_3 is at least pseudorelevant because a vacuously false statement (in a classical sense) does not justify everything. Please note that though $(A \wedge \neg A) \rightarrow B$ is a tautology in K_3 , it does – due to the failure of the general deduction theorem – not hold that $\{A, \neg A\} \Vdash_{\overline{K_3}} B$, for every B.

Myhill describes in [Norman and Sylvan, 1989], the relevance system \mathbf{E} to be 'designed to formalize the insights of philosophers who oppose to the notion that a true proposition is implied by everything and a false proposition implies everything'. Clearly, K_3 accounts for the second problem but not for the first: if $\Phi \in Cn_3(X)$ then $\Psi \to \Phi \in Cn_3(X)$, even though there is no real connection between Φ and Ψ . Anderson and Belnap get rid of this problem by rejecting the axiom scheme $\Phi \to (\Psi \to \Phi)$, which is called by the relevantists *Positive Paradox*.

In K₃ we do not only have that Positive Paradox is a valid axiom scheme, but $\{A\} \Vdash_{\overline{K_3}}(B \to A)$ for every *B*. Hence, contrary to the situation $(A \land \neg A) \to B$, we do not get rid of the irrelevance aspects of Positive Paradox. This is clearly a point where the hardcore relevantists would reject K₃ as a genuine relevance logic¹³.

Even though K_3 is not a genuine relevance logic, it has the variable sharing property which is very important for relevance logics. According to Dunn, the variable sharing condition expresses some commonality of meaning between the antecedent and the succedent. The variable sharing property is also called *weak* relevance.

Proposition 2.32 (Variable Sharing Property). Let Φ be non-tautological. If $\Phi \in Cn_3(X)$ then there is a formula $\Psi \in X$ such that Φ and Ψ share some sential variable.

Proof. Since Φ is non-tautological, there is a preferred assignment I such that $I(\Phi) = f$. Assume to the contrary that there is no $\Psi \in X$ sharing a variable

¹³ I would like to annotate that Anderson and Belnap, though they refuse Positive Paradox, admit that if A is true then it is 'safe to infer A from an arbitrary B, since we run no risk of uttering a falsehood in doing so' (Entailment 1).

with Φ . Then there is a preferred model J of X such that J(A) = I(A) for every atomic A appearing in X. Hence, $J(\Phi) = f$ and $\Phi \notin Cn_3(X)$ – a contradiction. \Box

Corollary 2.2. If $(\Phi \land \neg \Phi) \to \Psi$ is a tautology which satisfies the normalisation condition and Ψ is not a tautology, then Φ and Ψ share some sential variable.

Corollary 2.3. If $\Phi \to \Psi$ is a tautology which satisfies the normalisation condition and Ψ is not a tautology, then Φ and Ψ share some sential variable.

In addition to the variable sharing property, Theorem 2.10 gives an answer to the question of what *literals* L are entailed by X (provided that X is in INF): the literals L and $\neg L$ are entailed by X if and only if both have been asserted *explicitly*, i.e. $L \in X$ and $\neg L \in X$. If L is entailed by X and $L \notin X$, then L must be based on some consistent argument, i.e. there is a subset $X' \subseteq X$ such that no formula of X' takes the value \top in any preferred model of X and L follows from X'.

2.7 Discussion

2.7.1 Historical Remarks

We have presented a neo-classical approach to handle inconsistent information: take Kleene's strong three-valued logic and supply it with a Shoham-like preference relation. With Belnap's work in the background, there is nothing spectacular about this approach and hence, it is no wonder that similar preference relations were independently discovered by [Priest, 1991], [Kifer and Lozinskii, 1992] and [Weber and Bell, 1994].

In Kifer and Lozinskii's work, we find two implicational symbols: one to denote ontological implication and another one for epistemological implication. The latter corresponds to our notion of preferred entailment.

Whereas Kifer and Lozinskii are mainly concerned about an inconsistency handling via annotated logics programs, Priest's logic LP(m) (minimal logic of paradox) is indeed identical to our logic K_3 .

Quite recently D. Batens suggested a class of logics called *Adaptive Logics* [Batens, 1997]. The main aspects are:

- A given set X is interpreted as consistently as possible. This means that if X is consistent, then the consequences should be identical to the classical consequences.

- An inconsistency adaptive logic, which - according to Batens - cannot be monotonic is based on a monotonic paraconsistent logic.

 K_3 enjoys the first and the second property. However, K_3 is not considered by Batens to be an adaptive logic (personal communication). This is since $\{A, \neg A, A \lor \neg A \to B\}$ does not K_3 -entail B whereas B should follow from the above set by means of an adaptive logic (Batens, personal communication).

2.7.2 Open Problems

Let me shortly recall the problems mentioned this chapter which are still open.

- 1. Is there a complexity class C for which the decision problem $\Phi \in Cn_3(X)$ is complete?
- 2. Is there some restriction of \mathcal{L} such that deciding $\Phi \in Cn_3(X)$ becomes tractable?
- 3. Is Cn_3 weakly compact?
- 4. Which is the largest monotonic logic beyond Cn_3 ?

2.7.3 Conclusion

From a pure logical aspect, the Relevance Logics \mathbf{E} and \mathbf{R} might be the most desirable systems for paraconsistent reasoning. There are two arguments against their use in AI: First, they are not decidable and second, they are too far away from classical logic. As an alternative to these logics one could consider the paraconsistent logics of da Costa (not to be discussed here) which are perfectly monotonic and the Rescher-based approaches and their deviants as, for example, discussed in [Benferhat et al., 1995]. The latter lack many desired properties like Reflexivity and AND. Da Costa's logics on the other hand are not close enough to classical logic. Their negation operation, for example, does not obey the rule of double negation.

This is now where K_3 comes into play. K_3 has all the nice mathematical properties, is very close to classical logic and is also from a computational point of view not worse than belief revision.

Beside these formal aspects, K₃ reflects important principles of its rivals.

Belief Revision systems must satisfy the so-called postulate of Minimal Change; this has been claimed Gärdenfors. This implies that the changes made to an inconsistent theory in order to become consistent should be minimal. In K_3 Minimal Change is comparable to the cautiousness of Cn_3 .

In fact, as the analysis of the computational complexity shows: Belief Revision and reasoning in K_3 are *optimisation problems*. Belief revision maximises consistency and K_3 to minimises inconsistency.

- Relevance Logics. K_3 can be categorised as a pseudo relevant system with variable sharing property. Even though it is not a genuine relevance logic, it has some aspects of relevance; given the non-decidability result of relevance logics, K_3 seems to be a good compromise between the pure logical and the computational aspects.
- Rescher-Based Approaches. Although Rescher does not give us much background information on his premises, I think it is reasonable to assume that Rescher wants his conditional entailment to be equivalent to classical entailment if the set of premises is consistent. This hypothesis stems merely from the fact that the only problematic conditionals are those whose antecedent has a doubtful truth-status. The equivalence to classical logic in the consistent case is also an important feature of K_3 .

Summarising, we have shown that Cn_3 is a reasonable basis for paraconsistent reasoning. We shall see that it can be combined with a possible worlds approach to account for contradicting and unknown information.

CHAPTER 3

Paraconsistency: The First-Order Logic FOK₃

We shall now extend the ideas and notions of Chapter 2 to the first-order case. As the reader will see, this is merely a technical exercise. However, we shall obtain some new results and solutions to open problems mentioned by Priest in [Priest, 1991]. As already mentioned before, Priest gave a proof of the Preferred Model Existence Theorem. An open problem, questioned by Priest, is whether the first-order analogue of this theorem holds. We shall see that there are theories which do not necessarily have a preferred model. If, however, we restrict ourselves to universal theories, then we can show that every such theory has a preferred model. We further investigate whether Herbrand's theorem holds for FOK₃.

3.1 Terminology and Notations

First-order logic requires the introduction of many additional concepts. If we want to be precise, it is unavoidable to become technical. However, we assume the reader to be familiar with the basic concepts and ideas of first-order logic so that getting used to my terminology should not be too difficult. The notations to be presented in this section are very close to those used by [Thiele, 1996].

The first-step is to extend the language \mathcal{L} to a first-order language \mathcal{L}_1 . A first-order signature Σ is a tuple $\Sigma = (\mathsf{PS}, \mathsf{FS}, \alpha)$, where PS is a countable set of predicate symbols, FS a countable set of function symbols and α is function which assigns to each predicate or function symbol an arity. If Σ is a first-order signature, V a countable set of variables we can define the language $\mathcal{L}_1(\Sigma, V)$ over Σ and V as usual. We shall omit the reference to Σ or V when it is clear from the context or not important.

Let us now turn to some basic notations of first-order semantics. A Σ algebraic structure \mathfrak{A} is a tuple $\mathfrak{A} = [\Sigma, D, F, P]$, where D is a non-empty set (also called domain), F an injective mapping which assigns a *total* function to each symbol $f \in FS$; $P = \{P^+, P^-\}$ where P⁺ and P⁻ are mappings which assign to every *n*-ary predicate symbol a subset from D^n , i.e. we can think of P⁺ and P⁻ as sets which contain relations.

Instead of using the truth-values $\{t, f, \top\}$ we shall now deal with models and counter-models. This means that for each predicate symbol $P \in \mathsf{PS}$ we have two interpreting sets: $\mathsf{P}_{\mathsf{P}}^+$ and $\mathsf{P}_{\mathsf{P}}^-$. The first set contains all objects from the domain for which P holds; the latter contains all objects for which P does not hold. It is not required that $\mathsf{P}_{\mathsf{P}}^+$ and $\mathsf{P}_{\mathsf{P}}^-$ are complementary or even disjunct. We only require that $P_{\rm P}^+ \cup P_{\rm P}^- = D^n,$ i.e. there are no truth-value gaps: each tuple is in $P_{\rm P}^+$ or in $P_{\rm P}^-$ (or, maybe in both sets).

We shall now explain how terms and formulas are interpreted.

Definition 3.1 (State of a Variable, Interpretation of Terms).

1. σ is called \mathfrak{A} -state of the variables from V if and only if

$$\sigma: V \to \mathsf{D}$$

- 2. $\operatorname{EL}(T,\mathfrak{A},\sigma)$ denotes the element from D which corresponds to the term T: (a) $\operatorname{EL}(v, \mathfrak{A}, \sigma) =_{def} \sigma(v)$, if $v \in V$.

 - (b) $\operatorname{EL}(f,\mathfrak{A},\sigma) =_{def} \operatorname{F}_f$, if f is a 0-ary function symbol (i.e. a constant). (c) $\operatorname{EL}(f(T_1,\ldots,T_n),\mathfrak{A},\sigma) =_{def} \operatorname{F}_f(\operatorname{EL}(T_1,\mathfrak{A},\sigma),\ldots,\operatorname{EL}(T_n,\mathfrak{A},\sigma))$ if f is a function symbol of arity at least 1.

Let T be a term of Σ containing no variables (i.e. T is a ground term). Let $c \in D$ (note that by definition D must be non-empty) and define $\sigma_0(x) = d$, for all $x \in V$. Then T corresponds to $EL(T, \mathfrak{A}, \sigma_0) =: EL(T, \mathfrak{A})$.

The above definition shows how terms are interpreted. Let us now turn to the interpretation of formulas. For a given \mathfrak{A} -state σ and a fixed $v \in V, \xi \in \mathsf{D}$ we define

$$\sigma \langle v := \xi \rangle =_{\scriptscriptstyle def} \begin{cases} \sigma(v') & , \text{ if } v' \neq v \\ \xi & , \text{ if } v = v' \end{cases}$$

Definition 3.2. Let $\mathfrak{A} = [\Sigma, D, F, P]$ be a Σ -algebraic structure. We define a relation $\models_{\overline{k}}$ among Σ -algebraic structures and the first-order language $\mathcal{L}_1(\Sigma, V)$ as follows:

We say that Φ is *valid* in \mathfrak{A} if and only if $\mathfrak{A}, \sigma \models_{\overline{k_3}} \Phi$, for all σ . We say that \mathfrak{A} is a model of a set X (denoted by $\mathfrak{A} \models_{\overline{k_3}} X$) if and only if $\mathfrak{A} \models_{\overline{k_3}} \Phi$ for every $\Phi \in X$.

Remark 3.1. If we require that $P_{\rm P}^+ = D \setminus P_{\rm P}$ holds for all $P \in PS$ the above definition coincides with the satisfiability relation for classical, first-order logic.

The above definition allows a formula to be both, true and false. It does however not allow for any truth-value gap. That is for each formula Φ we have that Φ is true under some assignment or false (or both).

Proposition 3.1. For any structure \mathfrak{A} and any assignment σ we have: $\mathfrak{A}, \sigma \models_{\overline{\mathsf{K}_3}} \Phi$ or $\mathfrak{A}, \sigma \models_{\overline{\mathsf{K}_3}} \Phi$

Proof. Assume that Φ is atomic; the case of compound follows immediately by structural induction. By definition a Σ -algebraic structure, we have that

$$\mathsf{P}_{\Phi}^{+} \cup \mathsf{P}_{\Phi}^{-} = \mathsf{D}^{n} \tag{(*)}$$

Assume that Φ has the terms t_1, \ldots, t_n as arguments. By (*) we thus have

$$(\mathrm{EL}(t_1,\mathfrak{A},\sigma),\ldots,\mathrm{EL}(t_n,\mathfrak{A},\sigma))\in\mathsf{P}_{\Phi}^+\text{ or }\\(\mathrm{EL}(t_1,\mathfrak{A},\sigma),\ldots,\mathrm{EL}(t_n,\mathfrak{A},\sigma))\in\mathsf{P}_{\Phi}^-.$$

Hence, $\mathfrak{A}, \sigma \mid_{\overline{\mathbf{k}_3}} \Phi$ or $\mathfrak{A}, \sigma \mid_{\overline{\mathbf{k}_3}} \neg \Phi$. Thus, $\mathfrak{A}, \sigma \mid_{\overline{\mathbf{k}_3}} \Phi$ or $\mathfrak{A}, \sigma \underset{\overline{\mathbf{k}_3}}{=} |\Phi$.

Example 3.1. Let $\Sigma = (\{P, Q\}, \{a, b\}, \alpha)$ be a signature. P and Q are one-place predicate symbols, and a, b 0-ary function symbols. Let $X = \{Pa, \neg Pa, Pb, \forall x (Px \rightarrow Qx)\}$. The following structures are (among others) models of X:

$$\begin{aligned} \mathfrak{A}_{1} &= \begin{bmatrix} \varSigma, & & \\ \mathsf{D} &= \{a', b'\}, & & \\ \mathsf{F}(a) &= a', \mathsf{F}(b) &= b', & \\ \mathsf{P}_{\mathrm{P}}^{+} &= \{a', b'\}, \mathsf{P}_{\mathrm{P}}^{-} &= \{a', b'\}, & \\ \mathsf{P}_{\mathrm{Q}}^{+} &= \varnothing, \mathsf{P}_{\mathrm{Q}}^{-} &= \{a', b'\}, & \\ \mathsf{P}_{\mathrm{Q}}^{+} &= \varnothing, \mathsf{P}_{\mathrm{Q}}^{-} &= \{a', b'\}, & \\ \end{bmatrix} \end{aligned}$$

As we see in the above example, \mathfrak{A}_1 does not satisfy Qb. As in the propositional case, the problem is that \mathfrak{A}_1 judges too many formulas to be paraconsistent, because it satisfies $Pa, \neg Pa, Pb, \neg Pb$. The set X, however, does not mention that Pb should be paraconsistent. What we need is a measure for the degree of paraconsistency of a structure, i.e. we have to extend the relation \sqsubset to Σ -algebraic structures.

Definition 3.3 (\sqsubseteq , $|_{\overline{K_3}}$). Let $\mathfrak{A} = (\Sigma, D, F, P)$, $\mathfrak{A}' = (\Sigma, D, F, P')$ be two structures. Define

 $\mathfrak{A} \sqsubseteq \mathfrak{A}'$ if and only if $\mathsf{P}_{\mathsf{P}}^+ \cap \mathsf{P}_{\mathsf{P}}^- \subseteq \mathsf{P}_{\mathsf{P}}^{+'} \cap \mathsf{P}_{\mathsf{P}}^{-'}$ for every $P \in \mathsf{PS}$ $\mathfrak{A} \sqsubseteq \mathfrak{A}'$ if and only if $\mathfrak{A} \sqsubseteq \mathfrak{A}'$ and NOT $\mathfrak{A}' \sqsubseteq \mathfrak{A}$.

Analogously to the propositional case we say that \mathfrak{A} is a preferred model for X (denoted by $\mathfrak{A} \models_{\mathbf{k}} X$) if and only if \mathfrak{A} is a least model for X (according to \Box)

The structure \mathfrak{A}_2 of Example 3.1 is less informed than \mathfrak{A}_1 (i.e. $\mathfrak{A}_2 \sqsubset \mathfrak{A}_1$). It is easy to see that \mathfrak{A}_2 is a \sqsubset -minimal structure which satisfies $X = \{Pa, \neg Pa, Pb, \forall x (Px \rightarrow Qx)\}$, because minimising the intersection $\mathsf{P}_{\mathsf{P}}^+ \cap \mathsf{P}_{\mathsf{P}}^-$ would yield the empty set and thus either Pa or $\neg Pa$ would not be satisfied.

Note that the relation \sqsubset differs from the substructure relation, which can be found in any textbook on model theory (e.g. [Chang and Keisler, 1977]) and is defined as follows:

Definition 3.4 (Substructure, Union). Let $\mathfrak{A} = [\Sigma, D, F, P^+, P^-], \mathfrak{A}' = [\Sigma, D', F', P^+', P^-']$ be two structures. We say that \mathfrak{A}' is a *substructure* of $\mathfrak{A}, \mathfrak{A}' \subseteq \mathfrak{A}$, if and only if $D' \subseteq D$ and

- 1. Each *n*-placed relation P' of P^+' or P^-' is the restriction to D' of the corresponding relation P of P^+ or P^- , i.e. $P' = P \cap (D')^n$.
- 2. Each *m*-placed function f' of F' is the restriction to D' of the corresponding function f of F, i.e. $f' = f \downarrow (D')^m$

 \mathfrak{A}' is a *strict substructure* of \mathfrak{A} , $\mathfrak{A}' \subset \mathfrak{A}$ if and only if the inclusion $D \subseteq D'$ is strict.

The union $\mathfrak{A}' \cup \mathfrak{A}$ is defined as $[\Sigma, \mathsf{D} \cup \mathsf{D}', \mathsf{F} \cup \mathsf{F}', \mathsf{P}^+ \cup \mathsf{P}^+', \mathsf{P}^- \cup \mathsf{P}^-']$.

Proposition 3.2. $\mathfrak{N} \subseteq \mathfrak{M}$ does not imply $\mathfrak{N} \sqsubseteq \mathfrak{M}$. Moreover, $\mathfrak{N} \sqsubseteq \mathfrak{M}$ does not imply $\mathfrak{N} \subseteq \mathfrak{M}$.

Proof. Note that the relation \sqsubseteq requires that both structures have the same universe D. This is not required by the definition of a substructure. For the converse direction, note that if \mathfrak{N} and \mathfrak{M} have the same domain and $\mathfrak{N} \subseteq \mathfrak{M}$, we must have $\mathfrak{N} = \mathfrak{M}$. But we do not have that $\mathfrak{N} \sqsubseteq \mathfrak{M}$ implies $\mathfrak{N} = \mathfrak{M}$ even though the universes of both structures coincide.

Remark 3.2. Let $\mathfrak{A}_1 \sqsubset \mathfrak{A}_2$ be two Σ -structures. Then for every term T and every state σ we have $\operatorname{EL}(T, \mathfrak{A}_1, \sigma) = \operatorname{EL}(T, \mathfrak{A}_2, \sigma)$. Moreover, if T is a ground term of Σ we have $\operatorname{EL}(T, \mathfrak{A}_1) = \operatorname{EL}(T, \mathfrak{A}_2)$ (cf. Definition 3.1)
We can now define a nonmonotonic entailment relation $\parallel_{\overline{K_3}}$

Definition 3.5 $(\Vdash_{\overline{K_3}}, Cn_{\mathsf{K}})$. Let $X \subseteq \mathcal{L}_1$, Φ be a first-order formula. We say that X K₃-entails Φ (denoted by $X \Vdash_{\overline{K_3}} \Phi$) if and only if every preferred model of X is a model of Φ . Further, $Cn_{\mathsf{K}}(X) =_{def} \{\Phi \mid X \Vdash_{\overline{K_3}} \Phi\}$.

Example 3.1. (continued) The following models are (among others) \sqsubset -minimal for X:

$$\begin{aligned} \mathfrak{A}_{1} &= \left[\begin{array}{cc} \varSigma, & \mathfrak{A}_{2} = \left[\begin{array}{c} \varSigma, \\ \mathsf{D} &= \{a', b', c'\}, & \mathsf{D} = \{a', b'\}, \\ \mathsf{F}(a) &= a', \mathsf{F}(b) &= b', & \mathsf{F}(a) &= a', \mathsf{F}(b) &= b', \\ \mathsf{P}_{\mathrm{P}}^{\mathrm{H}} &= \{a', b', c'\}, \mathsf{P}_{\mathrm{P}}^{\mathrm{H}} = \{a', b'\}, & \mathsf{P}_{\mathrm{P}}^{\mathrm{H}} = \{a', b', \mathsf{P}_{\mathrm{P}}^{\mathrm{H}} = \{a', b'\}, \\ \mathsf{P}_{\mathrm{Q}}^{\mathrm{H}} &= \{a'\}, \mathsf{P}_{\mathrm{Q}}^{\mathrm{H}} = \{b'\} & \mathsf{P}_{\mathrm{Q}}^{\mathrm{H}} = \{b'\}, \mathsf{P}_{\mathrm{Q}}^{\mathrm{H}} = \{a'\} \end{aligned}$$

 \mathfrak{A}_1 contains an object c' which is not needed. Thus, \mathfrak{A}_2 is a substructure of \mathfrak{A}_1 . Beside \mathfrak{A}_1 and \mathfrak{A}_2 there is another preferred model of X:

$$\begin{aligned} \mathfrak{A}_{3} = \; [\; \varSigma, \\ \mathsf{D} = \{a', b'\}, \\ \mathsf{F}(a) = a', \mathsf{F}(b) = b', \\ \mathsf{P}_{\mathrm{P}}^{+} = \{a', b'\}, \mathsf{P}_{\mathrm{P}}^{-} = \{a'\}, \\ \mathsf{P}_{\mathrm{Q}}^{+} = \{a', b'\}, \mathsf{P}_{\mathrm{Q}}^{-} = \varnothing \end{aligned}$$

 \mathfrak{A}_2 and \mathfrak{A}_3 are up to isomorphism the only preferred models of X. Thus, $Qb \in Cn_{\mathsf{K}}(X)$ but $Qa, \neg Qa \notin Cn_{\mathsf{K}}(X)$.

We know that in propositional K_3 , any set of formulas X which is propositionally consistent, has the same consequences in K_3 as in classical propositional logic. This guarantees that we get only nonmonotonic effects, if we add to any set X a formula Φ which is contradictory to some information in X. This property does also hold for the 1st-order case, as stated in the following proposition.

Proposition 3.3. Let $X \subseteq \mathcal{L}_1$. If X has a classical first-order model, then $\operatorname{Cn}_{cl}^{1st}(X) = \operatorname{Cn}_{\kappa}(X)$, where $\operatorname{Cn}_{cl}^{1st}$ is the consequence operator of classical first-order logic.

Proof. If X has a classical first-order model, then in any preferred model \mathfrak{A} of X we have that $\mathsf{P}^+ \cap \mathsf{P}^-$ is empty. Since $\mathsf{P}^+_{\mathsf{P}} \cup \mathsf{P}^-_{\mathsf{P}} = \mathsf{D}^n$ we must have that $\mathsf{P}^-_{\mathsf{P}}$ is the complement of $\mathsf{P}^+_{\mathsf{P}}$. Hence, we have X is satisfied by a classical first-order structure if and only if it is valid in some preferred model \mathfrak{A} of X. \Box

From the above proposition it follows that the set of tautologies of classical first-order logic coincides with the set of first-order Kleene tautologies. We thus have the following

Corollary 3.1. $\operatorname{Cn}_{\operatorname{cl}}^{\operatorname{1st}}(\emptyset) = \operatorname{Cn}_{\mathsf{K}}(\emptyset).$

Let us denote the first-order version of K_3 by FOK_3 .

3.2 Properties of FOK_3

We begin with some properties for the records:

Proposition 3.4. The operator Cn_{K} is a nonmonotonic, cumulative preclosure operator, *i.e.*

- 1. $X \subseteq Cn_{\mathsf{K}}(X)$ (Inclusion)
- 2. $\operatorname{Cn}_{\mathsf{K}}(X) = \operatorname{Cn}_{\mathsf{K}}(\operatorname{Cn}_{\mathsf{K}}(X))$ (Idempotency)
- 3. $\Phi, \Psi \in Cn_{\mathsf{K}}(X)$ implies $\Phi \in Cn_{\mathsf{K}}(X \cup \{\Psi\})$ (Cumulativity)
- 4. $X \subseteq Y$ does not imply $\operatorname{Cn}_{\mathsf{K}}(X) \subseteq \operatorname{Cn}_{\mathsf{K}}(Y)$ (Nonmonotonicity)

5. Cn_{K} is not compact.

Proof. Part 1 and 2 follow immediately from the Definition of Cn_{κ} . Part 3 is similar to Theorem 2.5. Part 4 and 5 is by counter-example.

3.2.1 Herbrand-structures

A structure $\mathfrak{A} = [\Sigma, \mathsf{D}, \mathsf{F}, \mathsf{P}]$ is said to be an Herbrand-structure if and only if

- 1. D is the set of all variable free terms (ground terms) constructible from Σ .
- 2. for every 0-ary function symbol $f \in FS$ we have $\Phi_f = f$
- 3. for every *n*-ary (n > 0) function symbol and every variable free term T_1, \ldots, T_k we have $\Phi_f(T_1, \ldots, T_k) = f(T_1, \ldots, T_k)$.

We summarise some well-known results on Herbrand-structures.

Theorem 3.1. Let $X \subseteq \mathcal{L}_1(\Sigma, V)$ be a universal theory.

- 1. X has a classical model if and only if X has a classical Herbrand-model.
- 2. X has a classical model if and only if the set $\{\Phi \mathbf{c} \mid \Phi x \in X\}$ of all ground instances of X has a classical model.

3. If Σ contains at least one constant symbol and $\exists x \Phi x \in \operatorname{Cn}_{cl}^{1\mathrm{st}}(X)$ then there are constant symbols c_1, \ldots, c_n such that $\Phi c_1 \vee \ldots \vee \Phi c_n \in \operatorname{Cn}_{cl}^{1\mathrm{st}}(X)$.

The last point is also called Herbrand's Theorem. The main aim of this section is to investigate whether Herbrand's Theorem holds for FOK₃. To see that this question is not beyond any scope, note that the Herbrand's Theorem does not hold for minimal reasoning as used in the field of Logic Programming. Here one considers only models which are (1) consistent, in a classical sense (that is for any $P \in PS$ we have $P_P^+ \cap P_P^- = \emptyset$) and (2) in which the number of true propositions is reduced to a minimum. That is two structures $\mathfrak{A}_1 = [\Sigma, D, F, P_1]$, $\mathfrak{A}_2 = [\Sigma, D, F, P_2]$ are compared via the relation \triangleleft , where $\mathfrak{A}_1 \triangleleft \mathfrak{A}_2$ if and only if for every $P \in PS$ we have $P_{1_P} \subseteq P_{2_P}$. Based on this notion we can define a consequence operator

$$Cn_{\triangleleft}(X) =_{def} \{ \Phi \mid \text{if } \mathfrak{A} \text{ is a } \triangleleft -\text{minimal model of } X, \\ \text{then } \mathfrak{A} \text{ is a model of } \Phi. \}$$

As a counter-example (which is due to Dix and taken from [Herre, 1990]) consider a signature containing constant symbols $\{1, 2, 3, ...\}$ and the set

$$X = \{ Pi \lor Pj \mid i \neq j \quad i, j \le \omega \}$$

We have that $\exists x \neg Px$ holds in every \triangleleft -minimal model of X. There is, however, no finite disjunction $\varphi_k = \neg Pc_1 \lor \ldots \lor \neg Pc_k$, where $c_i \in \{1, 2, 3, \ldots\}$. Thus, $\exists x Px \in Cn_{\triangleleft}(X)$ but $\varphi_k \notin Cn_{\triangleleft}(X)$, for every $k \in \mathbb{N}$.

Unfortunately, there is a counter-example to Herbrand's Theorem for FOK_3 as well.

Theorem 3.2. Herbrand's Theorem does not hold for FOK_3 .

Proof. Consider again a signature containing all natural numbers $\{1, 2, 3, ...\}$ as constant symbols. Let

$$\begin{split} X =_{def} \{ (\neg Pi \lor \neg Pj) \land \\ Pi \land Pj \land \\ Pi \to Qi \land \\ Pj \to Qj \mid i \neq j, \quad i, j \leq \omega \} \end{split}$$

The above set has infinitely many preferred FOK_3 -Herbrand-models. For every such model, there is some constant c such that $c \notin P_P^+ \cap P_P^-$. Hence, $\exists xQx \in Cn_K(X)$, but there is no $\varphi_k = Qc_1 \vee \ldots \vee Qc_k$, $k < \omega$ such that $\varphi_k \in Cn_K(X)$. We close this section with a proposition which states that if a FOK₃-Herbrand structure \mathfrak{H} satisfies exactly the same grounded formulas as the structure \mathfrak{A} , then \mathfrak{H} satisfies also all universal sentences which are satisfied by \mathfrak{A} .

Proposition 3.5. Let Σ be a signature such that the set of ground terms is not empty. Further, let X be a set of universal formulas and \mathfrak{A} a model of X. Let $\mathfrak{H} = [\Sigma, \mathsf{D}, \mathsf{F}, \mathsf{P}]$ be a Herbrand structure such that for all quantifier-free sentences Φ we have $\mathfrak{A} \models_{\overline{\mathsf{K}}_3} \Phi \Leftrightarrow \mathfrak{H} \models_{\overline{\mathsf{K}}_3} \Phi$. Then, $\mathfrak{H} \models_{\overline{\mathsf{K}}_3} X$.

Proof. Assume to the contrary that $\mathfrak{H} \not\models_{\mathfrak{K}_3} X$. Since X contains only universal formulas, we must have $\mathfrak{H} \not\models_{\mathfrak{K}_3} \forall \mathbf{x} \Theta$, for some $\forall \mathbf{x} \Theta \in X$. Without loss of generality we can assume that the quantifier-free formula Θ is a disjunction of literals; we further assume that Θ consists of exactly one literal $\varphi(t_1, \ldots, t_n)$. It follows that for some \mathfrak{H} -state σ we have

$$\mathfrak{H}, \sigma \not\models \varphi(t_1, \ldots, t_n)$$

Assume that $\varphi(t_1, \ldots, t_n)$ is a positive literal. Hence,

$$[\mathrm{EL}(\mathfrak{H},\sigma,t_1),\ldots,\mathrm{EL}(\mathfrak{H},\sigma,t_n)]\not\in P_{\omega}^+.$$

Since \mathfrak{H} is a Herbrand structure, there must be a quantifier-free sentence $\varphi \mathbf{c}$ which is satisfied by \mathfrak{A} but not by \mathfrak{H} – a contradiction.

We shall now analyse some fundamental properties of our paraconsistent first-order extension.

3.2.2 Preferred Model Existence

Priest proved the following weak version of preferred model existence:

Proposition 3.6 (Priest, 1991). Let \mathfrak{A} be a finite structure, i.e. the domain is finite, such that $\mathfrak{A} \models_{\overline{K}} X$. Then there is a \mathfrak{A}' such that $\mathfrak{A}' \models_{\overline{K}} X$.

For the general case, the existence of a preferred model cannot always be guaranteed:

Proposition 3.7. There are finite sets of formulas which do not have a preferred FOK_3 -model. *Proof.* The following counter-example is due to Heinrich Herre. Consider the theory of dense linear orderings plus the following formulas:

$$\begin{split} & \varPhi_1 : \forall x \forall y (Px \land (x < y)) \to Py \\ & \varPhi_2 : \exists Px \\ & \varPhi_3 : Px \leftrightarrow (Qx \land \neg Qx) \end{split}$$

The set has the following model:

$$\begin{split} \mathfrak{A} &= \left[\begin{array}{cc} \varSigma, \\ \mathsf{D} &= I\!\!N, \\ \mathsf{F} &= \varnothing, \\ \mathsf{P}_{<}^{+} &= \{(a,b) \mid a < b; \quad a,b \in I\!\!N\}, \mathsf{P}_{<}^{-} &= \varnothing, \\ \mathsf{P}_{\mathsf{P}}^{+} &= I\!\!N, \mathsf{P}_{\mathsf{P}}^{-} &= \varnothing, \\ \mathsf{P}_{\mathsf{Q}}^{+} &= I\!\!N, \mathsf{P}_{\mathsf{Q}}^{-} &= I\!\!N \\ \end{split} \right]$$

We can restrict P_P^+ , P_Q^+ and P_Q^- to $I\!N \setminus \{0\}$. This yields a model \mathfrak{A}' with $\mathfrak{A}' \sqsubset \mathfrak{A}$. By removing the minimal element of the domain of \mathfrak{A}' we can again get a model which is less informed than \mathfrak{A}' and so on. Obviously, there is a chain $\mathfrak{A} \sqsupset \mathfrak{A}' \sqsupset \mathfrak{A}' \sqsupset \mathfrak{A}'' \sqsupset \ldots$ which has no minimal element. Hence there is no preferred model for the above theory.

Fortunately, there are theories, which always have a preferred model:

Theorem 3.3. Let Σ be a signature containing at least one constant symbol, X a set of universal sentences and \mathfrak{H}_0 a Herbrand-model of X. Then, there is a preferred Herbrand-model \mathfrak{H} of X such that $\mathfrak{H} \sqsubset \mathfrak{H}_0$.

Proof. Since X is a universal, we know that each element of X has the form

$$\forall x_1 \forall x_2 \dots \forall x_n \Phi(x_1, \dots, x_n)$$

where \varPhi is quantifier-free. We can transform \varPhi to a sort of conjunctive normal form

$$\forall x_1 \forall x_2 \dots \forall x_n \Phi_1(x_1, \dots, x_n) \land \dots \land \Phi_k(x_1, \dots, x_n)$$

where each Φ_i is a disjunction of first-order literals. Distributing the quantifiers yields

$$\forall x_1 \forall x_2 \dots \forall x_n \Phi_1(x_1, \dots, x_n)$$

$$\land \quad \forall x_1 \forall x_2 \dots \forall x_n \Phi_2(x_1, \dots, x_n)$$

$$\vdots$$

$$\land \quad \forall x_1 \forall x_2 \dots \forall x_n \Phi_k(x_1, \dots, x_n)$$

We thus can assume that each element of X is a universal formula of the form $\forall x \Phi_i$ where Φ_i is a disjunction of literals. Consider an Herbrand-model $\mathfrak{H}_0 = [\Sigma, \mathsf{D}, \mathsf{F}, \mathsf{P}]$ of X. Define for every universal formula $\Phi = \forall \mathbf{x} \varphi \mathbf{x}$ a set $Z_{\Phi} = \{\varphi \mathbf{c} \mid \varphi \mathbf{c} \text{ is a ground instance of } \Phi\}$. Let $Z = \{Z_{\Phi} \mid \Phi \in X\}$. Obviously, for every structure $\mathfrak{B} = [\Sigma, \mathsf{D}, \mathsf{F}, \ldots]$ we have

$$\mathfrak{B} \models_{\overline{\mathrm{K}_3}} X \Leftrightarrow \mathfrak{B} \models_{\overline{\mathrm{K}_3}} X \cup Z$$

Let $Z = \{ Z_{\Phi} \mid \Phi \in X \}.$

Consider now a descending chain $\mathfrak{H}_0 \supseteq \mathfrak{H}_1 \supseteq \mathfrak{H}_2 \supseteq \ldots$ of Herbrand-models of X. Define

$$\operatorname{Inc}(\mathfrak{H}_{i}, P) =_{{}_{def}} \mathsf{P}^{+}_{\mathbf{i}, \mathsf{P}} \cap \mathsf{P}^{-}_{\mathbf{i}, \mathsf{P}}$$

and let $\operatorname{Inc}(P) =_{def} \bigcap \operatorname{Inc}(\mathfrak{H}_i, P)$. Let Y be a subset of Z such that

$$Y = \{ \varphi \mid \varphi \in Z \text{ for every atomic } P(c_1, \dots, c_n) \text{ appearing as a subformula} \\ \text{ in } \varphi : (\text{EL}(c_1, \mathfrak{H}_0), \dots, \text{EL}(c_n, \mathfrak{H}_0)) \notin \text{Inc}(P) \}$$

Every finite subset of Y has a classical two-valued model. By compactness of classical first-order logic we know that Y has a classical model and by Theorem 3.1 it has a classical Herbrand-model $\mathfrak{H}_Y = [\Sigma, \mathsf{D}, \mathsf{F}, (\mathsf{P}^+_Y, \mathsf{P}^-_Y)]$. We extend \mathfrak{H}_Y in order to become a model of Z. Define $\mathfrak{H}_Z = [\Sigma, \mathsf{D}, \mathsf{F}, \mathsf{P}_Z]$ with

$$\begin{split} \mathsf{P}_{Z} &=_{\scriptscriptstyle def} (\mathsf{P}^{+},\mathsf{P}^{-}), \text{ where } \\ \mathsf{P}_{\mathrm{P}}^{+} &=_{\scriptscriptstyle def} \mathsf{P}_{\mathrm{P}Y}^{+} \cup \operatorname{Inc}(P) \\ \mathsf{P}_{\mathrm{P}}^{-} &=_{\scriptscriptstyle def} \mathsf{P}_{\mathrm{P}Y}^{-} \cup \operatorname{Inc}(P) \end{split}$$

We have $\mathfrak{H}_Z \models_{\overline{K_2}} Z$: let $\varphi \in Z$ be a disjunction of literals; there are two cases:

 $\varphi \in \mathbf{Y}$. Since \mathfrak{H}_Z is an extension of \mathfrak{H}_Y we have $\mathfrak{H}_Z \models_{\overline{\mathbf{X}}_2} \varphi$.

 $\varphi \notin \mathbf{Y}$. In this case there must be an atomic subformula $P(c_1, \ldots, c_n)$ appearing in φ such that $(\operatorname{EL}(c_1, \mathfrak{H}_0), \ldots, \operatorname{EL}(c_n, \mathfrak{H}_0)) \in \operatorname{Inc}(P)$. From the construction of \mathfrak{H}_Z it follows that $\mathfrak{H}_Z \models_{\overline{\mathfrak{H}}_3} P(c_1, \ldots, c_n)$ and hence $\mathfrak{H}_Z \models_{\overline{\mathfrak{H}}_3} \varphi$.

But \mathfrak{A}_Z is also a model of X: assume to the contrary that $\mathfrak{A}_Z \not\models_{\mathsf{K}_S} X$. Hence, there must be $\Phi = \forall \mathbf{x} \varphi \mathbf{x} \in X$ such that $\mathfrak{A}_Z \not\models_{\mathsf{K}_S} \Phi$. Without loss of generality assume that $\varphi \mathbf{x}$ is atomic. There must be a Σ -state $\sigma : V \to \mathsf{D}$ such that

 $\mathfrak{A}_Z, \sigma \not\models_{\mathsf{K}_2} \forall \mathbf{x} \varphi \mathbf{x}.$

By Definition 3.2 that there is a $\xi \in D$ such that

$$\mathfrak{A}_Z, \langle v := \xi \rangle \not\models_{\kappa} \varphi \mathbf{x}$$

Since \mathfrak{H}_Z is a Herbrand-model there must be a ground instance $\varphi \mathbf{c}$ of $\varphi \mathbf{x}$ such that $\mathfrak{H}_Z \not\models_{\mathsf{MK}} \varphi \mathbf{c}$. This implies $\mathfrak{H}_Z \not\models_{\mathsf{MK}} Z$ – a contradiction. Hence, \mathfrak{H}_Z is a model of X. Moreover, $\mathfrak{H}_Z \sqsubset \mathfrak{H}_i$, for every \mathfrak{H}_i of the descending

Hence, \mathfrak{H}_Z is a model of X. Moreover, $\mathfrak{H}_Z \sqsubset \mathfrak{H}_i$, for every \mathfrak{H}_i of the descending chain. It follows from Zorn's lemma that the above chain has a minimal element. \Box

The following proposition is now easy to show:

Proposition 3.8. Let X be a set of universal sentences, Φ a universal sentence. If $\mathfrak{A} \models_{\overline{K}_3} X$ and $\mathfrak{A} \models_{\overline{K}_3} \Phi$ then $\mathfrak{A} \models_{\overline{K}_3} X \cup \{\Phi\}$.

Proof. Similar to the propositional case.

3.2.3 Deduction Theorems

We already saw that the general version of the deduction theorem fails for propositional K₃. Hence, we cannot expect that the full version holds for the first-order extension FOK₃. Moreover, the full version of the deduction theorem fails also for standard first-order logic, provided that the concept of entailment is defined as 'X entails Φ if and only if every structure which satisfies X does also satisfy Φ '. As a counter-example consider: $Px \Vdash \forall xPx$ which holds in classical logic; but obviously we do not have that $Px \to \forall xPx$ is a theorem.

The weak version of the deduction theorem, however, does hold for classical first-order logic:

 $X \cup \{\Phi\} \Vdash \Psi \text{ implies } X \Vdash \Phi \to \Psi$

where ⊢ denotes standard first-order validity entailment, cf.[Avron, 1991].

Standard first-order logic (FOL) satisfies the following version of the deduction theorem, which is called 'sential' because it deals only with sentences, i.e. formulas where each variable is in the scope of some quantifier.

Theorem 3.4 (Sential Deduction Theorem for FOL). Let Φ be a sentence, i.e. Φ contains no free variables. Then the following holds: $X \cup \{\Phi\} \Vdash \Psi$ implies $X \Vdash \Phi \to \Psi$.

A analogous version holds for FOK_3 :

Theorem 3.5 (Sential Deduction Theorem for FOK₃). Let Φ be a sentence, i.e. Φ contains no free variables. Then the following holds: $X \cup \{\Phi\} \Vdash_{\overline{K_3}} \Psi$ implies $X \Vdash_{\overline{K_3}} \Phi \to \Psi$.

Proof. We have to show that $\Phi \to \Psi$ is valid in every preferred model \mathfrak{A} of X. Suppose that $\mathfrak{A} \models_{\overline{\mathfrak{K}}} X$. Thus, we have to prove that for every state σ we have $\mathfrak{A}, \sigma \models_{\overline{\mathfrak{K}}} \Phi \to \Psi$, i.e.

$$\mathfrak{A}, \sigma \equiv \Phi \text{ or } \mathfrak{A}, \sigma \models \Psi.$$

By Proposition 3.1 we have that for each $\mathfrak{A}, \sigma : \mathfrak{A}, \sigma \models_{\overline{k_3}} \Phi$ or $\mathfrak{A}, \sigma \models_{\overline{k_3}} \Phi$. In the latter case we are done. Thus assume that $\mathfrak{A}, \sigma \models_{\overline{k_3}} \Phi$. Since Φ is a sentence we have $\mathfrak{A} \models_{\overline{k_3}} \Phi$. By Proposition 3.8 we know that $\mathfrak{A} \models_{\overline{k_3}} X$, $\mathfrak{A} \models_{\overline{k_3}} \Phi$ implies $\mathfrak{A} \models_{\overline{k_3}} X \cup \{\Phi\}$. Thus, $\mathfrak{A} \models_{\overline{k_3}} \Psi$. Hence, $\mathfrak{A} \models_{\overline{k_3}} \Phi \to \Psi$.

Note, that the restriction to sentences in the above theorem is not dramatical at all because first-order logic's entailment relation is invariant w.r.t. generalisation. That is, $X \Vdash \Phi$ if and only if $\operatorname{GEN}(X) \Vdash \Phi$, where $\operatorname{GEN}(X)$ is the *generalisation* of X, i.e. each free variable occurring in a formula Φ of X is bounded by a universal quantifier.

Likewise in the propositional case, the converse direction of the sential deduction theorem does not hold. However, there is a normalised version which does hold. The normalisation condition for FOK₃ is analogous to the one for K₃. $\Phi \rightarrow \Psi$ satisfies the normalisation condition, if and only if for every \mathfrak{A}, σ we have

 $\mathfrak{A}, \sigma \not\models_{\overline{k}} \Phi \to \Psi \quad \Leftrightarrow \quad \mathfrak{A}, \sigma \models_{\overline{k}} \Phi \text{ and } \mathfrak{A}, \sigma \not\models_{\overline{k}} \Psi \quad \text{Normalisation Condition (NC)}$

Theorem 3.6 (Normalised Deduction Theorem). Let $\Phi \to \Psi$ be a firstorder tautology which satisfies the normalisation condition. We then have

 $X \Vdash_{\overline{\kappa}_3} \Phi \to \Psi$ if and only if $X \cup \{\Phi\} \Vdash_{\overline{\kappa}_3} \Psi$

Proof. Similar to the propositional case.

The normalised deduction theorem closes our considerations on the purely mathematical aspects of FOK_3 . Let us now turn to a more philosophical aspect of FOK_3 .

3.3 Extracting Consistent Information

The propositional logic K_3 has – at least from a philosophical point of view – an important feature: if a literal L is K_3 -entailed by X, then L is based on a consistent argument in X or $L \in X$, provided that X is a set of clauses. This consistent argument can be identified with a consistent subset of X.

Selecting the consistent part of a propositional database was done by

- 1. the prerequisite that X is in clausal normal-form and
- 2. eliminating all clauses from X in which no literal L takes the value \top in some preferred model of X.

In the first-order case the situation is somewhat more complicated because there is nothing like a conjunctive normal-form and hence there is nothing like INF. Suppose we have $\exists x Px \land \neg Px$. The existential quantifier glues the subformulas Px and $\neg Px$ together, so that we cannot generate a proper CNF. However, it easy to see that the following formulas are tautologies which satisfy the normalisation condition:

$$\exists x P x \land \neg P x \to \exists x P x \\ \exists x P x \land \neg P x \to \exists x \neg P x \end{cases}$$

In this section we shall show that each consistent formula which is entailed by an inconsistent formula α is already entailed by a consistent 'part' of α , i.e. by a subformula which has a two-valued model.

The following proposition establishes the link between inconsistent information and consistent conclusions.

Proposition 3.9. Let α be a sentence, $\{\alpha\} \Vdash_{\overline{\mathsf{K}_3}} \Phi$ and Φ be consistent, i.e. Φ has a classical two-valued model. There is a consistent β such that

1. $\alpha \rightarrow \beta$ is a tautology which satisfies NC 2. $\beta \rightarrow \Phi$ is a tautology. *Proof.* There are two cases: either α has a two-valued model or not. The first case is easy. Since α contains no free variables, we have by Theorem 3.4 that $\alpha \to \Phi$ is a tautology. Choose $\beta = \alpha$.

In the second case, we know that α is inconsistent. Let \mathfrak{A} be a model for α . It is easy to see that there are structures \mathfrak{A}_1 , \mathfrak{A}_2 of \mathfrak{A} with $\mathfrak{A}_1 \sqsubset \mathfrak{A}$ and $\mathfrak{A}_2 \sqsubset \mathfrak{A}$ such that

$$\begin{aligned} \mathfrak{A}_1 &\models_{\overline{k}_3} \alpha \text{ and } \mathfrak{A}_2 &\models_{\overline{k}_3} \alpha \\ \mathfrak{A}_1 \cup \mathfrak{A}_2 &= \mathfrak{A} \\ \mathfrak{A}_1 &\models_{\overline{k}_3} \beta_1 \text{ and } \mathfrak{A}_1 &\models_{\overline{k}_3} \beta_1 \\ \mathfrak{A}_2 &\models_{\overline{k}} \beta_2 \text{ and } \mathfrak{A}_2 &\models_{\overline{k}} \beta_2 \end{aligned}$$

for some β_1, β_2 . Roughly speaking, \mathfrak{A}_1 and \mathfrak{A}_2 extract the consistent parts β_1 and β_2 of \mathfrak{A} . For example, if $\alpha \equiv \exists x(Px \land \neg Px)$ then any \mathfrak{A} with $\mathsf{P}_{\mathsf{P}}^+ = \{c\} = \mathsf{P}_{\mathsf{P}}^$ satisfies α , for some object c. Let \mathfrak{A}_1 be the same as \mathfrak{A} except that $\mathsf{P}_{\mathsf{P}}^+$ is empty; analogously let \mathfrak{A}_2 be the same as \mathfrak{A} except that $\mathsf{P}_{\mathsf{P}}^-$ is empty. Clearly, \mathfrak{A}_1 and \mathfrak{A}_2 are both less informed than \mathfrak{A} and neither \mathfrak{A}_1 nor \mathfrak{A}_2 satisfies α , but $\mathfrak{A}_1 \models_{\mathsf{K}_3} \exists x Px$ and \mathfrak{A}_2 and $\mathfrak{A}_2 \models_{\mathsf{K}_3} \exists x \neg Px$. Hence, we have $\beta_1 \equiv \exists x Px$ and $\beta_2 \equiv \exists x \neg Px$.

Since β_1 and β_2 is valid in every structure which satisfies α we have that both $\alpha \to \beta_1$ and $\alpha \to \beta_2$ are tautologies which satisfy NC.

It remains to show that $\beta \to \Phi$ is a tautology. Assume to the contrary that for all consistent β such that $\alpha \Vdash_{\overline{K_3}} \beta$ we have that $\beta \to \Phi$ is no tautology, i.e. $\beta \Vdash_{\overline{K_3}} \Phi$. Since $\alpha \Vdash_{\overline{K_3}} \Phi$ it follows that Φ cannot be consistent – a contradiction. \Box

The following example illustrates the above proposition.

Example 3.2. Let $\alpha \equiv A \land (B \land \neg B) \land (A \to C)$. Clearly, $\alpha \Vdash_{\overline{k_3}} C$. By Proposition 3.9 we know that there must be a consistent formula β such that $\beta \Vdash_{\overline{k_3}} \Phi$ and hence, $\beta \to \Phi$ is a tautology. Moreover, $\alpha \to \beta$ must satisfy NC. The choice, $\beta \equiv A \land A \to C$ is appropriate.

3.4 Discussion

We have seen that the first-order logic FOK₃ is much more complicated than its propositional counterpart K₃. The main drawback is that in general we cannot guarantee that a theory X has a preferred model even though every X has a first-order model. This yields that the entailment relation is explosive, i.e. that there are theories X for which $X \Vdash_{\mathbf{K}_3} \Phi$ is vacuously true. This is clearly not

desired for a paraconsistent logic. However, if we restrict ourselves to universal theories, we can guarantee that there is always a preferred model.

Another drawback is that Herbrand's theorem fails for FOK_3 . This might not be dramatic for applications in Computing Science, for example deductive databases, because here we only deal with finite sets of formulas. The above negative results might, however, be crucial, if one has an application of FOK_3 in Mathematics, e.g. number theory in mind.

CHAPTER 4

Syntactical Characterisation and Proof Theory

We have seen that Kleene's strong three-valued logic can be modified very easily in order to obtain a logic which treats contradictory information cautiously and which behaves identical to classical logic, if the information provided is consistent. In this chapter I shall present some systems of deduction for K_3 .

Systems of deduction allow for a mechanisation of a logic, or more precisely: an entailment relation. This aspect is of great importance for Computing Science. There are, however, other important arguments for developing a deductive system or a proof theory which I would like to account for. For example:

- 1. A formal calculus which axiomatises a given entailment relation is a 'compressed' version of the entailment relation and can facilitate the study of the logic in question, as for example Hilbert's and Ackermann's calculus did for classical logic.
- 2. A formal calculus can reflect valid patterns of inference used by mathematicians as well as in the commonsense world. The historical origin of treating logic as the determination of valid inference patterns are Aristotle's systems of syllogisms. But also Gentzen's calculi of sequents and natural deduction belong to this type of calculi.

Since K_3 is nonmonotonic, we cannot expect to obtain the same easy-going proof theory as in the case of classical logic. Moreover, as the following theorem shows, the set of K_3 -consequences might not even in the propositional case be recursively enumerable. In fact, the failure of compactness indicates that the set $Cn_3(X)$ might not be recursively enumerable (r.e.) when X is r.e. As the following theorem shows, $Cn_3(X)$ is in fact Σ_2^0 -hard, i.e. not r.e.¹

Theorem 4.1. There is a recursive set of sentences X such that $\operatorname{Cn}_3(X)$ is Σ_2^0 -hard.

Proof. Let $\Gamma =_{def} \{n \mid \exists x \forall y \neg R(n, x, y)\}$ be a Σ_2^0 -complete set, where R is a recursive relation. We shall construct a recursive set of sentences X such that Γ is many-to-one reducible to $Cn_3(X)$.

Let $\{A_n \mid n \in \omega\} \cup \{B_{\langle n, x \rangle} \mid n, x \in \omega\}$ be a set of pairwise distinct atomic formulas. Let

$$X =_{{}_{def}} \{A_n \lor (B_{\langle n, x \rangle} \land \neg B_{\langle n, x \rangle}) \mid n, x \in \omega\} \cup \{\bigwedge_{i=0}^y \beta_i(n, x) \mid R(n, x, y)\},\$$

¹ A similar result holds for reasoning with minimal classical models, where f < t induces the corresponding partial ordering. C.f. [Papalaskari and Weinstein, 1990].

where each $\beta_i(n, x) =_{def} B_{\langle n, x \rangle} \wedge \neg B_{\langle n, x \rangle}$. Then the following holds:

- 1. for each $n, x \in \omega$ we have $B_{\langle n, x \rangle} \wedge \neg B_{\langle n, x \rangle} \in Cn_3(X)$ if and only if $\exists y \ R(n, x, y)$, and
- 2. $A_n \notin Cn_3(X)$ if and only if for all n we have $B_{\langle n,x \rangle} \wedge \neg B_{\langle n,x \rangle} \in Cn_3(X)$.

Hence,

$$A_n \not\in Cn_3(X) \Leftrightarrow \text{ for all } x: \exists y R(n, x, y)$$

Thus,

$$A_n \in Cn_3(X) \Leftrightarrow \neg \forall x \exists y \ R(n, x, y)$$
$$\Leftrightarrow \exists x \forall y \neg R(n, x, y)$$
$$\Leftrightarrow n \in \Gamma.$$

It is an open question whether Cn_3 is also a member of Π_2^0 , i.e. whether Cn_3 is Π_2^0 -complete.

Of course, if we restrict ourselves to finite sets X_{fin} , then $Cn_3(X_{\text{fin}})$ becomes decidable.

Proposition 4.1. Let X be finite. Then $Cn_3(X)$ is recursive.

Proof. There are only finitely many models of X and therefore only finitely many preferred models. We can thus easily check whether any arbitrary Φ holds in all preferred models or not.

The above discussion shows, that it is highly questionable whether there are sound and complete r.e. syntactical characterisations for Cn_3 and Cn_K . However, there are several non-trivial approximations. By 'non-trivial' we mean that the consequence relation which approximates Cn_3 fulfils several requirements. Let us therefore introduce the concept of *recursively enumerable approximation*.

Definition 4.1 (R.E. Approximation). Let Cn be a nonmonotonic consequence operator. An operator Cons is a r.e. approximation of Cn if

- 1. Cons is embedding, idempotent, cumulative and nonmonotonic.
- 2. For any r.e. X we have that Cons(X) is r.e.
- 3. $Cons(X) \subseteq Cn(X)$.

4. If X has a two-valued model, then $Cn_{cl}(X) = Cons(X)$. In Chapter 2 we call this property 'conservativity'.

Most of the proof systems to be presented in this chapter are r.e. approximations of Cn_3 or Cn_K , respectively. Only the tableau-calculus gives us an effective procedure for deciding $\Phi \in Cn_3(X)$, provided that X is finite. We shall discuss the following types of proof-systems:

- A Hilbert-System for Monotonic K_3 which characterises Kleene's logic, based on Bolzano's entailment relation, Cn_{Bolz} .
- A Hilbert-Style Formalisation which tries to characterise Cn_3 with the help of a set of axioms and a variant of modus ponens as the only rule of inference. We shall also discuss a Hilbert-style formalisation of FOK₃.
- A Tableau Procedure which gives us an effective and machine oriented method for deciding $Cn_3(X)$.
- A Sequent-Style Calculus which models Cn_3 with the help of valid patterns of inference.

As mentioned before, we cannot expect a nonmonotonic deduction operation to be as straight forward as a monotonic one. As a consequence, nonmonotonic inference rules have an additional component, called *nonmonotonic precondition*, or PC for short. We shall see that the amount of additional information given in the precondition varies from one deductive system to another. For example, we can build a sound and complete tableau procedure for K_3 only by changing the concept of a tableau proof. The Hilbert-style calculus requires the modification of the Modus Ponens inference rule. The situation becomes a little bit more difficult for the sequent calculus. Let me stress right from the beginning that I do not consider the sequent calculus for K_3 to be an appropriate reasoning mechanisation of Cn_3 . But: This sequent calculus represents a sound set of principles showing which inference patterns (in the above sense) are valid.

4.1 A Hilbert System for Monotonic K₃

Recall that

$$Cn_{Bolz} = \{ \Phi \mid I \models X \text{ implies } I \models \Phi \}$$

is a compact closure operator. Moreover, every formula Φ is semantically equivalent to some formula $\Phi_{\rm CNF}$.

We begin with a basic observation of how formulas are entailed.

Proposition 4.2. Let X be a set of clauses, D a nontautological clause. $D \in Cn_{Bolz}(X)$ if and only if there is some clause $D' \in X$ such that $D' \subseteq D$.

Proof.

'⇒' Assume to the contrary that there is no such D'. Let $D = L_1 \lor \ldots \lor L_n$. We shall show that there is some model I of X such that I(D) = f. We can assume that all literals L_i appearing in D also appear in some clause C of X (otherwise we can assign I(L) = f, i.e. I(A) = f if A = L or I(A) = t if $\neg A = L$).

By assumption, there is no clause $C \in X$ such that C consists solely of of literals from $\{L_1, \ldots, L_n\}$. That is, for every clause C of X there is some literal $L_C \in C$ such that $L_C \notin D$. Define an interpretation I such that

$$I(L_C) = \top \quad \text{i.e. } I(A) = \top, \text{ for } A \in \operatorname{ATOM}(L_C)$$
$$I(L_i) = f \quad \text{i.e. } I(A) = f \text{ if } L_i = A \text{ and } I(A) = t \text{ if } L_i = \neg A.$$

Since D is nontautological I is well defined and I(D) = f. Moreover, $I \models X$. Hence, $D \notin Cn_{Bolz}(X)$ – a contradiction. ' \Leftarrow ' Immediately.

The above shows that we need only a few rules to describe the entailment of nontautological sentences Φ_{CNF} from a set of clauses X. However, Modus Ponens is not a valid rule of inference, though a valid rule of proof. This means: if Φ and $\Phi \to \Psi$ are both tautological, so is Ψ , whilst we cannot conclude that Ψ is valid in a given interpretation, if both Φ and $\Phi \to \Psi$ is valid.

Since Kleene's logic has exactly the same tautologies as classical propositional logic, we can use a classical Hilbert-system to generate the set of tautologies. However, we cannot use this system to generate the consequences of a nonempty set X, i.e. we do not have $X \models_{\mathbb{H}} \Phi$ implies $X \Vdash \Phi$, where $\models_{\mathbb{H}}$ is Hilbert's classical provability relation. We shall therefore distinguish between rules of proof (for generating the set of tautologies) and rules of inference (for generating the consequences of a given set X).

Consider the following axiomatic system: **Axioms:** The set Ax of classical axioms, i.e.

1. $A \to (B \to A)$ 2. $(A \to B) \to ((A \to (B \to C)) \to (A \to C))$ 3. $A \to (B \to (A \land B))$ 4.
$$(A \land B) \rightarrow A$$

5. $(A \land B) \rightarrow B$
6. $A \rightarrow (A \lor B)$
7. $B \rightarrow (A \lor B)$
8. $(A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow (A \lor B \rightarrow C))$
9. $(A \rightarrow B) \rightarrow ((A \rightarrow \neg B) \rightarrow \neg A)$
10. $\neg \neg A \rightarrow A$

Rules of Proof:

$$\frac{\Phi \quad \Phi \to \Psi}{\Psi} \ (MP)$$

Rules of Inference:

We shall write $X \models_{\text{INF}} \Phi$ if there is a proof of Φ from X which uses only inference rules.

Theorem 4.2. Let X be a set of clauses, Φ a nontautological formula in CNF. $X \models_{\text{INF}} \Phi$ if and only of $\Phi \in \text{Cn}_{\text{Bolz}}(X)$.

Proof. It is easy to check that the above rules of inference are sound. To see that they are also complete, we have to show that we are able to prove each conjunct C_i of Φ with help of the above rules. I.e. we have to show that $X \parallel \vdash C_i$ implies $X \models_{\text{INF}} C_i$, where C_i is a disjunction of literals. By Proposition 4.2 we know that C_i is an element of X or there is a clause $C_0 \in X$ such that $C_0 \subseteq C_i$ (if we treat the disjunction C_i of literals as a set). We can apply $(\vee -I)$ to obtain C_i .

After we have generated all the C_i as described above, we use $(\land -I)$ to generate Φ .

In order to prove for arbitrary sets X and arbitrary formulas Φ whether $\Phi \in Cn_{\text{Bolz}}(X)$ holds, we need to transform X into clausal form and Φ into CNF. That is, for an arbitrary φ we have to generate an equivalent formula in CNF. Transforming φ in CNF can be done with the following rules:

$$\frac{\Phi \to \Psi}{\neg \Phi \lor \Psi} (MI) \qquad \qquad \frac{\neg \neg \Phi}{\Phi} (DN)$$
$$\frac{\neg (\Phi \land \Psi)}{\neg \Phi \lor \neg \Psi} (DM) \qquad \qquad \frac{\Phi \lor (\Psi \land \Theta)}{(\Phi \lor \Psi) \land (\Phi \lor \Theta)} (DIS)$$

We wish to add these rules to the inference rules of the above proof-system. We shall write $X \models_{\text{Bolz}} \Phi$ if there is a proof of Φ from X using only inference rules from the above system.

Theorem 4.3 (Soundness and Completeness). Let Φ be a nontautological formula. $X \models_{\text{Bolg}} \Phi$ if and only if $\Phi \in \text{Cn}_{\text{Bolg}}(X)$.

Proof. It is easy to see that the rules MI (Material Implication), DN (Double Negation), DM (De Morgan), DIS (Distributivity) are sound. Moreover the given rules are sufficient to transform any formula into CNF (cf.[Bergmann and Noll, 1977], p. 75). Hence, we can transform X into clausal normal form and Φ in CNF. By Theorem 4.2 we have that the above proof-system is complete.

4.2 A Hilbert Style Approximation for K₃

We have mentioned in Chapter 2, that, due to the non-normality of the implication, Modus Ponens (MP) is not a valid rule of inference. We shall now explain how to weaken Modus Ponens in order to become valid; the modified version of Modus Ponens will, together with a set of axioms, yield a syntactical characterisation of the semantical entailment operator Cn_3 .

4.2.1 Context-dependent reasoning

The syntactical characterisation of a nonmonotonic entailment relation is not as straightforward as its monotonic counterpart. This because the applicability of some inference rules does not only depend on what has been inferred, but on some context-dependent conditions. For example, in the very beginning of our analysis of paraconsistent reasoning we said that we wish that Ψ should be inferred from $\Phi, \Phi \to \Psi$ provided that there is no doubt about Φ 's real true-status. This means that we have to guarantee at stage *i* the deduction process that Φ will not turn out to be doubtful at any stage *j*, *j* > *i*. Marek and Truszczyński describe this behaviour as follows [Marek and Truszczyński, 1993]:

We use the term 'jumping to conclusions' to denote the following inference technique: if there is no evidence that would contradict φ , conclude φ . It should be quite clear that 'jumping to conclusions' is not monotonic. [...] This mode of reasoning² is based on the concept of contextdependent proof or derivation. The idea is to relativize the concept of a proof using a context to control the applicability of rules.

We shall model the context by using the device PC(...). An inference rule, thus has the following general form:

$$\frac{X;\varphi_1,\ldots,\varphi_n:\mathsf{PC}(X,\varphi_1,\ldots,\varphi_n)}{\varphi_n}$$

A nonmonotonic inference rule is *valid*, if $\varphi_1, \ldots, \varphi_n$ is valid in every preferred model of X and $\mathsf{PC}(X, \varphi_1, \ldots, \varphi_n)$ holds implies that φ holds in every preferred model of X.

Definition 4.2. Given a set X of formulas, a set \mathcal{R} of nonmonotonic inference rules, we define the set of nonmonotonic consequences, $C_{\mathcal{R}}(X)$, to be the smallest set which contains X and which is closed under the following condition: if $\frac{X;\varphi_1,\ldots,\varphi_n:\mathsf{PC}(X,\varphi_1,\ldots,\varphi_n)}{\varphi} \in \mathcal{R}, \varphi_1,\ldots,\varphi_n \in C_{\mathcal{R}}(X)$ and $\mathsf{PC}(X,\varphi_1,\ldots,\varphi_n)$ then $\varphi \in C_{\mathcal{R}}(X)$.

The above definition coincides with the usual inductive definition of topological operators, i.e.

$$\begin{array}{l} X_{0} =_{\scriptscriptstyle def} X \\ X_{i} =_{\scriptscriptstyle def} X_{i-1} \cup \{ \varphi \mid \frac{X; \varphi_{1}, \dots, \varphi_{n}: \mathsf{PC}(X, \varphi_{1}, \dots, \varphi_{n})}{\varphi} \in \mathcal{R} \text{ and} \\ \varphi_{1}, \dots, \varphi_{n} \in X_{i-1} \text{ and } \mathsf{PC}(\dots) \text{ holds} \end{array} \right\}$$

We have

$$\bigcup_{i < \omega} X_i = C_{\mathcal{R}}(X).$$

The Hilbert-style and Gentzen-style calculi to be presented in this chapter will base on context-dependent inference rules. For the Gentzen-style calculus we shall modify only the Modus Ponens rule. This modification is similar to the restriction of the deduction theorem and Right Weakening. As the discussion on the deduction theorem and Right Weakening shows, it is possible to restrict both properties such that their restricted variant holds in K_3 . In both cases the restriction was to consider only those formulas of the type $\Phi \to \Psi$ for which the normalisation condition (NC) holds.

² [Batens, 1997] calls this dynamic proof-theory.

4.2.2 Restricted Modus Ponens

Let us start with restricting Modus Ponens by an extra condition (PC(...)). The resulting rule is called *restricted Modus Ponens* (RMP).

$$\frac{X; \varPhi, \varPhi \to \Psi: \mathsf{PC}(\varPhi \to \Psi)}{\Psi} \qquad \mathbf{RMP}$$

The precondition $\mathsf{PC}(\Phi \to \Psi)$ means that $\Phi \to \Psi \in \mathsf{Taut}_{\mathsf{NC}}$ i.e. $\Phi \to \Psi$ is a tautology which satisfies the normalisation condition. Thus RMP is restricted to formulas $\Phi \to \Psi$ which are not satisfied if and only if the antecedent takes a designated value and the succedent takes a non-designated one. Note, that this precondition is not context-dependent, since it does not refer to X. Hence, RMP lies between classical MP and nonmonotonic inference rules described in Section 4.2.1. In Section 4.2.4 we shall present a context-dependent variant of RMP.

As the next proposition states, RMP is not as restrictive as it might seem. Together with a set Ax of classical axioms, RMP is able to generate all classical tautologies, and therefore all K_3 tautologies. Let Ax be the set of classical axioms.

Proposition 4.3. Ax and RMP generate all K_3 -tautologies.

Proof. Assume to the contrary that there is a tautology Ψ such that Ψ is not generated by Ax and RMP. Since Ψ is a tautology, it can be generated by using MP instead of RMP. Thus there is a derivation $\Phi, \Phi \to \Psi$ of Ψ , where $\Phi \to \Psi$ does not satisfy the normalisation condition. Hence, there is a preferred model I of $\{\Phi, \neg \Phi\}$, which satisfies $\neg \Psi$. Hence, $\{\Phi, \neg \Phi\} \not\Vdash \Psi$. Therefore, Ψ cannot be a tautology – a contradiction.

4.2.3 Relatively Consistent Modus Ponens

Even though RMP generates together with Ax the set of all tautologies, it is too restrictive to generate the set $Cn_3(X)$ when given the set $X \cup Ax$. For example, it is impossible to generate B from $\{A, A \to B\}$ only by using RMP.

We are in a little dilemma here: RMP is too restrictive (precisely: $\mathsf{RMP}(X) \subseteq Cn_3(X)$) and MP is explosive ($Cn_3(X) \subseteq \mathsf{MP}(X)$, where $\mathsf{RMP}(X)$ is the smallest set containing $\mathsf{Ax} \cup \{X\}$ and is closed under the rule RMP; $\mathsf{MP}(X)$ is defined analogously). The solution which resolves the dilemma is to invoke RMP only in those cases where MP is too explosive. MP becomes invalid, when applied to

formulas $\Phi, \Phi \to \Psi$ where Φ is inconsistent w.r.t. the given set X. We shall first clarify what we mean by ' Φ is inconsistent w.r.t. X'.

Let X be a set of formulas in clausal normal form. A literal L is consistent w.r.t. X iff and only if there is no preferred model I of X such that $I(L) = \top$. A clause C is consistent w.r.t. X iff Φ contains at least one literal which is consistent w.r.t. X. A formula $\Phi = C_1 \land \ldots \land C_n$ in CNF is consistent w.r.t. X iff every clause C_i is consistent w.r.t. X.

For any arbitrary set X of formulas and any arbitrary Φ we say that Φ is consistent w.r.t. X if there is a set X' in clausal normal form, X and X' semantically equivalent such that $\Phi_{\rm CNF}$ is consistent w.r.t. X' where $\Phi_{\rm CNF}$ is a formula in conjunctive normal form which is semantically equivalent to Φ . Instead of saying that ' Φ is consistent w.r.t. X' we shall also say that ' Φ and X are relatively consistent'.

The concept of relativised consistency expresses that a formula Φ does not cause a new inconsistency when added to a set X. To illustrate this concept, consider the following example.

Example 4.1. Let $X = \{A, B, C, \neg C\}$.

- 1. $\neg A \lor \neg B$ is inconsistent w.r.t. X.
- 2. $C \lor D$ is consistent w.r.t. X
- 3. $\neg A \lor D$ is consistent w.r.t. $X \cup \{\neg A \lor \neg B\}$.

We strengthen the concept of relative consistency a little bit.

Definition 4.3 (Strongly Relatively Consistent). A formula Φ is *strongly* relatively consistent w.r.t. X if and only if Φ takes the value t in every preferred model I of $X \cup \{\Phi\}$.

Proposition 4.4. If Φ is strongly relatively consistent w.r.t. X then Φ is relatively consistent w.r.t. X.

Proof. Without loss of generality we can assume that Φ is in CNF. Since Φ is assigned t in every preferred model I of $X \cup \{\Phi\}$, we have by Proposition 2.3 that Φ takes a value from $\{t, f\}$ in every preferred model of X. By the same argument, each a conjunct D of Φ takes a value from $\{t, f\}$ in any preferred model of X. It follows that for every preferred model I of X there is a literal L appearing as a disjunct in D such that $I(L) \in \{t, f\}$. Hence, every D is relatively consistent w.r.t. X and therefore Φ is relatively consistent w.r.t. X.

The converse of Proposition 4.4 does not hold. For example, $A \lor B$ is consistent w.r.t. $\{A, \neg A\}$ but not strongly consistent. This reflects our intuition: $A \lor B$ does not cause a new inconsistency when added to X.

With help of the above concepts we can formulate the following sound rule of inference (the name 'SRCMP' stand for 'Strongly Relatively Consistent MP')

$$\frac{X; \Phi, \Phi \to \Psi : \mathsf{PC}(X, \Phi)}{\Psi} (\mathbf{SRCMP})$$

where the precondition $\mathsf{PC}(X, \varPhi)$ means: \varPhi is strongly relatively consistent w.r.t. X.

Definition 4.4 (\mathcal{R} , \mathcal{R} -provable). Let \mathcal{R} denote the system [Ax, (RMP, SRCMP)], i.e. \mathcal{R} is a Hilbert-style system with Ax as the set of axioms and RMP and SR-CMP as rules of inference. We say that Φ is \mathcal{R} -provable from X if Φ can be generated from $X \cup Ax$ with help of RMP and SRCMP.

Proposition 4.5. \mathcal{R} is embedding and idempotent.

Proof. Immediately by definition of \mathcal{R} .

Proposition 4.6 (Soundness of \mathcal{R}). If Ψ is \mathcal{R} -provable from X, then $\Psi \in Cn_3(X)$.

Proof. It is sufficient to show that the application of RMP and SRCMP is sound. For RMP the soundness follows directly from the Normalised Deduction Theorem. Thus, we need only to consider SRCMP. Since Φ and X are strongly relatively consistent, we know that Φ takes the value t in every preferred model of $X \cup \{\Phi\}$. But since $X \Vdash \Phi$ we also have that Φ takes the value t in every preferred model of X: for suppose there is a preferred model I of X such that $I(\Phi) = \top$. By Lemma 2.1 we have that I is also a preferred model of $X \cup \{\Phi\}$ in which Φ takes the value \top , which contradicts the requirement that Φ and X are relatively consistent. Hence, Φ takes the value t in every preferred model X. By $X \Vdash \Phi \rightarrow \Psi, \Psi$ must take the value t or \top in every preferred model of X. Thus, $\Psi \in Cn_3(X)$.

It is an open question whether the system \mathcal{R} is also complete with respect to finite sets X. But \mathcal{R} might be a r.e. approximation of Cn_3 . We shall investigate if this the case. In order to be a r.e. approximation, the set

 $\mathcal{R}(X) =_{def} \{ \Phi \mid \Phi \text{ is } \mathcal{R}\text{-provable from } X \}$

must be r.e., whenever X is r.e. This means, that the precondition in SRCMP must be decidable. We can force the precondition to be decidable, if we restrict ourselves to finite sets X of premisses. On the other hand, \mathcal{R} must be idempotent. However, if we restrict ourselves to finite sets X, we have to clarify what idempotency should mean in this context.

A consequence relation is said to be *finitely idempotent* if and only if for any finite X, Y with $X \subseteq Y \subseteq Z$:

$$X \vdash Y$$
 and $Y \vdash Z \implies X \vdash Z$.

Note that with $Cn(X) = \{ \Phi \mid X \vdash \Phi \}$ we have for any finite Y such that $X \subseteq Y \subset Cn(X)$:

 $\mathsf{Cn}(X) = \mathsf{Cn}(Y).$

Proposition 4.7. Let Cn be an embedding consequence relation. Cn is finitely idempotent if and only if Cn is cumulative.

Proof.

- '⇒' Let Cn(X) = {Φ | X ⊢ Φ} be finitely idempotent. Let φ be a formula such that $X \vdash \varphi$. Since Cn is embedding, ⊢ is reflexive and we have $X \vdash X \cup {\varphi}$ and by finite idempotence we have for any Φ such that $X \vdash Φ$: Cn(X) = Cn(X ∪ {\varphi}). Hence, $X ∪ {\varphi} \vdash Φ$. But this means that $X \vdash φ$ and $X \vdash Φ$ implies $X ∪ {\varphi} \vdash Φ$, which is nothing other than cumulativity.
- '⇐' We have to show that if Cn is cumulative then $Cn(X) = Cn(X \cup \{\Phi\})$, for any $\Phi \in Cn(X)$. Let $\Phi \in Cn(X)$. By cumulativity we have

$$\{\varphi \mid X \vdash \varphi\} = \{\varphi \mid X \cup \{\Phi\} \vdash \varphi\}$$

and hence

$$\mathsf{Cn}(X) = \mathsf{Cn}(X \cup \{\Phi\}).$$

Definition 4.5 (Finite R.E. Approximation). Let Cons be an arbitrary consequence operator. We say that Cons is a finite r.e. approximation of Cn, if Cons satisfies all requirements for a r.e. approximation except that idempotency is replaced by finite idempotency, i.e. if for any finite Y with $X \subseteq Y \subset Cons(X)$ we have

$$\operatorname{Cons}(X) = \operatorname{Cons}(Y).$$

Theorem 4.4. Let X be finite. The system $\mathcal{R} = [Ax, (RMP, SRCMP)]$ is a finite r.e. approximation of Cn_3 .

Proof. Let

$$\mathcal{R}(X) = \{ \Phi \mid \Phi \text{ is } \mathcal{R}\text{-provable from } X \}.$$

Inclusion. By definition of \mathcal{R} .

- **Nonmonotonicity.** To see that the operator \mathcal{R} is nonmonotonic consider the following counter-example to monotonicity: $X = \{A, A \to B\}$. We can apply SRCMP and thus B is \mathcal{R} -provable from X. If we extend X by $\neg A$ we can no longer apply SRCMP since A is not strongly consistent w.r.t. $X' = X \cup \{\neg A\}$. Because RMP is too weak to generate B from X, we have $B \notin \mathcal{R}(X')$. Hence, \mathcal{R} is nonmonotonic.
- **Cumulativity.** We have to show that if Φ and Ψ are both \mathcal{R} -provable from X, then Ψ is \mathcal{R} -provable from $X \cup \{\Phi\}$. From Proposition 4.6 we know that $\Phi \in Cn_3(X)$. Therefore, by Lemma 2.1 we have $\mathsf{PMOD}(X) = \mathsf{PMOD}(X \cup \{\Phi\})$. Hence, a formula φ is strongly consistent w.r.t. X if and only φ is strongly consistent w.r.t. X if and only φ is strongly consistent w.r.t. $X \cup \{\Phi\}$.
- **Recursively enumerable.** Since we restricted ourselves to finite sets X, the validity of the preconditions in each rule is recursive. Hence $\mathcal{R}(X)$ is r.e. $\mathcal{R}(X) \subseteq Cn_3(X)$. Follows from Proposition 4.6.

Finite Idempotence. By equivalence to cumulativity (Proposition 4.7).

Conservativity. Assume that X has a two-valued model. Hence, no atomic formula takes the value \top in any preferred K₃-model of X. In this case any application of SRCMP to X is identical to an application of Modus Ponens to X. Hence, $\mathcal{R}(X)$ coincides with $Cn_{cl}(X)$.

The following proposition shows that \mathcal{R} generates at least all consequences of the monotonic version of Kleene's logic.

Proposition 4.8 (Supradeductivity). $Cn_{Bolz}(X) \subseteq \mathcal{R}(X)$.

Proof. By Theorem 4.2 we have only to show that we are able to simulate the effects of the rules (\lor -I), (\land -I) and (\land -E). Since the axioms $\Phi \to (\Phi \lor \Psi)$, ($\Phi \land \Psi$) $\to \Phi$ and ($\Phi \land \Psi$) $\to \Psi$ are all in Taut_{NC} we can simulate (\land -I) and (\land -E) with help of RMP. To see that we can also handle (\land -I), note that the axiom $A \to (B \to A \land B)$

is in $Taut_{NC}$. Consider the following RMP-proof:

$$\frac{\Psi}{\frac{\Psi \rightarrow \Psi \rightarrow \Psi \rightarrow \Psi}{\Psi \rightarrow \Phi \wedge \Psi}} (\text{RMP})}{\frac{\Phi \rightarrow \Psi}{\Phi \wedge \Psi}} (\text{RMP})$$

As already mentioned in Section 2.2.4, congruence plays an important role in the theory of inference frames. Recall that Cn_{Bolz} denotes the monotonic operator which was syntactically characterised in Section 4.1. We shall not only show that \mathcal{R} and Cn_{Bolz} are congruent but that Cn_{Bolz} is a deductive basis of \mathcal{R} .

Proposition 4.9. Cn_{Bolz} is a deductive basis of \mathcal{R} , i.e.

- 1. for all X we have $\operatorname{Cn}_{\operatorname{Bolz}}(X) \subseteq \mathcal{R}(X)$
- 2. for all X we have $\operatorname{Cn}_{\operatorname{Bolz}}(\mathcal{R}(X)) = \mathcal{R}(X)$
- 3. for all X, Y we have $\operatorname{Cn}_{\operatorname{Bolz}}(X) = \operatorname{Cn}_{\operatorname{Bolz}}(Y)$ implies $\mathcal{R}(X) = \mathcal{R}(Y)$.

Proof. The first two points follow from supradeductivity (Proposition 4.8). For the third point, congruence, we first show that for arbitrary Z we have

$$\mathcal{R}(Cn_{\text{Bolz}}(Z)) = \mathcal{R}(Z). \tag{(*)}$$

For the inclusion $\mathcal{R}(Cn_{\text{Bolz}}(Z)) \subseteq \mathcal{R}(Z)$, we know that $Z \subseteq Cn_{\text{Bolz}}(Z) \subseteq \mathcal{R}(Z)$. Because \mathcal{R} is cumulative we have $\mathcal{R}(Z) \subseteq \mathcal{R}(Cn_{\text{Bolz}}(Z) \cup Z)$ and since Cn_{Bolz} is embedding we get $\mathcal{R}(Z) \subseteq \mathcal{R}(Cn_{\text{Bolz}}(Z))$. The converse direction does also follow from supradeductivity (Proposition 4.8) and idempotency (Proposition 4.5) of \mathcal{R} . Hence \bigstar .

Since $Cn_{Bolz}(X) = Cn_{Bolz}(Y)$ we have $\mathcal{R}(Cn_{Bolz}(X)) = \mathcal{R}(Cn_{Bolz}(Y))$. It follows from \star that $\mathcal{R}(X) = \mathcal{R}(Y)$.

4.2.4 Relativised Normalisation

The concept of an implicational tautology which satisfies the normalisation condition can be generalised as follows:

Definition 4.6 (Relativised Normalisation). Let X be a set of formulas. $\Phi \to \Psi$ is said to satisfy the normalisation condition (NC) w.r.t. X if and only if for every preferred model I of X we have:

$$I \not\models \Phi \to \Psi \Leftrightarrow I \models \Phi \ , \ I \not\models \Psi.$$

The set of all implicational formulas which satisfy NC w.r.t. X is denoted by NC(X). Clearly, for any X we have $Taut_{NC} \subseteq NC(X)$. Moreover, $Taut_{NC} = NC(\emptyset)$.

Consider the following inference rule, RNMP - Relativised Normalisation MP:

$$\frac{X; \varPhi, \varPhi \to \Psi : \mathsf{PC}(X, \varPhi \to \Psi)}{\Psi} \left(\mathbf{RNMP} \right)$$

where $\mathsf{PC}(X, \Phi \to \Psi)$ means: $\Phi \to \Psi \in \mathsf{NC}(X)$.

Proposition 4.10. RNMP is sound.

Proof. Since Φ is valid in every preferred model of X, we know that Φ takes a designated truth-value in every preferred model of X. Since $\Phi \to \Psi \in \mathsf{NC}(X)$ we also know that Ψ takes a designated value in every preferred model of X. Thus, $X \models \Psi$.

With help of the rule RNMP, we can define another Hilbert-style inference system:

Definition 4.7 (\mathcal{R}_{NC} , \mathcal{R}_{NC} -provable). Denote by \mathcal{R}_{NC} the system [Ax, RNMP]. We say that Φ is \mathcal{R}_{NC} -provable from X if and only if Φ can be generated from $X \cup Ax$ by means of RNMP.

Proposition 4.11. If Ψ is \mathcal{R} -provable from X then Ψ is \mathcal{R}_{NC} -provable from X.

Proof. By induction on the length l of the proof. If l = 0 we have that $\Phi \in X$. Assume that l = k + 1. We have two cases:

- Ψ is generated by RMP. We know that there are formulas $\Phi, \Phi \to \Psi$ with a proof-length of at most k. Since $\mathsf{Taut}_{\mathsf{NC}} \subseteq \mathsf{NC}(X)$ we have that $\Phi \to \Psi \in \mathsf{NC}(X)$. Hence, RNMP is applicable and Ψ is $\mathcal{R}_{\mathsf{NC}}$ -provable from X.
- Ψ is generated by SRCMP. We know that there are formulas $\Phi, \Phi \to \Psi$ with a proof-length of at most k. By the precondition of SRCMP, we know that Φ takes the value t in every preferred model of $X \cup \{\Phi\}$. Since Φ is \mathcal{R} -provable we know by soundness of \mathcal{R} that Φ takes the value t in every preferred model of X. Moreover $\Phi \to \Psi$ is \mathcal{R} -provable. Hence, Ψ takes a value from $\{t, \top\}$ in every preferred model of X. Therefore, $\Phi \to \Psi \in \mathsf{NC}(X)$.

Proposition 4.12 (Supradeductivity). $\operatorname{Cn}_{\operatorname{Bolz}}(X) \subseteq \mathcal{R}_{\operatorname{NC}}(X)$.

Proof. By $Cn_{Bolz}(X) \subseteq \mathcal{R}(X)$ (Proposition 4.8) and $\mathcal{R}(X) \subseteq \mathcal{R}_{NC}(X)$ (Proposition 4.11).

The above propositions show that the rule RNMP is indeed very powerful. This is due to the extremely strong precondition. Of course, there is no general restriction on the type of preconditions that should be allowed for syntactical characterisations and which type of characterisation contains too much semantical information.

We close our small discussion on RNMP with the following theorem.

Theorem 4.5. Let X be finite. The system $\mathcal{R}_{NC} = [Ax, RNMP]$ is finite r.e. approximation of Cn_3 .

Proof. Let

 $\mathcal{R}_{\mathsf{NC}}(X) = \{ \Phi \mid \Phi \text{ is } \mathcal{R}_{\mathsf{NC}} \text{-provable from } X \}.$

The inclusion $X \subseteq \mathcal{R}_{\mathsf{NC}}(X)$ is immediate.

Nonmonotonicity. By our standard counter-example: let $X = \{A, A \rightarrow B\}$ from which B is \mathcal{R}_{NC} -provable. However, we have $B \notin \mathcal{R}_{NC}(X \cup \{\neg A\})$.

Cumulativity. Assume that $\Phi \in \mathcal{R}_{NC}(X)$. We show that every Ψ which is derivable from X at stage *i* is derivable from $X \cup \{\Phi\}$ at stage *i*. If i = 0 we have $\Psi \in X$ and hence, $\Phi \in X \cup \{\Phi\}$. For i > 0 we have to show that

$$\varphi \to \Psi \in \mathsf{NC}(X) \text{ implies } \varphi \to \Psi \in \mathsf{NC}(X \cup \{\Phi\}).$$

Thus, it suffices to show that $\mathsf{PMOD}(X) = \mathsf{PMOD}(X \cup \{\Phi\})$. But since RNMP is sound we have by $\Phi \in \mathcal{R}_{\mathsf{NC}}(X)$ that $X \Vdash \Phi$. Hence, by Lemma 2.1 we have $\mathsf{PMOD}(X) = \mathsf{PMOD}(X \cup \{\Phi\})$.

Recursively enumerable. By finiteness of X.

 $\mathcal{R}_{NC}(X) \subseteq Cn_3(X)$. By soundness of RNMP.

Finite Idempotence. By cumulativity.

Conservativity. We have to show: if X has a two-valued model, then $Cn_{cl}(X) = \mathcal{R}_{NC}(X)$. Since $\Phi \in Cn_{cl}(X)$ we know that there is a classical MP proof of Φ . We shall show that RNMP can simulate the effect of MP. Consider an application of MP

$$\frac{\varphi, \varphi \to \Psi}{\Psi} (\mathrm{MP})$$

By the consistency of X we know that the set $\mathsf{PMOD}(X)$ coincides with the set of classical two-valued model of X. Thus, $\varphi \to \Psi \in \mathsf{NC}(X)$.

4.2.5 Paraconsistent Extensions

It is quite convenient to think of a nonmonotonic reasoner as an agent who has several beliefs about the world. In our case, the agent has some beliefs about what he or she knows consistently and what might be contradictory. Sometimes it is quite clear which information is consistent and which is not. For example, when given

$$\{A, \neg A, B, B \rightarrow C\}$$

the agent believes that the information about A is contradictory, whereas the information on C is not. However, when given

$$X = \{A, B, A \to \neg B, B \to \neg A\}$$

the agent has no clear idea whether the information A is non-contradictory or not. If our agent minimises the contradictory information, he or she could either believe that A is not contradictory and B is, or that B is not contradictory and A is. Thus, the above information gives raise to attribute two plausible epistemic states to the agent.

Suppose the agent infers additional information on the basis of the following rule: if you know $\Phi, \Phi \to \Psi$ and you do not believe that $\neg \Phi$, then infer Ψ . Let us write this rule in a first step as:

$$\frac{X \succ \Phi, \Phi \to \Psi : \mathsf{PC}(X \not \succ \neg \Phi)}{X \succ \Psi} \,(\mathbf{MP^*})$$

Do not be confused by the sloppy presentation of the above rule; especially not by the reference to something like non-deducibility. It should just serve as an illustration. Suppose the agent acts on the basis of MP* and has been given the above set X. This yields two different paraconsistent epistemic states:

State 1 He or she can ap-	State 2 He or she can ap-
ply MP* to $A, A \rightarrow \neg B$.	ply MP* to $B, B \rightarrow \neg A$.
Thus, B becomes contra-	Thus, A becomes contra-
dictory and we cannot ap-	dictory and we cannot ap-
ply MP* to $B, B \to \neg A$.	ply MP* to $A, A \to \neg B$.

When constructing an epistemic state, the agent does only refer to what he already knows but also to what he or she does not know. E.g. when constructing State 1 the agent assumes that A is consistently known.

Instead of characterising the construction of an epistemic state E by something like non-deducibility, we refer to E itself and reformulate MP* as follows:

$$\frac{(X,E) \succ \Phi, \Phi \to \Psi : \mathsf{PC}(\Phi,E)}{(X,E) \succ \Psi} (\mathbf{MP}^*)$$

where $\mathsf{PC}(\Phi, E)$ means: $\Phi \notin E$.

This is very similar to Reiter's Default Logic (cf. [Brewka, 1991]). Like in Default Logic, we only have a so-called quasi-inductive fixpoint definition of an extension. Technically, this means that we (nondeterministically) guess a set E and verify whether E is an extension.

Definition 4.8 (Paraconsistent Extension). Let X be a set of formulas. A set $E \supseteq X$ is said to be a *paraconsistent extension* of X if and only if $E = \Gamma(X, E)$ where

$$\begin{split} &\Gamma_0(X,E) =_{\scriptscriptstyle def} X \cup \mathsf{Ax} \\ &\Gamma_i(X,E) =_{\scriptscriptstyle def} \Gamma_{i-1}(X,E) \cup \{ \Psi \mid \varPhi, \varPhi \to \Psi \in \varGamma_{i-1}(X) \text{ and } \text{ if } \neg \varPhi \in E \text{ then} \\ & \varPhi \to \Psi \in \mathsf{Taut}_{\mathsf{NC}} \} \\ &\Gamma(X,E) =_{\scriptscriptstyle def} \bigcup_{n=0}^\infty \varGamma_n(X,E) \end{split}$$

The inductive step corresponds to the application of the following rule, which is in fact a combination of RMP and MP*. This rule is called *Cautious Modus Ponens*:

$$\frac{(X,E);\Phi,\Phi\to\Psi:\mathsf{PC}(E,\Phi,\Phi\to\Psi)}{\Psi}(\mathbf{CMP})$$

where $\mathsf{PC}(E, \Phi, \Phi \to \Psi)$ means: $\neg \Phi \in E$ implies $\Phi \to \Psi \in \mathsf{Taut}_{\mathsf{NC}}$ Let us illustrate how the above fixpoint construction works:

Example 4.2. Consider again $X = \{A, A \rightarrow B, A \rightarrow \neg B\}$. We have just seen that the set of preferred models of X can be partitioned into two sets each judging

different formulas to be paraconsistent. There are two paraconsistent extensions: $E_1 = X \cup \{A, \neg A, A \land \neg A \dots\}$ and $E_2 = X \cup \{A, B, \neg B, \dots\}$. Moreover, $B, \neg B \notin E_1$ and $\neg A \notin E_2$.

It is easy to see that $B, \neg B \in \Gamma(X, E_2)$: since $\neg A \notin E_2$ we can infer B from $A, A \to B$. After contraposing $B \to \neg A$ we can infer $\neg B$ from $A, A \to \neg B$.

To see that $A, \neg A \in \Gamma(X, E_1)$ is more difficult: we have to show how $\neg A$ can be derived, i.e. we have to show that there is $\Gamma(X, E_1)$ which contains $\neg A$. One possibility how $\neg A$ can enter $\Gamma(X, E_1)$ is by the fact that $(A \to B) \land (B \to \neg A) \to \neg A$ is a tautology. Unfortunately, this tautology does not satisfy the normalisation condition, so we have to show that the negation of the antecedent, i.e. $\neg((A \to B) \land (B \to A)))$, is not contained in E_1 . The negation of the antecedent is semantically equivalent to $(A \land \neg B) \lor (B \land A)$. But since neither B nor $\neg B$ is contained in E_1 . Further, there is $\Gamma_{i-1}(X, E_1)$ which contains X and the above tautology. Hence, $\neg A \in \Gamma_i(X, E_1)$.

Proposition 4.13. Let $I \models X$ and $E =_{def} \{ \Phi \mid I \models \Phi \}$. Then, $\Gamma(X, E) \subseteq E$.

Proof. We show by induction on the construction of $\Gamma(X, E)$ that for all i we have

$$\Gamma_i(X, E) \subseteq E.$$

For i = 0 the proposition is immediately clear. Suppose that the proposition holds for i = n. Consider $\Phi \in \Gamma_{n+1}(X, E)$. If Φ was already a member of $\Gamma_i(X, E)$ we are done. Suppose $\Phi \notin \Gamma_i(X, E)$. If Ψ has been inserted by applying RMP we are done. Assume that Ψ has been inserted by MP*. Since $\Phi \in E$ and $\neg \Phi \notin E$ we have that $I(\Phi) = t$. Thus, $I(\Psi) = t$. This proves the soundness of MP* and therefore the soundness of CMP.

4.3 A Hilbert-Style Approximation for FOK₃

The preclosure operator Cn_3 of propositional K_3 can be syntactically approximated by means of $\mathcal{R} = [Ax, (RMP, SRCMP)]$. As we have seen in Corollary 3.1 FOK₃ has the same tautologies as classical first-order logic. This gives rise to the question, whether we can just add quantifier-handling rules and axioms to \mathcal{R} in order to obtain a syntactical characterisation of FOK₃. In classical first-order logic, quantifiers are handled by the following rules:

$$rac{\Phi o \Psi}{orall x \Phi o \Psi} \mathbf{A} \ \mathbf{Gen} \qquad \qquad rac{\Phi o \Psi^{\dagger}}{\Phi o orall x \Psi} \mathbf{S} \ \mathbf{Gen}$$

The \dagger is the Eigenvariable-condition (i.e. x does not belong to the set of A's free variables). Beside the above rules we have the following rules for substituting variables:

- **V** Rep Suppose we are given Φ . If the formula Ψ results from Φ by substituting each bounded variable x in Φ by some x', then we can conclude Ψ .
- **T** Sub Suppose we are given Φ and $\Phi x/t$ is the result substituting simultaneously each occurrence of the variable x by the term t, then we can conclude Φ^x/t .

It is known that the axiom schemes Ax and the rules (A Gen, S Gen, V Sub, T Sub, MP) axiomatise classical first-order entailment. It is also easy to see that the rules A Gen, S Gen, V Rep, T Sub are sound w.r.t. FOK₃. Moreover RMP is sound w.r.t. FOK₃.

Lemma 4.1. Let \mathfrak{A} be a structure. Then

- 1. If $\mathfrak{A} \models_{\overline{K}} \Phi$ and $\Phi \to \Psi$ is a tautology which satisfies the normalisation condition, then $\mathfrak{A} \models_{\overline{K}} \Psi$.
- 2. If $\mathfrak{A} \models_{\overline{\mathbf{k}_3}} \Phi \to \Psi$ then $\mathfrak{A} \models_{\overline{\mathbf{k}_3}} \forall x \Phi \to \Psi$. 3. If $\mathfrak{A} \models_{\overline{\mathbf{k}_3}} \Phi \to \Psi$ and x does not belong to Φ 's free variables then $\mathfrak{A} \models_{\overline{\mathbf{k}_3}} \Phi \to \forall x \Psi$.
- 4. If $\mathfrak{A} \models \Phi$ and Ψ is the result of substituting one of Φ 's bounded variables x by x' then $\mathfrak{A} \models_{\overline{K}_2} \Psi$.
- 5. If $\mathfrak{A} \models_{\overline{K}_3} \Phi$ and Φ^x/t is the replacement of x by the term t, then $\mathfrak{A} \models_{\overline{K}_3} \Phi^x/t$.

Proof. Part 1 is by Theorem 3.6. Part 2-5 is immediately by definition of validity (Definition 3.2).

Proposition 4.14. The system [Ax, (RMP, A Gen, S Gen, V Rep, T Sub)] generates all classical first-order tautologies and therefore all FOK₃ tautologies.

Proof. Analogously to the propositional case (Proposition 4.3).

 $\mathcal{R}_{\mathsf{FO}}$ does not generate all FOK_3 -consequences of the set X. There is no procedure which recursively enumerates all elements of $Cn_{\kappa}(X)$, for any X. This is, like in the propositional case, due to failure of compactness. However, in the propositional case we could restrict X to be finite in order guarantee that $Cn_3(X)$ is r.e. Unfortunately, this trick does not work because there are finite sets of first-order formulas, which have FOK_3 -models but no preferred FOK_3 models. But even if we guarantee that a finite set X has a preferred model, there is no guarantee that $Cn_{\mathsf{K}}(X)$ is r.e. In order to obtain this result, we need an additional concept. Let Φ be a formula, P a unary predicate symbol. The *relativisation* of Φ to P, denoted by Φ^P , is inductively defined as follows:

$$\begin{split} \varPhi^{P} =_{\scriptscriptstyle def} P & \text{if } \varPhi \text{ is atomic} \\ (\varPhi_{1} \land \varPhi_{2})^{P} =_{\scriptscriptstyle def} \varPhi_{1}^{P} \land \varPhi_{2}^{P} \\ (\varPhi_{1} \lor \varPhi_{2})^{P} =_{\scriptscriptstyle def} \varPhi_{1}^{P} \lor \varPhi_{2}^{P} \\ (\neg \varPhi)^{P} =_{\scriptscriptstyle def} \neg (\varPhi^{P}) \\ ((\exists x) \varPhi)^{P} =_{\scriptscriptstyle def} (\exists x) (Px \land \varPhi^{P}) \\ ((\forall x) \varPhi)^{P} =_{\scriptscriptstyle def} (\forall x) (Px \to \varPhi^{P}) \end{split}$$

It is easy to see that the definition of the \exists - and the \forall -case comply with the equivalence $\exists x \Phi \equiv \neg \forall x \neg \Phi$. Literally speaking, Φ^P is interpreted as Φ with all quantifiers restricted to P. The book [Chang and Keisler, 1977] shows how to axiomatise Bernays' Set Theory with help of relativisation.

Proposition 4.15. There is a finite set X such that for every $I \models_{\overline{K_3}} X$ there is $I_0 \sqsubseteq I$, $I_0 \models_{\overline{K_3}} X$ with $\operatorname{Cn}_{\mathsf{K}}(X)$ is not recursively enumerable.

Proof. Consider Peano's formal number theory Arith, i.e. the axioms

$$\{ \begin{array}{l} \forall x \forall y \forall z [x = y \rightarrow (x = z \rightarrow y = z)], \\ \forall x \forall y [s(x) = s(y) \rightarrow x = y], \\ \forall x \forall y [x = y \rightarrow s(X) = s(y)], \\ \forall x [\neg (s(x) = 0)], \\ \forall x [\neg (s(x) = 0)], \\ \forall x [x + 0 = x], \\ \forall x \forall y [x + s(y) = s(x + y)], \\ \forall x [x \cdot 0 = 0], \\ \forall x \forall y [x \cdot s(y) = x \cdot y + x], \\ \} \end{array}$$

Let Ax(P,Q) denote the following set of axioms:

$$\{ \begin{array}{l} P0 \wedge \neg P0, \\ \forall x [Qx \leftrightarrow (Px \wedge \neg Px)], \\ \forall x [Qx \rightarrow Qs(x)], \\ Q0 \\ \} \end{array}$$

Let $T = \operatorname{Arith} \cup \operatorname{Ax}(P, Q)$. Since T is universal, it has by Theorem 3.3 a preferred model $\mathfrak{A} = [\Sigma, \mathsf{D}, \mathsf{F}, \mathsf{P}]$ such that $\mathsf{P}_Q^+ = \mathbb{N}$. Let \mathfrak{N} be a standard model of number theory. Then, for any sentence φ we have $\mathfrak{N} \models \varphi$ if and only if $\varphi^Q \in Cn_{\mathsf{K}}(T)$. Since the set $\{\Phi \mid \mathfrak{N} \models \Phi\}$ is not recursively enumerable, we have that $Cn_{\mathsf{K}}(T)$ is not r.e.

Nevertheless we can try to find nonmonotonic approximation of Cn_{K} . Consider the following rule

$$\frac{X; \Phi, \Phi \to \Psi : \mathsf{PC}(X, \Phi)}{\Psi} \operatorname{\mathbf{SRCMP}_{FO}}$$

where $\mathsf{PC}(X, \Phi)$ means: there is no preferred model \mathfrak{A} of X such that $\mathfrak{A} \models_{\overline{\mathsf{K}_3}} \Phi \wedge \neg \Phi$. Let $\mathcal{R}_{\mathsf{FO}}$ denote $[\mathsf{Ax}, (\mathsf{RMP}, \mathsf{SRCMP}_{\mathsf{FO}}, \mathsf{A} \mathsf{Gen}, \mathsf{S} \mathsf{Gen}, \mathsf{V} \mathsf{Rep}, \mathsf{T} \mathsf{Sub})]$.

Proposition 4.16 (Soundness of \mathcal{R}_{FO}). If Φ is \mathcal{R}_{FO} -provable from X, then $\Phi \in Cn_{K}(X)$.

Proof. Soundness of RMP and the quantifier rules is immediate. It thus remains to show that SRCMP_{FO} is sound. Assume that $\Phi, \Phi \to \Psi$ holds in every preferred model \mathfrak{A} of X. By the precondition it is assured that $\neg \Phi$ is invalid in every such \mathfrak{A} . Hence, $\mathfrak{A} \models_{\overline{\mathbf{k}}} \Psi$.

The question is now, whether the deductive system \mathcal{R}_{FO} is a r.e. approximation of Cn_{K} . The properties of Nonmonotonicity, Inclusion, Idempotence are immediate. The most important question is whether \mathcal{R}_{FO} is r.e. The rule SRCMP_{FO} requires that we check whether no preferred model of X satisfies $\Phi \wedge \neg \Phi$. However, the set of formulas, which are not paraconsistent in any preferred model \mathfrak{A} of a given set X are not recursively enumerable. For example if X is empty this set coincides with the satisfiable formulas of first-order logic. This set is known not to be recursively enumerable (otherwise, first-order logic would be decidable).

Theorem 4.6. The set of formulas which are \mathcal{R}_{FO} -provable is not r.e.

Proof. By the above discussion.

4.4 Tableaux Systems

The Hilbert-style characterisations are not well suited for a mechanisation of answering the question whether Φ can be inferred from X. We shall now discuss a constructive method to decide whether Φ is entailed by X, or not. This method is called *tableau method*, or *tableau procedure* and has gained enormous popularity during the last decades. Tableaux systems were independently discovered in the mid-fifties by Beth, Hintikka and Schütte.³

A tableau system is a refutation system, i.e. in order to prove that Φ follows from X, we show systematically that there is no (preferred) model of X in which Φ takes the value f.

Even though the idea is quite simple, it requires that we introduce a new type of formula as well as a concept of validity for these formulas.

Definition 4.9 (Signed Formula, Tableau). Let $\Phi \in \mathcal{L}$ be a formula. A signed formula is an expression of the form Φ :t, Φ :f or Φ :T. A tableau is a tree whose nodes are signed formulas.

The intended meaning of Φ : **t** is that ' Φ is true' (Φ : **f** and Φ : **T** correspondingly). We shall decompose a signed formula like $(A \land B)$: **t** by a *tableau decomposition rule* in order to determine the possible truth-values of its subformulas A and B. We have two kinds of tableau decomposition rules: α -rules and β -rules which have the following form:

$$\frac{\alpha}{\alpha_1, \alpha_2} \qquad \qquad \frac{\beta}{\beta_1 \parallel \beta_2}$$

An α -rule extends the tableau by adding two nodes α_1 , α_2 without generating a new branch. A β -rule extends a tableau by generating a branch containing β_1 and another branch containing β_2 .

³ For classical logic, the corresponding tableaux systems can be seen as a variant of the Sequent Calculus presented by Gentzen. Like the resolution method, tableaux systems can be understood 'as attempts to exploit the power of CUT-elimination theorems in Gentzen-type calculi' ([Avron, 1993]).

We have the following set R of tableau decomposition rules.

$$\frac{(\varPhi \land \Psi) : \mathbf{t}}{\varPhi : \mathbf{t}, \, \Psi : \mathbf{t}} \, (\land \mathbf{t}) \qquad \qquad \frac{(\varPhi \land \Psi) : \mathbf{f}}{\varPhi : \mathbf{f} \, \| \, \Psi : \mathbf{f}} \, (\land \mathbf{f})$$

$$\frac{(\varPhi \lor \varPsi) {:} \mathbf{t}}{\varPhi {:} \mathbf{t} \, \| \, \varPsi {:} \mathbf{t}} \, (\lor \mathbf{t}) \qquad \qquad \frac{(\varPhi \lor \varPsi) {:} \mathbf{f}}{\varPhi {:} \mathbf{f}, \, \varPsi {:} \mathbf{f}} \, (\lor \mathbf{f})$$

$$\frac{(\varPhi \to \Psi): \mathbf{t}}{\varPhi: \mathbf{f} \parallel \Psi: \mathbf{t}} \left(\to \mathbf{t} \right) \qquad \qquad \frac{(\varPhi \to \Psi): \mathbf{f}}{\varPhi: \mathbf{t}, \Psi: \mathbf{f}} \left(\to \mathbf{f} \right)$$

$$\frac{\neg \Phi: \mathbf{t}}{\Phi: \mathbf{f}} (\neg \mathbf{t}) \qquad \qquad \frac{\neg \Phi: \mathbf{f}}{\Phi: \mathbf{t}} (\neg \mathbf{f})$$

At the moment, there are no rules for modifying formulas which have a T sign. We shall introduce them later. For an introduction we wish to stick to the two-valued propositional case. As an illustrating example, of how tableaux systems work, assume we want to prove that $A \to (B \to A)$ is a tautology. We start with the initial formula $A \to (B \to A)$: **f**, i.e. we try to find an interpretation in which $A \to (B \to A)$ takes the value f. If we succeed, $A \to (B \to A)$ cannot be a tautology. On the other side: if we fail, then $A \to (B \to A)$ must be a tautology.

We apply $(\rightarrow \mathbf{f})$ to the initial formula $A \rightarrow (B \rightarrow A)$: \mathbf{f} . This yields A: \mathbf{t} and $(B \rightarrow A)$: \mathbf{f} . The tableau-proof tree is given in the figure below.



The formula A occurs in the above path with two different signs. This means that we failed to find an assignment under which $A \to (B \to A)$ takes the value f. Therefore $A \to (B \to A)$ must be a tautology.

As a more complex example consider the formula $((A \to B) \to A) \to A$ (Peirce's law). Again we start with $(((A \to B) \to A) \to A)$:f. Application of $(\to \mathbf{f})$ yields $((A \to B) \to A)$:t and A:f. Now, rule $(\to \mathbf{t})$ tells us that if $((A \to B) \to A)$:t then A:t (i.e. A must be true) or $(A \to B)$:f (i.e. $A \to B$ must be false). The or is coded by \parallel . In this case our prooftree branches, i.e. we generate several alternatives by applying a rule of type β .



Each alternative branch in the above proof tree contains a formula which has two different truth-signs. In other words: each branch contains a contradiction and hence, there is no assignment under which $((A \to B) \to A) \to A$ gets the value f.

Branches containing a contradiction are also said to be *closed*. Consequently, a tableau is said to be closed if and only if all of its branches are closed.

Definition 4.10 (Closed, Open). A branch is *closed* if and only if it contains a formula with at least two different signs. A tableau is closed if and only if each branch is closed. A tableau or a branch is *open* if and only if they are not closed.
We can now define what a tableau-proof is:

Definition 4.11. A *tableau-proof* for Φ is a closed tableau containing Φ : **f** as the initial formula.

We shall now show that Φ is a classical, propositional tautology if and only if there is a tableau-proof for Φ . The tableau-proof procedure operates on sets of signed formulas. Similar to the satisfiability relation between interpretation functions and propositional sentences, we can define when an interpretation function satisfies a set of signed formulas. This notion enables us to talk about the soundness (and later, completeness) of the tableau modification rules R.

Definition 4.12 (Satisfiability of Signed Formulas). A signed formula Φ :s is satisfiable if and only if there is an interpretation I such that $I(\Phi) = v(s)$ where $v : \{\mathbf{f}, \mathbf{t}, \mathsf{T}\} \to \{f, t, \mathsf{T}\}$ is defined as follows

$$v(s) := \begin{cases} t, & \text{if } s = \mathbf{t} \\ f, & \text{if } s = \mathbf{f} \\ \top, & \text{if } s = \top \end{cases}$$

Similarly, a branch is satisfiable if and only if each of its nodes is satisfiable; a tableau is satisfiable if and only if it contains at least one satisfiable branch.

It is easy to see that the above rules are sound, i.e. whenever we have a satisfiable tableau τ and apply rule R to τ , then $R(\tau)$ is satisfiable.

Proposition 4.17 (Soundness (Fitting)). Any application of rule R to a satisfiable tableau yields a satisfiable tableau.

Corollary 4.1. If Φ has a tableau proof, then Φ is a classical tautology.

On the other hand, the above tableau-system is complete.

Proposition 4.18 (Completeness (Fitting)). If Φ is a classical tautology, then Φ has a tableau proof.

We shall skip the proofs because they are a special case of the soundness and completeness proofs of the tableau procedure for K_3 . The reader who is interested in the completeness proofs for the above calculus may find them in [Fitting, 1990].

Let us now turn to a tableau proof procedure for our nonmonotonic paraconsistent logic. Since there is an elegant tableau procedure for Kleene's strong three-valued logic with Bolzano's entailment relation, we have only to modify this procedure to account for our preferential entailment relation. The tableau procedure for Kleene's strong three-valued logic given in [Bloesch, 1993] is a straight forward extension of the classical propositional procedure presented in the preceding section.

4.4.1 Bloesch's Tableau System

Bloesch extends the set of truth-signs to $\{\mathbf{t}, \mathbf{f}, \overline{\mathbf{t}}, \overline{\mathbf{f}}\}$ with $\overline{\mathbf{t}}$ meaning 'at least true', $\overline{\mathbf{f}}$ meaning 'at least false', \mathbf{t} meaning 'definitely true' and \mathbf{f} meaning 'definitely false'. Thus, a signed formula Φ :s is satisfiable if and only if there is an interpretation I such that $I(\Phi) \in v(s)$ where $v(s) : \{\mathbf{t}, \mathbf{f}, \overline{\mathbf{t}}, \overline{\mathbf{f}}\} \to 2^{\{t, f, \top\}}$ is defined as follows:

$$v(s) := \begin{cases} \{t\}, & \text{if } s = \mathbf{t} \\ \{f\}, & \text{if } s = \mathbf{f} \\ \{t, \top\}, & \text{if } s = \overline{\mathbf{t}} \\ \{f, \top\}, & \text{if } s = \overline{\mathbf{f}} \end{cases}$$

Note that the truth-signs denote sets of truth-values. This approach is also called *sets-as-signs strategy*.

$$\frac{(\varPhi \land \Psi):\mathbf{t}}{\varPhi:\mathbf{t}, \Psi:\mathbf{t}} (\land \mathbf{t}) \qquad \frac{(\varPhi \land \Psi):\mathbf{f}}{\varPhi:\mathbf{f} \parallel \Psi:\mathbf{f}} (\land \mathbf{f}) \qquad \frac{(\varPhi \land \Psi):\overline{\mathbf{t}}}{\varPhi:\overline{\mathbf{t}}, \Psi:\overline{\mathbf{t}}} (\land \overline{\mathbf{t}}) \qquad \frac{(\varPhi \land \Psi):\overline{\mathbf{f}}}{\varPhi:\overline{\mathbf{f}} \parallel \Psi:\overline{\mathbf{f}}} (\land \overline{\mathbf{f}})$$

$$\frac{(\varPhi \lor \Psi):\mathbf{t}}{\varPhi : \mathbf{t} \, \| \, \Psi: \mathbf{t}} \, (\lor \mathbf{t}) \qquad \frac{(\varPhi \lor \Psi):\mathbf{f}}{\varPhi : \mathbf{f}, \, \Psi: \mathbf{f}} \, (\lor \mathbf{f}) \qquad \frac{(\varPhi \lor \Psi):\overline{\mathbf{t}}}{\varPhi : \overline{\mathbf{t}} \, \| \, \Psi: \overline{\mathbf{t}}} \, (\lor \overline{\mathbf{t}}) \qquad \frac{(\varPhi \lor \Psi):\overline{\mathbf{f}}}{\varPhi : \overline{\mathbf{f}}, \, \Psi:\overline{\mathbf{f}}} \, (\lor \overline{\mathbf{f}})$$

$$\frac{(\varPhi \to \Psi) : \mathbf{t}}{\varPhi : \mathbf{f} \, \| \, \Psi : \mathbf{t}} \left(\to \mathbf{t} \right) \quad \frac{(\varPhi \to \Psi) : \mathbf{f}}{\varPhi : \mathbf{t}, \, \Psi : \mathbf{f}} \left(\to \mathbf{f} \right) \quad \frac{(\varPhi \to \Psi) : \overline{\mathbf{t}}}{\varPhi : \overline{\mathbf{f}} \, \| \, \Psi : \overline{\mathbf{t}}} \left(\to \overline{\mathbf{t}} \right) \quad \frac{(\varPhi \to \Psi) : \overline{\mathbf{f}}}{\varPhi : \overline{\mathbf{t}}, \, \Psi : \overline{\mathbf{f}}} \left(\to \overline{\mathbf{f}} \right)$$

$$\frac{\neg \Phi: \mathbf{t}}{\Phi: \mathbf{f}} (\neg \mathbf{t}) \qquad \qquad \frac{\neg \Phi: \mathbf{f}}{\Phi: \mathbf{t}} (\neg \mathbf{f}) \qquad \qquad \frac{\neg \Phi: \overline{\mathbf{t}}}{\Phi: \overline{\mathbf{f}}} (\neg \overline{\mathbf{t}}) \qquad \qquad \frac{\neg \Phi: \overline{\mathbf{f}}}{\Phi: \overline{\mathbf{t}}} (\neg \overline{\mathbf{f}})$$

The number of rules may seem too inflationary, but a closer look yields that for example, the rule $\wedge t$ is almost identical to its overlined counterpart $\wedge \overline{t}$.

Because truth-signs denote a set of truth-values, we need other closing conditions than in the classical case where the fact that two formulas had different truth-signs was sufficient; cf. Definition 4.10.

Definition 4.13 (Closed). A branch is *closed* if and only if there is a formula Φ such that

- 1. $\boldsymbol{\varPhi}{:}\mathbf{t}$ and $\boldsymbol{\varPhi}{:}\mathbf{f},$ or
- 2. Φ : t and Φ : \overline{f} , or
- 3. Φ : **f** and Φ : **t** is contained in the sequent or branch.

As usual, a tableau is closed if and only if every branch is closed.

It is easy to check that the above conditions correspond to the case where the signed formula Φ is not satisfiable.

The following Lemma is taken from Bloesch.

Lemma 4.2 (Bloesch, 1993).

- 1. Every application of a tableau modification rule to a satisfiable tableau yields a satisfiable tableau.
- 2. If there is a closed tableau for a set of signed formulas X, then X is not satisfiable.

Theorem 4.7 (Bloesch, 1993). Φ is true in every model of X if and only if every tableau τ for $X: \overline{t}, \Phi: f$ is closed.

4.4.2 Extending Bloesch's Tableau Procedure

Instead of finding an arbitrary model of X in which Φ is false, we have – in order to characterise Cn_3 – to find a preferred model of X in which Φ is false. If the search for such a model fails, then we know that $\Phi \in Cn_3(X)$.

Suppose we are given the initial tableau containing the signed formulas $\Xi: \overline{\mathbf{t}}$, $\Phi: \mathbf{f}$ (Ξ is the conjunction of all elements of X). Clearly, $\Xi: \overline{\mathbf{t}}$ is satisfiable. In order to find a preferred model for $\Xi: \overline{\mathbf{t}}$ it suffices to proceed as follows:

- 1. Expand every tableau branch completely. If the tableau cannot be further expanded, then
- 2. Determine which branches of the complete tableau are \Box -minimal, where $S \sqsubset S'$ holds by definition if and only if

$$\{A \mid A: \overline{\mathbf{t}} \in S \text{ and } A: \overline{\mathbf{f}} \in S\} \subset \{A \mid A: \overline{\mathbf{t}} \in S' \text{ and } A: \overline{\mathbf{f}} \in S'\}$$

for all atomic A.

3. Delete all branches which are not \Box -minimal; call the resulting tableau final.

Example 4.3. Let us prove that $\{A, \neg A, A \rightarrow B\} \not\Vdash B$.





there a \sqsubset -minimal branch which is open, the above tableau shows that B is not provable from $\{A, \neg A, A \rightarrow B\}$.

As another example consider the proof of $\{A, A \to B\} \parallel B$:



Again we have two branches. The left one, S_1 overinterprets A. It is therefore ruled out by Step 3 in the above procedure. Thus only the right branch remains. Since it is closed we have a tableau for B from $\{A, A \to B\}$.

Consider now a branch S of the final tableau resulting from the last step. There is a preferred model I of Ξ such that: $I(\Psi) = \top$ iff $\Phi: \overline{\mathbf{t}}$ and $\Phi: \overline{\mathbf{f}}$ appears on the branch S. Moreover, if the final tableau is open, then it contains an open branch which corresponds to some preferred model of Ξ in which Φ is false. To establish this result, we need the notion of signed Hintikka sets.

A signed Hintikka set H is a set of signed formulas such that for every atomic A we have that neither $A:\mathbf{t}, A:\overline{\mathbf{f}}$ nor $A:\mathbf{f}, A:\overline{\mathbf{t}}$ nor $A:\mathbf{t}, A:\mathbf{f}$ is contained in H. Moreover, if $(\Phi \land \Psi):\mathbf{t} \in H$ then $\Phi:\mathbf{t} \in H$ and $\Psi:\mathbf{t} \in H$. The rest is according to the modification rules. The clue about signed Hintikka sets is that (1) each open, complete branch corresponds to some Hintikka set and (2) each Hintikka set has a model. The last point is also called *model existence theorem*.

The following Lemma is fairly obvious:

Lemma 4.3. Let $S \sqsubset S'$ be two complete and open branches of a tableau τ for X and I_S be a preferred model of S. Then the following holds: for every preferred model $I_{S'}$ of S' we have $I_S \sqsubset I_{S'}$.

Proof. It is easy to verify that the set of all nodes from S is a Hintikka set; the same holds for the set of nodes of S'. Moreover, S is satisfied by the following

assignment:

$$I_{S}(A) = \begin{cases} t & , \text{ if } A: \overline{\mathbf{t}} \in S, A: \overline{\mathbf{f}} \notin S \\ f & , \text{ if } A: \overline{\mathbf{f}} \in S, A: \overline{\mathbf{t}} \notin S \\ \top & , \text{ if } A: \overline{\mathbf{t}} \in S, A: \overline{\mathbf{f}} \in S \\ t & , \text{ if } A: \overline{\mathbf{t}} \notin S, A: \overline{\mathbf{f}} \notin S \end{cases}$$

The mapping $I_{S'}$ is defined analogously. Moreover, $I_{S'}$ is a model for S' and for any model J of S' we have $I_{S'} \sqsubset J$ and $J \sqsubset I_{S'}$. By definition of satisfiability of signed formulas, it follows that $I_S \sqsubset I_{S'}$ and hence $I_S \sqsubset J$ for every model Jof S'.

A *BF-tableau proof* of Φ from X is any tableau τ such that

- 1. τ contains the initial tableau $X: \overline{\mathbf{t}}, \Phi: \mathbf{f}$ and
- 2. every branch of τ is complete and
- 3. every \sqsubset -minimal branch of τ is closed.

Theorem 4.8 (Soundness and Completeness for BF-proofs). Let X be finite. Φ is BF-provable from X if and only if $\Phi \in Cn_3(X)$.

Proof. First note that by Theorem 4.7 we know that I is a preferred model of X if and only if there is a branch S in the tableau construction for $X:\overline{\mathbf{t}}$ such that I satisfies S.

For the direction from right to left, assume that Φ is BF-provable from X. It follows from Lemma 4.2 and the definition of a BF-tableau proof that $X: \overline{\mathbf{t}}, \Phi: \mathbf{f}$ is not satisfiable by any preferred model of X. Hence, $X \Vdash_{\mathsf{K}_3} \Phi$.

For the opposite direction, assume to the contrary that $\Phi \in Cn_3(X)$ but Φ is not BF-provable from X. This implies that $X: \overline{\mathbf{t}}, \Phi: \mathbf{f}$ produces a tableau in which some \Box -minimal branch is open and complete. Thus, there is a preferred model I of X which satisfies $X: \overline{\mathbf{t}}, \Phi: \mathbf{f} - \mathbf{a}$ contradiction.

The above tableau procedure for Cn_3 has several advantages and disadvantages. It is certainly a plus that the procedure can be obtained easily from the one Bloesch gave for Kleene's logic with Bolzano's entailment relation. The brute force search for a preferred model is easy to implement and thus, the strong machine oriented character of tableau systems is preserved. In [Weber, 1996b] you may find an enhancement which constructs \Box -minimal branches in a bottom-up manner (instead of generating all branches in a brute force manner and then deciding which ones are \Box -minimal). Even though tableaux procedures are very natural for a machine oriented proof theory, they do not give us much insight into the valid patterns of inference. For example, if we know that X entails A and B, then we also know that X entails $A \wedge B$. Whereas the rules for treating disjunction and conjunction might be fairly obvious⁴, the situation becomes complicated in the case of implication and negation. This is because Modus Ponens is not a valid rule of inference. The following section is thus devoted to question: Which rules of inference are needed to characterise the inferential behaviour of K₃?

4.5 A Sequent Calculus

A classical sequent, or short CL-sequent, is a pair (X, Y) of finite sets of formulas. We denote sequents by $X \vdash Y$, where ' \vdash ' is a new metalogical symbol which should not be confused with any provability symbol. We shall omit the CL if it is clear from the context. The intended meaning of a sequent is very close to the notion of provability: a CL-sequent $X \vdash Y$ is valid if and only if every twovalued interpretation which satisfies all the elements of X does at least satisfy one element of Y. Thus, the sequent $X \vdash Y$ is valid if and only if

$$\xi_1 \wedge \ldots \wedge \xi_{|X|} \Vdash \psi_1 \vee \ldots \vee \psi_{|Y|}$$

holds, where $X = \{\xi_1, \dots, \xi_{|X|}\}, Y = \{\psi_1, \dots, \psi_{|Y|}\}.$

For example, for every formula Φ , the sequent $\Phi \vdash \Phi$ is valid. The task of a Sequent Calculus is to produce all valid sequents.

Sequent Systems have been introduced by Gentzen in his dissertation where he presented a sequent calculus called LK for classical logic and one for intuitionistic logic as well as their first-order versions.

Let me now briefly introduce Gentzen's sequent calculus $L{\sf K}$ for classical propositional logic.

Notational convention We write X, Y for $X \cup Y$ and X, Φ for $X \cup \{\Phi\}$ to improve readability.

Axioms: All sequents of the form $\Phi \vdash \Phi$. Logical Rules:

$$\frac{X \vdash \Phi, Y}{X, \neg \Phi \vdash Y} \text{ NEA} \qquad \qquad \frac{X, \Phi \vdash Y}{X \vdash Y, \neg \Phi} \text{ NES}$$

⁴ In fact we already have them implicitly in our set of tableau rules, if we read the rules upside down.

$$\frac{X \vdash Y, \Phi}{X \vdash Y, \Phi \lor \Psi} \text{ OES 1}$$

$$\frac{X \vdash Y, \Phi \lor \Psi}{X \vdash Y, \Phi \lor \Psi} \text{ OES 1}$$

$$\frac{X \vdash Y, \Psi}{X \vdash Y, \Phi \lor \Psi} \text{ OES 2}$$

$$\frac{X, \Psi \vdash Y}{X, \Phi \land \Psi \vdash Y} \text{ UEA 1} \qquad \qquad \frac{X \vdash Y, \Phi \land X \vdash Y, \Psi}{X \vdash Y, \Phi \land \Psi} \text{ UES}$$
$$\frac{X, \Psi \vdash Y}{X, \Phi \land \Psi \vdash Y} \text{ UEA 2}$$

$$\frac{X \vdash X', \varPhi \quad Y, \varPsi \vdash Y'}{X, Y, \varPhi \to \varPsi \vdash X', Y'} \text{ IEA} \qquad \qquad \frac{X, \varPhi \vdash Y, \varPsi}{X \vdash Y, \varPhi \to \varPsi} \text{ IES}$$

Structural Rules:

$$\frac{X \vdash Y}{X, \varPhi \vdash Y} VL \qquad \qquad \frac{X \vdash Y}{X \vdash Y, \varPhi} VR$$

Cut:

-

$$\frac{X\vdash \Phi, X' \quad Y, \Phi\vdash Y'}{X, Y\vdash X', Y'} \operatorname{Cut}$$

The rules IEA and IES are originally called FEA and FES in [Gentzen, 1934]. The following two results are due to Gentzen:

Theorem 4.9. [Gentzen, 1934]

- 1. The above calculus LK is sound and complete w.r.t. classical propositional logic.
- 2. Any sequent provable in the above calculus is also provable without using the Cut-rule.

The second point of Theorem 4.9 is the famous Cut-elimination theorem.

Some of the above rules become invalid for K_3 , others remain valid. For example, the two OES rules seem to be unproblematic whereas Cut can obviously not be sustained.

Definition 4.14 (K₃-Sequent, Validity). A K₃-sequent is a pair (X, Y) of finite sets of formulas, denoted by $X \models_{K_3} Y$. A K₃-sequent $X \models_{K_3} Y$ is valid if and only if each preferred model of X satisfies at least one element of Y.

The proof theory for K_3 relies on the following observation which is another (weakened) version of the deduction theorem called *ultra normalised*, because it requires that the antecedent always takes the truth-value t in every preferred model of a given set $X \cup \{\Phi\}$.

Theorem 4.10 (Ultra Normalised Deduction Theorem). If $X \Vdash \Phi \to \Psi$ and Φ does not take the value \top in any preferred model of $X \cup \{\Phi\}$ then $X \cup \{\Phi\}$ lime Ψ .

The validity of the above theorem is immediate. Corollary, the following version of the CUT rule is valid for K_3

$$\frac{X \vdash_{\mathsf{K}_3} \varPhi \quad X', \varPhi \vdash_{\mathsf{K}_3} \varPsi \quad : \mathsf{PC}(X, \varPhi)}{X, X' \vdash_{\mathsf{K}_2} \varPsi} \ \operatorname{Cut}$$

The precondition $\mathsf{PC}(X, \Phi)$ means 'provided that Φ does not take the value \top in any preferred model of $X \cup X'$ '. In the following we shall denote ' Φ does not take the value \top in any preferred model of X' by $X \downarrow \Phi$.

Using the above condition, we can now give the axioms and rules of the calculus LK_3 .

4.5.1 The Calculus LK_3

The calculus LK_3 has the following axioms and rules:

Axioms: All sequents of the form $\Phi, X \vdash_{\kappa_3} \Phi$.

Logical Rules: (the names are derived from Gentzen's original terminology).

$$\frac{X, \varPhi \vdash_{\mathsf{K}_3} Y}{X \vdash_{\mathsf{K}_3} Y, \neg \varPhi} \text{ NES}$$

$$\frac{X_{ \vdash_{\mathsf{K}_{3}}}Y, \phi \lor \Psi}{X_{ \vdash_{\mathsf{K}_{3}}}Y, \phi \lor \Psi} \text{ OES } \qquad \qquad \frac{X_{ \vdash_{\mathsf{K}_{3}}}Y, \phi \lor \Psi}{X_{ \vdash_{\mathsf{K}_{3}}}Y, \phi, \Psi} \text{ OBS }$$

$$\frac{X, \Phi, \Psi \vdash_{\mathsf{K}_{3}} Y}{X, \Phi \land \Psi \vdash_{\mathsf{K}_{3}} Y} \text{ UEA} \qquad \qquad \frac{X \vdash_{\mathsf{K}_{3}} \Phi, Y \quad X \vdash_{\mathsf{K}_{3}} \Psi, Y}{X \vdash_{\mathsf{K}_{3}} \Phi \land \Psi, Y} \text{ UES}$$

$$\frac{X \vdash_{\mathsf{K}_{3}} Y, \Phi \quad X', \Psi \vdash_{\mathsf{K}_{3}} Y' \quad :\mathsf{PC}(X \cup X' \downarrow \neg \Phi)}{X, X', \Phi \rightarrow \Psi \vdash_{\mathsf{K}_{3}} Y, Y'} \text{ IEA} \qquad \frac{X, \Phi \vdash_{\mathsf{K}_{3}} Y, \Psi}{X \vdash_{\mathsf{K}_{3}} Y, \Phi \rightarrow \Psi} \text{ IES}$$

$$\frac{X, \Phi \land \Psi \vdash_{\mathsf{K}_{3}} Y}{X, \Phi, \Psi \vdash_{\mathsf{K}_{3}} Y} \text{ UBA} \qquad \qquad \frac{X \vdash_{\mathsf{K}_{3}} Y, \Phi \land \Psi}{X \vdash_{\mathsf{K}_{3}} Y, \Phi, \Psi} \text{ UBS}$$

$$\frac{X \vdash_{\mathsf{K}_3} Y, \Phi \to \Psi : \mathsf{PC}(X \cup \{\Phi\} \downarrow \neg \Phi)}{X, \Phi \vdash_{\mathsf{K}_3} Y, \Psi} \text{ IBS}$$

Structural Rules:

$$\frac{\not{ \forall \vdash_{\mathsf{K}_{3}} \Phi}}{X \vdash_{\mathsf{K}_{3}} \Phi} \operatorname{Inclusion} \qquad \frac{X \vdash_{\mathsf{K}_{3}} Y}{X \vdash_{\mathsf{K}_{3}} Y, \Phi} \operatorname{VR}$$
$$\frac{X \vdash_{\mathsf{K}_{3}} \Phi \quad X \vdash_{\mathsf{K}_{3}} \Psi}{X, \Phi \vdash_{\mathsf{K}_{3}} \Psi} \operatorname{Cumulativity}$$

Weak Cut:

$$\frac{X \vdash_{\mathsf{K}_3} \Phi \quad X', \Phi \vdash_{\mathsf{K}_3} \Psi : \mathsf{PC}(X \cup X' \downarrow \neg \Phi)}{X, X' \vdash_{\mathsf{K}_3} \Psi} \text{ Weak Cut}$$

We say that Φ is LK_3 -provable from X if and only if the sequent $X \vdash_{K_3} \Phi$ can be generated.

NB. The axioms of LK_3 are more general than the axioms for Gentzen's calculus. They represent the inclusion property of the semantical entailment relation $\parallel \vdash$.

Comments on the rules: there are three rules, namely NEA, VL, and Cut, of the classical sequent calculus LK which do not hold for K_3 . The failure of CUT is obvious. To see that NEA, which is

$$\frac{X \models_{\mathsf{K}_3} Y, \varPhi}{X, \neg \varPhi \models_{\mathsf{K}_3} Y} \text{ NEA}$$

is invalid, consider the application of NEA to the valid sequent $A, A \wedge B \vDash_{\kappa_3} B, A$ would yield $\neg A, A, A \rightarrow B \nvDash_{\kappa_3} B$ which is invalid. The rule NEA corresponds to the reductio ad absurdum principle which is invalid in κ_3 .

The rule IBS, which is a version of Weak Cut, can be obtained from the other rules. We wish to list it explicitly in order to make the proofs simpler.

The rule UBS has no analogue in Gentzen's calculus, though being valid for classical propositional logic. The reason is that it can be obtained quite easily in LK:

$$\frac{X \vdash \Phi \land \Psi}{X \vdash \Phi} \frac{\frac{\Phi \vdash \Phi}{\Phi \land \Psi \vdash \Phi} (\text{UEA 1})}{(\text{Cut})}$$

In the calculus LK_3 we can simulate the effect of UEA 1 but not the full power of Gentzen's Cut. We therefore list the UBS explicitly.

Theorem 4.11 (Soundness). Let X be a set of formulas. Then, for any Φ , $X \models_{\mathsf{K}_3} \Phi$ implies $X \Vdash \Phi$.

Proof. It is easy to check that the above rules are sound. The logical rules NES, UEA, IES, UES and UBA are immediate. Soundness of Weak Cut, IEA and IBS follows from Theorem 4.10. Soundness of Inclusion and VR is immediate.

Theorem 4.12. If Φ is a $(K_3$ -)tautology, then the sequent $\emptyset \models_{K_3} \Phi$ can be generated.

Proof. It is easy to see that if Φ is an axiom of the classical propositional calculus then $\emptyset \models_{\kappa_3} \Phi$ and by Inclusion $X \models_{\kappa_3} \Phi$ can be generated. For the inductive step we will simulate modus ponens. Suppose we have generated $\emptyset \models_{\kappa_3} \Psi$ and $\emptyset \models_{\kappa_3} \Psi \to \Phi$. Then we can prove $\emptyset \models_{\kappa_3} \Phi$ as follows (we use $\models_{\kappa_3} \Phi$ as a shorthand for $\emptyset \models_{\kappa_3} \Phi$):

The rules IBS and Weak Cut can be applied, because by soundness of LK_3 we know that if $\models_{K_3} \Psi$ holds then Ψ has a two-valued model. Thus, we have shown that if Φ is a tautology, then the sequent $\models_{K_3} \Phi$ can be generated. \Box

Example 4.4. Let us show that $A \wedge B$ is provable from $\{A, \neg A, A \rightarrow B, C, C \rightarrow B, E, \neg E, F \rightarrow E\}$. Let $X' := \{A, \neg A, A \rightarrow B, E, \neg E, F \rightarrow E\}$, $X'' := X \setminus \{A\}$.

$$\frac{X'', A \vdash_{\mathsf{K}_3} A}{X \vdash_{\mathsf{K}_3} A} \underset{X \vdash_{\mathsf{K}_3} A \to B}{\overset{X', C \vdash_{\mathsf{K}_3} C \to B \vdash_{\mathsf{K}_3} B}{X', C, C \to B \vdash_{\mathsf{K}_3} B}} (\text{IEA}) \overset{C \text{ does not take}}{\underset{\{C\}}{\overset{\text{the value } T \text{ in any}}{fc}}}_{\{C\}} (\text{UES})$$

Proposition 4.19. If X is finite, then the set of formulas which are LK_3 -provable from X is recursively enumerable.

Proof. It is sufficient to note, that for each rule, the precondition PC(...) – if there is any – is decidable. But since X is always finite, this must be the case. \Box

Theorem 4.13. LK_3 is a finite recursively enumerable approximation of $Cn_3(X)$.

Proof. By soundness of LK_3 's rules we have inclusion w.r.t. to Cn_3 . Moreover by the Proposition 4.19 we have that LK_3 's provability relation is r.e.

Inclusion. By axiom.

Nonmonotonicity. We can use our standard counter-example to monotonicity. Let $X = \{A, A \rightarrow B\}$. We can generate $X \models_{\aleph_3} B$ by using IEA. It is simple to check that IEA is the only rule which admits the introduction of \rightarrow in the antecedent. However, the application of IEA is blocked when augmenting X by $\neg A$. Hence, B is not $L\aleph_3$ -provable from $X \cup \{\neg A\}$.

Cumulativity. By the corresponding rule.

Finite Idempotency. Follows by Proposition 4.7 from inclusion and cumulativity.

Conservativity. We have to show that if X has a two-valued model, then $\{\Phi \mid X \Vdash_{\overline{K_3}} \Phi\} = Cn_{c1}(X)$. The inclusion from left to right follows from soundness of LK_3 (Theorem 4.11). For the converse direction, we know that since $\Phi \in Cn_{c1}(X)$, there is a classical Modus Ponens proof of Φ from X. It suffices to show that if X has a two-valued model, then from $X \vDash_{\overline{K_3}} \Phi$ and $X \vDash_{\overline{K_3}} \Phi \to \Psi$ we can infer $X \vDash_{\overline{K_3}} \Psi$. We apply the same argument as in Theorem 4.12:

$$\frac{X \vdash_{\mathbf{k}_{3}} \Psi \to \Phi}{X \vdash_{\mathbf{k}_{3}} \Phi} (\text{IBS}) \\ \frac{X \vdash_{\mathbf{k}_{3}} \Psi}{X \vdash_{\mathbf{k}_{3}} \Phi} (\text{Weak Cut})$$

It is easy to see that IBS and Weak Cut are applicable because X is consistent and $X \models_{\mathsf{K}_3} \Phi$ is valid. The following proposition is the analogue of Proposition 4.8.

Proposition 4.20 (Supradeductivity). If $\Phi \in \operatorname{Cn}_{\operatorname{Bolz}}(X)$ then $X \models_{\kappa_3} \Phi$.

Proof. Like in the proof of Proposition 4.8 it suffices to show that the rules (\wedge -I), (\wedge -E) and (\vee -I) of the calculus for Cn_{Bolz} given in Section 4.1 can be simulated. (\wedge -I) can be handled by UES and (\wedge -E) by UBS.

The above sequent-style calculus shows which patterns of inference are legal for K_3 . It gives us a good insight into how contradictory information is treated in our paraconsistent logic: in the very moment in which some formula can be reasonably suspected to be paraconsistent, i.e. as soon as there is at least one preferred model in which this formula takes the value \top , some inferences are not valid. This shows again the extreme cautious character of the inference relation. However, testing whether Φ can take the value \top is clearly of semantical nature and it is debatable whether the above calculus is completely in the spirit of Gentzen's sequent calculi or not.

Gentzen did pay special attention to a formal system which 'is as close as possible to mathematical reasoning'. We can now ask how close the rules of LK_3 are to common-sense reasoning. I have got no knocking down philosophical arguments, but I think that our reasoning in the presence of contradictions is only cautious if we are aware of these contradictions – no matter how we found out about them. This does also hold for mathematical reasoning. For example, assume that number theory or set theory is not consistent. Our whole reasoning within and about these formal theories is not affected by fact that they are inconsistent as long as we do not know about this. Once we are conscious of inconsistencies we are extremely careful about which sentences they should justify. This is reflected by the rules with the extra condition $X \downarrow \Phi$. Thus, I would confirm that LK_3 reflects the patterns of a certain type of common-sense reasoning.

Another interesting point has been brought up by Priest in his paper on Minimally Inconsistent LP. Priest asks whether a logic like K_3 can be used to model mathematical reasoning. For example, if number theory is inconsistent then reasoning with K_3 prevents us from sanctioning the belief in every sentence. Hence, the argument for using K_3 instead of classical logic is similar to the one which favours intuitionistic logic to classical logic: if a theory is decidable, then no intuitionist would reject an indirect proof. If, however, a theory is not decidable, then indirect proofs will be rejected. As for K_3 we could say that if we know that a theory is consistent, then there is no problem with classical logic. If, however, we cannot guarantee the consistency of a theory (e.g. number theory), then it is safer to reason in a K_3 -manner.

4.6 Conclusion

We investigated K_3 's proof-theory. We have shown that the set of consequences is Σ_2^0 -hard. i.e. not recursively enumerable. We thus introduced the concept of a r.e. approximation. We then discussed three proof-systems: Hilbert-style, semantical tableaux and Gentzen-type.

The Hilbert-style and Gentzen-type systems were all r.e. approximations of Cn_3 , whereas the tableau-system is a sound and complete procedure for deciding whether Φ is entailed by a finite set X. This suggests that the tableau procedure is the most appropriate device for a mechanisation of K₃. We have, however, argued that the other proof-systems do not become obsolete: the Hilbert-style systems show how to modify the modus ponens rule in order to obtain a syntactical approximation of K₃ and the Gentzen-type system shows which patterns of inference are valid for K₃. It is an open question, whether the Hilbert- and Gentzen-like systems are also complete.

CHAPTER 5

Reasoning About Unknown Information

Up to now we have discussed the problem of how to deal with contradictory information. Contradictory information means that we have too much information. On the other side we can also have the situation where we have a lack of information or data. This means that the database has neither information on $\overline{\Phi}$ nor on $\neg \Phi$.

The problem is how to find out whether the database (which has a certain deduction mechanism) has information of Φ or not? The easiest way would be to put the query Φ ? If this query fails (i.e. if the database answers 'No.'), we could ask for $\neg \Phi$? If both queries fail, then we know that the database does not have information on Φ .

This procedure enables us to *check* whether there is information on Φ or not. But can we say that the database itself has information on whether Φ is known or unknown? The answer is 'No', because there is no sentence like 'unknown Φ ' contained in the database. If the language of the database is a classical first-order or propositional language, then the database cannot even represent a fact like 'unknown Φ ' because there is no operator 'unknown'.

The mission to be accomplished in this chapter is to develop a logic which enables a database to reason about its own content or data. Assume that we have enriched our language by an operator \Box with the intended meaning that $\Box \Phi$ should represent something like 'it is known to the database that Φ ' or ' Φ can be derived from the database' (by means of its deduction mechanism). The intended meaning of $\Box \Phi$ is similar to that of Φ if none of them is negated: there should be no difference between telling the database that ' Φ holds' or telling it 'you *know* that Φ holds'. If, however, the database has *no* information on Φ , for example when it is empty, then $\neg \Box \Phi$ should be entailed. This is different from a database which entails $\neg \Phi$, because in general such a database could not be empty.

A logic which enables a database to reason about its own content and identifies the sentences which are *not* known has a great impact on AI applications. For example, machine learning algorithms, could really improve when having the explicit information that A is unknown. This has been reported by, e.g. Hirsh [Hirsh, 1990] (cf. also the transcript of the discussion on meta-reasoning in [Brazdil and Konolige, 1990]). Moreover, there are machine learning systems and knowledge-acquisition systems whose inference engine provide an unknown operator (cf. [Morik and Wrobel, 1933], [Emde, 1991]). Unfortunately there is a severe theoretical drawback of these systems: they lack a clear semantics. One part of the semantics is algebraical or set theoretical while the semantics for the unknown operator is given in a proof theoretical manner. For example, the semantics for the unknown-operator in [Morik and Wrobel, 1933] is described as (\dots) unknown [A] evaluates to true if and only if the proposition A (\dots) cannot be proved. This is a reading rather than a semantics. The logic presented in this chapter will fix this problem.

The plot is as follows: after having presented some preliminaries we will investigate what is meant by ' Φ is unknown', thus developing a formal semantics which eliminates the aforementioned problems with earlier approaches. In the ensuing section we shall study fundamental properties of our logic. Again, closure properties and compactness will play an important role. We will then start to give a syntactical characterisation of our entailment relation. This characterisation enables us to compare in Section 5.4 to Moore's autoepistemic logic and the nonmonotonic logic NML-2 of McDermott. It will turn out that our logic, which was originally intended to model the semantics of an unknown-operator used by some AI-systems, is located between autoepistemic logic and McDermott's nonmonotonic S5.

5.1 Semantical Investigations

5.1.1 Terminology

We will extend the propositional language \mathcal{L} by the modal operators \Diamond and \Box . A formula preceded by \Diamond or \Box is called a *modal formula*. A *nonmodal formula* is a formula which has no modal subformula. Again, we shall omit the reference to a special propositional signature Σ when it is clear from the context or not important and then just talk about \mathcal{L}_M .

As usual, a two-valued propositional interpretation function of a signature Σ is a mapping $I : \Sigma \to \{t, f\}$. A Kripke structure \mathfrak{M} is a tuple $(M, V)^1$ where Mis an index set (also called set of possible worlds or states) and $V = \{I_1, I_2, \ldots\}$ is a set of two-valued propositional interpretation functions such that there is a bijection between V and M. Throughout this chapter we shall only deal with two-valued functions I. We shall thus omit the attribute 'two-valued'. The set of all Kripke structures interpreting a signature Σ is denoted by $\mathrm{STRUCT}(\Sigma)$.

¹ Normally, Kripke structures have an accessibility relation $R \subseteq M \times M$. Throughout this chapter we assume R to be complete (i.e. $R = M \times M$); it follows that M is also universal (this implies that it is reflexive, symmetric and transitive).

5.1.2 Unknown Means Satisfiability

Let us return to our database. We say that the database $X \subseteq \mathcal{L}_M$ does not have any information about a sentence $\Phi \in \mathcal{L}_M$ if and only if neither Φ nor $\neg \Phi$ can be proved from the database². A careless transformation of this idea into the definition of validity could yield the potpourri semantics mentioned in the beginning of this chapter. The key idea to obtain an appropriate semantics is that if neither Φ nor $\neg \Phi$ is provable from X, then Φ as well as $\neg \Phi$ is satisfiable with respect to X. Thus, Φ is unknown w.r.t. the database X if both, $\{\Phi\} \cup X$ and $\{\neg \Phi\} \cup X$ are satisfiable. Such an unknown-operator could be modelled by using Kripke structures as interpreting structures. The validity relation \models_{MK} is identical to the validity relation of modal S5.

Definition 5.1 (Validity, Kripke-model). Let $\Phi, \Psi \in \mathcal{L}$ be formulas, $\mathfrak{M} = (M, V)$ a Kripke structure, $\alpha \in M$

$$\begin{split} \mathfrak{M} &\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Phi & \text{for an atomic } \Phi \text{ if } I_{\alpha}(\Phi) = t \\ \mathfrak{M} &\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Phi \wedge \Psi & \text{iff } \mathfrak{M} \mid_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Phi \text{ and } \mathfrak{M} \mid_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Psi \\ \mathfrak{M} &\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Phi \to \Psi & \text{iff } \mathfrak{M} \not\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Phi \text{ or } \mathfrak{M} \mid_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Psi \\ \mathfrak{M} &\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Diamond \Phi & \text{iff there is } \beta \in M \text{ such that } \mathfrak{M} \mid_{\overline{\mathsf{M}}\mathsf{K}\,\beta} \Phi \\ \mathfrak{M} &\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \neg \Phi & \text{iff } \mathfrak{M} \not\models_{\overline{\mathsf{M}}\mathsf{K}\,\alpha} \Phi \end{split}$$

The connectives \lor (disjunction), \oplus (exclusive or) and the operator \Box are defined as abbreviations in the usual way.

 \mathfrak{M} is a (Kripke-) model for Φ , alternatively Φ is valid in \mathfrak{M} , $(\mathfrak{M} \models_{\mathsf{MK}} \Phi)$ if for all $\alpha \in M$, we have $\mathfrak{M} \models_{\mathsf{MK} \alpha} \Phi$. We extend the relation \models_{MK} in the usual way to sets of formulas.

The next step is to define the concept of MK-entailment (denoted by the relational symbol $\parallel_{\overline{MK}}$). As said before, the goal is to have an entailment relation $\parallel_{\overline{MK}}$ such that $X \parallel_{\overline{MK}} \neg \Box \neg \Phi$ (or shorter, $X \Vdash_{\overline{MK}} \Diamond \Phi$), whenever Φ is satisfiable w.r.t. X or, in other terms, whenever $X \cup \{\Phi\}$ has a model. However, if we allow Φ to be an arbitrary sentence, this could yield counter-intuitive results: let $X = \{B\}$; there is a Kripke-structure \mathfrak{M} such that $\mathfrak{M} \models_{\overline{MK}} X \cup \{\Box C\}$. Thus, $X \cup \{\Box C\}$ is

² An alternative would be to restrict the non-provability condition to Φ , i.e. to say that we don't have any information on Φ if and only if we cannot prove Φ . We judge this to be a matter of taste.

satisfiable and we should have $X \Vdash_{\overline{\mathsf{MK}}} \Diamond \Box C$. But $\Diamond \Box C$ is semantically equivalent to $\Box C$. Hence, $X \Vdash_{\overline{\mathsf{MK}}} \Box C$. This contradicts our intuition because the conclusion '*C* is known' is entailed by a database which had no information about *C*. We will therefore restrict Φ to be a nonmodal sentence³.

We have to find out whether there is a model \mathfrak{M} of X such that $\Phi \in \mathcal{L}$ is true in at least one of \mathfrak{M} 's possible worlds. Fortunately, we can restrict the search for such a model to those models of X which contain a *maximal* set of possible worlds. To see why, note that if $X \cup \Diamond \Phi$ has a model then $X \cup \Diamond \Phi$ is valid in some model \mathfrak{M}_{\max} of X which contains a maximal set of possible worlds (remember that Φ is nonmodal). Generally,

Intuitively, X MK-entails Ψ if Ψ is valid in each structure $\mathfrak{M} = (M, V)$ with

- 1. $\mathfrak{M} \models_{MK} X$ and
- 2. *M* is maximal, that is *V* contains as many two valued interpretations as possible, that is \mathfrak{M} makes as many formulas of the form $\Diamond A$ valid as possible, where *A* is a nonmodal formula.

To ensure that we consider only maximal models (in the above sense), we must be able to compare arbitrary Kripke structures to find out which structure is a maximal model. This can be done by correlating all structures via the S5substructure relation. The idea behind S5-substructures is that, $\mathfrak{M} = (M, V)$ is a S5-substructure of $\mathfrak{N} = (N, W)$ if \mathfrak{N} 'extends' \mathfrak{M} , i.e if $M \subseteq N$ (and $V \subseteq W$). However, since M and N could be arbitrary index sets (i.e. M could be a set of natural numbers, while N could be a set of characters), we have to ensure, that they are 'comparable'. This will be guaranteed by existence of an isomorphism.

Definition 5.2 (S5-Substructure). Two structures $\mathfrak{M} = (M, V)$, $\mathfrak{N} = (N, W)$ are *isomorphic* (denoted by $\mathfrak{M} \cong \mathfrak{N}$) if and only if there is a bijection $I : M \to N$ such that $V_{\alpha} = W_{I(\alpha)}$, for every $\alpha \in M$.

We say that $\mathfrak{M} = (M, V)$ is an S5-substructure of $\mathfrak{N} = (N, W)$, denoted by $\mathfrak{M} \preccurlyeq \mathfrak{N}$, if and only if there is $\mathfrak{N}' = (N', W')$ such that $N' \subseteq N$, $W' \subseteq W$ and \mathfrak{N}' is isomorphic to \mathfrak{M} . We say that \mathfrak{M} is a strict S5-substructure of \mathfrak{N} , denoted by $\mathfrak{M} \prec \mathfrak{N}$, if and only if there is $\mathfrak{N}' = (N', W')$ such that $N' \subset N$, $W' \subset W$ and \mathfrak{N}' is isomorphic to \mathfrak{M} .

We can use \prec as a preference relation in the sense of [Shoham, 1988] and define preferred models on the basis of this preference relation.

³ This restriction can practically be justified by the fact that in the aforementioned AI systems the operator **unknown** can only be applied to propositions which do not already contain this operator.

Definition 5.3 (Preferred model). Let $\Phi \in \mathcal{L}_M, X \subseteq \mathcal{L}_M$. Define a relation $|\underline{\bullet}_{MK} \subseteq |\underline{\bullet}_{MK}$ such that

- $\mathfrak{M}' \models_{\mathcal{M}} X$; we say that \mathfrak{M} is a *preferred* or *maximal* model for X^4 .

Following Shoham [Shoham, 1988] we define an entailment relation $\parallel_{\overline{MK}}$ as follows:

Definition 5.4 ($\parallel_{\overline{M}K}$, Cn_{MK}). $X \Vdash_{\overline{M}K} \Phi$ if and only if, $\mathfrak{M} \models_{\overline{M}K} X$ implies $\mathfrak{M} \models_{\overline{M}K} \Phi$. $\operatorname{Cn}_{\mathsf{MK}}(X)$ is the set of all formulas entailed by X, i.e. $\operatorname{Cn}_{\mathsf{MK}}(X) =_{def} \{ \varPhi \mid X \Vdash_{\widetilde{\mathsf{MK}}} \varPhi \}^5$.

The following examples illustrate the above definitions

Example 5.1. Let $\Sigma = \{A, B\}$ be a signature, $X = \{\neg \Box A \rightarrow A\}$. There are two preferred models $\mathfrak{M}_1 = (\{\alpha, \beta\}, I_1)$ and $\mathfrak{M}_2 = (\{\gamma, \delta\}, I_2)$ for X:

$$\begin{aligned} I_{1_{\alpha}}(A) &= t \left| I_{1_{\beta}}(A) = t \right| \left| I_{2_{\gamma}}(A) = f \right| I_{2_{\delta}}(A) = t \\ I_{1_{\alpha}}(B) &= t \left| I_{1_{\beta}}(B) = f \right| \left| I_{2_{\gamma}}(B) = t \right| \left| I_{2_{\delta}}(B) = t \right| \end{aligned}$$

Thus, we have $X \Vdash_{\mathsf{MK}} \Diamond B \land \Diamond \neg B$ but $X \nvdash_{\mathsf{MK}} \Diamond A$ and $X \nvdash_{\mathsf{MK}} \Diamond \neg A$.

Example 5.2. Let $\Sigma = \{A, B\}$ be a propositional signature, $X = \emptyset$. Then $\{\Diamond A,$ $\langle B, \langle \neg A, \langle \neg B \rangle \subset \operatorname{Cn}_{\mathsf{MK}}(X).$

The following lemma states a basic property of the preferential validity relation and will be useful in a number of proofs.

Lemma 5.1. If $\mathfrak{M} \models_{\overline{\mathsf{M}}\mathsf{K}} X$ and $X \models_{\overline{\mathsf{M}}\mathsf{K}} \Phi$ then $\mathfrak{M} \models_{\overline{\mathsf{M}}\mathsf{K}} X \cup \{\Phi\}$.

Proof. Assume to the contrary that there is $\mathfrak{A}' \prec \mathfrak{A}$ such that $\mathfrak{A}' \models_{\mathsf{MK}} X \cup \{\Phi\}$. Hence $\mathfrak{A}' \models_{MK} X$ which contradicts the assumption that \mathfrak{A} is a preferred model of X.

Let us denote the logic based on the above entailment relation by MK, for minimal knowledge.

 $^{^4}$ Please note that the extension from formulas to sets of formulas in the above definition is not equivalent to $\mathfrak{M} \models_{\mathsf{MK}} X$ if and only if $\mathfrak{M} \models_{\mathsf{MK}} \Phi$, for every $\Phi \in X'$. ⁵ The idea of taking maximal S5 structures already appeared in [Lifschitz, 1991] who

called his system a logic of Minimal Knowledge (MK). To exploit the relationship between his work and mine, please see the section on related work at the end of this chapter.

5.2Properties

In this section we shall discuss some properties of our logic. The main questions we have in mind are: what general properties, which we know from classical logic, e.g. closure properties, do hold for our logic?

5.2.1**Basic Properties**

Proposition 5.1. Cn_{MK} is a nonmonotonic, cumulative, preclosure operator.

Proof. Clearly, Cn_{MK} is a preclosure operator. To see that it is cumulative, we have to show that

$$X \Vdash_{\overline{\mathsf{M}}\mathsf{K}} \Phi, \Psi \text{ implies } X \cup \{\Phi\} \Vdash_{\overline{\mathsf{M}}\mathsf{K}} \Psi$$

Let $\mathfrak{A} \models_{M_{\mathsf{K}}} X$. We then know that $\mathfrak{A} \models_{M_{\mathsf{K}}} \Phi, \Psi$ and thus by Lemma 5.1 we have $\mathfrak{A} \models_{M_{\mathsf{K}}} X \cup \{\Phi\}$ and hence $X \cup \{\Phi\} \parallel_{M_{\mathsf{K}}} \Psi$. The nonmonotonicity follows easily from a counter-example to monotonicity:

 $\emptyset \Vdash_{\overline{\mathsf{M}}\mathsf{K}} \Diamond \neg A \text{ but } \{A\} \Vdash_{\overline{\mathsf{M}}\mathsf{K}} \Diamond \neg A.$

Theorem 5.1 (Preferred Model Existence). If X has an MK-model \mathfrak{M} then X has a preferred MK-model \mathfrak{N} such that $\mathfrak{M} \preccurlyeq \mathfrak{N}$.

Proof. Consider the following chain of models of X:

$$\mathfrak{M} \prec \mathfrak{M}_0 \prec \mathfrak{M}_1 \prec \ldots$$

with $\mathfrak{M}_i = (M_i, V_i)$. We shall show that this chain has an upper bound \mathfrak{N} . Without loss of generality we can assume that $M_i \subseteq M_{i+1}$. Define, $N =_{def} \bigcup M_i$ and $W =_{def} \bigcup V_i$. Let $\mathfrak{N} = (N, W)$. We have $\mathfrak{N} \models_{\overline{M}_k} X$: let $\Phi \in X$ we show by induction on the degree of Φ that if $\mathfrak{M}_i \models_{\overline{K}_3} \Phi$ then $\mathfrak{N} \models_{\overline{K}_3} \Phi$. Let Φ be atomic. Suppose that for every \mathfrak{M}_i we have $\mathfrak{M}_i \models_{\overline{K}_3} \Phi$ but $\mathfrak{N} \not\models_{\overline{K}_3} \Phi$. Since \mathfrak{N} is not a model Φ , there must be a state $\alpha \in N$ such that $V_{\alpha}(\Phi) = f$. By construction of N, there must be a set M_i such that $\alpha \in M_i$ and hence, $\mathfrak{M}_j, \alpha \not\models \mathfrak{G}$. For the inductive step, we have several cases:

 $\Phi = \Diamond \varphi$. For every \mathfrak{M}_i there must be a state $\alpha \in M_i$ such that $\mathfrak{M}_i \models_{\overline{\mathbf{k}}_{\alpha} \alpha} \varphi$. By construction of \mathfrak{N} and the induction hypothesis we have $\mathfrak{N} \models_{\overline{\kappa}_{3\alpha}} \varphi$ and hence, $\mathfrak{N} \models_{\overline{\mathsf{K}}_3} \Diamond \varphi.$

 $\Phi = \varphi_1 \lor \varphi_2$ By induction hypothesis.

The other connectives can be reduced to one of the cases above.

Moreover, we have $\mathfrak{M}_i \preccurlyeq \mathfrak{N}$, for all \mathfrak{M}_i of the above chain. Thus \mathfrak{N} is an upper bound of the chain. It follows from Zorn's Lemma that there is a maximal element. \Box

Lemma 5.2. Let X be a set of formulas, \mathfrak{M} a model of X and L a literal such that $\mathfrak{M} \not\models_{\mathsf{M}_{\mathsf{X}}} L$. There is a preferred model \mathfrak{N} of X, $\mathfrak{M} \preccurlyeq \mathfrak{N}$ such that $\mathfrak{N} \not\models_{\mathsf{M}_{\mathsf{X}}} L$.

Proof. Let $\mathfrak{M} = (M, V)$ and $\mathfrak{N} = (N, W)$. Without loss of generality we assume that $M \subseteq N$. Since $\mathfrak{M} \not\models_{\mathsf{MK}} L$, there is a state α such that $V_{\alpha}(L) = f$. By $M \subseteq N$ we have that there is W_{α} such that $W_{\alpha}(L) = f$. Hence, $\mathfrak{N} \not\models_{\mathsf{MK}} L$. \Box

The following corollary follows by an easy structural induction proof from Lemma 5.2.

Corollary 5.1. Let X be a set of formulas, \mathfrak{M} a model of X and Φ a nonmodal formula such that $\mathfrak{M} \not\models_{\mathsf{MK}} \Phi$. There is a preferred model \mathfrak{N} of X, $\mathfrak{M} \preccurlyeq \mathfrak{N}$ such that $\mathfrak{N} \not\models_{\mathsf{MK}} \Phi$.

Note, that Proposition 5.2 does not hold for reasoning with minimal propositional model, i.e. where f is preferred to t. For example we have that I with I(A) = t and I(B) = t satisfies $\{A\}$ but is no minimal model for $\{A\}$. The minimal model for $\{A\}$ is J with J(A) = t and J(B) = f. We have $I \not\models \neg B$ and $J \models \neg B$ even though J is preferred to I.

Remark 5.1. The deduction theorem fails for MK. It does, however, already fail for the modal logics S4, S5.

5.2.2 Other Properties

Let us now state some closure properties.

Proposition 5.2. Define $\operatorname{Th}_{S5}(X)$ to be the smallest set containing X, all classical axioms, as well as all instances of the axiom schemes Ax

$\Box(\varPhi \to \Psi) \to (\Box \varPhi \to \Box \Psi)$	Κ
$\Box \Phi o \Phi$	Т
$\Box \varPhi \to \Box \Box \varPhi$	4
$\Diamond \Phi \to \Box \Diamond \Phi$	5

which is closed under application of modus ponens and the rule of necessitation, i.e. $\Phi/\Box\Phi$.

Let \mathfrak{M} be a structure and $T_{\mathfrak{M}} =_{def} \{ \Phi \mid \mathfrak{M} \models_{\mathsf{MK}} \Phi \}$. Then, $T_{\mathfrak{M}}$ is closed under $\mathrm{Th}_{\mathrm{S5}}$

Proof. Easy (omitted).

Let us now turn to an important property, namely compactness. The fact that $\models_{\overline{M}K}$ is compact follows from [Chellas, 1980], who calls it compactness of consistency. If Cn_{MK} were monotonic, then the compactness of entailment would follow immediately. However, in the presence of nonmonotonicity, compactness of entailment is a nontrivial property.

Theorem 5.2 (Compactness of Consistency). Let $X \subseteq \mathcal{L}_M$ be a possibly infinite set of formulas. X has a model if and only if every finite subset of X has a model.

Proof. See e.g. [Chellas, 1980]

Theorem 5.3 (Nonmodal Compactness). Let $X \Vdash_{\mathsf{MK}} \Phi$ for some X and some nonmodal Φ . Then there is a finite set $X_{\mathrm{fin}} \subseteq X$ such that $X_{\mathrm{fin}} \Vdash_{\mathsf{MK}} \Phi$.

Proof. Assume to the contrary that for all $X_{\operatorname{fin}} \subseteq X$ we have $X_{\operatorname{fin}} | \not_{\operatorname{MK}} \Phi$. Hence by [Chellas, 1980], Theorem 2.16 (15), for all X_{fin} there is $\mathfrak{M} \models_{\operatorname{MK}} X_{\operatorname{fin}} \cup \{\neg \Phi\}$. Thus, by Theorem 5.2, there is $\mathfrak{A} = (M, V)$ such that $\mathfrak{A} \models_{\operatorname{MK}} X \cup \{\neg \Phi\}$, hence $\mathfrak{A} \not\models_{\operatorname{MK}} \Phi$. Now, either \mathfrak{A} is a preferred model of X, or there is $\mathfrak{A}', \mathfrak{A} \preccurlyeq \mathfrak{A}'$ such that \mathfrak{A}' is a preferred model of X. But by Corollary 5.1 neither \mathfrak{A} nor \mathfrak{A}' can be a model of Φ . This contradicts the assumption that $X \parallel_{\operatorname{MK}} \Phi$.

Corollary 5.2. Let X be a set of sentences such that $X \Vdash_{\mathsf{MK}} \Box \Phi$, for some nonmodal Φ . There is some finite subset $X_{\mathrm{fin}} \subseteq X$ such that $X_{\mathrm{fin}} \Vdash_{\mathsf{MK}} \Box \Phi$. Let me briefly state what we have got so far: we have defined a cumulative, nonmonotonic entailment relation $\parallel_{\overline{MK}}$ and an operator \diamond such that for any nonmodal sentence A the database X entails $\diamond A$ if and only if $X \mid_{\overline{MK}} A$ (provided that X has a unique preferred model). Moreover, the consequence operator Cn_{MK} is inclusive, idempotent. A weakened version of the compactness theorem holds. It is an open question whether Cn_{MK} is compact.

5.3 Adequacy of the Entailment Relation

Now that we have defined the entailment relation $\Vdash_{\overline{\mathsf{MK}}}$ and discussed some of its basic properties let us examine the question whether this entailment relation does really express our intuition about 'unknown'. That is, does $X \Vdash_{\overline{\mathsf{MK}}} \Diamond \Phi$ whenever $X \Vdash_{\overline{\mathsf{MK}}} \neg \Phi$ and $X \Vdash_{\overline{\mathsf{MK}}} \Phi$ hold? The answer is 'yes' if X has a unique preferred model and 'no' otherwise. The reason that, in general, we could have $X \Vdash_{\overline{\mathsf{MK}}} \Diamond \Phi$ even if $X \Vdash_{\overline{\mathsf{MK}}} \Phi$ and $X \Vdash_{\overline{\mathsf{MK}}} \neg \Phi$ is that the concept of satisfiability can be expressed in the object language itself: let $X = \{\Diamond A \oplus \Diamond B\}$ (literally, either A is satisfiable or B is satisfiable). X has two preferred models, one satisfying $\Diamond A$ and $\neg B$ and the other one satisfying $\Diamond B$ and $\neg A$. Clearly, we have $X \Vdash_{\overline{\mathsf{MK}}} A$ and $X \nvDash_{\overline{\mathsf{MK}}} \neg A$ but also $X \nvDash_{\overline{\mathsf{MK}}} \Diamond A$.

I shall present two solutions to the problem of multiple preferred models: the first one is to identify sublanguages of \mathcal{L}_M which always admit unique preferred models, i.e. whenever X is a subset of one of the sublanguages, then X has at most one preferred model. Of course, it could happen that X does not have any model at all.

The identified sublanguages, however, are not strong enough to express things like 'A is unknown implies B'. This leads us to second approach, where we do not restrict the language but use a filter to select only a subset of the preferred models. This filter allows us to read implicational formulas like inference rules.

5.3.1 Sublanguages which allow for unique preferred models

The following states an important and basic fact about formulas already having a unique preferred model.

Proposition 5.3. Let $X \subseteq \mathcal{L}_M$, $\Phi \in \mathcal{L}_M$ such that both X and Φ has a unique preferred S5-model. $X \cup \{\Phi\}$ has a unique preferred S5-model if and only if $X \cup \{\Phi\}$ has an S5-model.

Proof. We show that for any two formula Φ and Ψ which both have a unique preferred model, we can construct a unique preferred model of $\{\Phi,\Psi\}$ (provided that $\Phi \wedge \Psi$ has indeed a model). Let $\mathfrak{A}_1 = (A_1, V_{A_1})$ (resp. $\mathfrak{A}_2 = (A_2, V_{A_2})$) the unique preferred model of Φ (resp. Ψ). Define $\mathfrak{A} = (A, V_A)$ as

$$A =_{def} A_1 \cap A_2$$
$$V_A =_{def} V_{A_1} \cap V_{A_2}$$

We claim that \mathfrak{A} is a unique preferred model for $\{\Phi, \Psi\}$. By assumption, $\Phi \wedge \Psi$ has a model and, hence, it has a preferred model, say $\mathfrak{M} = (M, V)$. Assume to the contrary, that $\mathfrak{A} \neq \mathfrak{M}$. This yields the following two cases:

- $\mathfrak{M} \not\subseteq \mathfrak{A}$ In this case there must be a state $\alpha \in M$ such that $\alpha \notin A_1 \cap A_2$. That is, $\alpha \notin A_1$ or $\alpha \notin A_2$. Assume without loss of generality that $\alpha \notin A_1$. By the preferred model existence theorem, we are able to expand \mathfrak{M} to obtain a preferred model $\mathfrak{M}' = (M', V')$ of Φ (in the case of $\alpha \notin A_2$ we have to expand \mathfrak{M} to a preferred model of Ψ). But then we have $\alpha \in M'$ and $\alpha \notin A_1$. Hence, $\mathfrak{M}' \neq \mathfrak{A}_1$. Since both, \mathfrak{M}' and \mathfrak{A}_1 are preferred models of Φ we have a contradiction to the assumption that there is only one preferred model of Φ .
- $\mathfrak{A} \not\subseteq \mathfrak{M}$ In this case there is $\alpha \in A_1 \cap A_2$ and $\alpha \notin M$. Again, we extend \mathfrak{M} to a preferred model \mathfrak{M}' of, without loss of generality, Φ . Obviously, α is no state of any extension \mathfrak{M}' (otherwise this would yield a contradiction to the assumption that \mathfrak{M} is a preferred model of $\Phi \wedge \Psi$). Therefore, $\mathfrak{M}' \neq \mathfrak{A}_1$ and thus, we have a contradiction to the assumption that \mathfrak{A}_1 is the only preferred model of Φ .

It follows from the above that for any two sets X, Y which both have a unique preferred model, we have that $X \cup Y$ has at most one preferred model.

Proposition 5.3 is a very powerful tool, since it says that once we have identified sublanguages \mathcal{L}_i of \mathcal{L}_M each admitting unique preferred models, then the union of \mathcal{L}_i admits unique preferred model. Of course, in order to be relevant for practical usage, we have to find sublanguages which admit unique preferred models.

Proposition 5.4. The following subsets X of \mathcal{L}_M have at most one unique model:

1. $X \subset \mathcal{L}$ 2. $X \subset \mathcal{L}_{\mathrm{NT}} =_{def} \{ \Phi \mid \Phi \equiv \Phi_1 \lor \ldots \lor \Phi_n \text{ and } \Phi_i = \neg \Box A \text{ where } A \in \mathcal{L}_M \}$ *Proof.* (1) is immediate. To verify (2) note that as soon as we have a disjunct saying 'A is not necessary' we can take the maximal S5 structure $\mathfrak{A}_{Max} = (M, V)$ where V is the set of *all* propositional valuations and M is large enough to refer to each valuation. Clearly, \mathfrak{A}_{Max} is a preferred model for $\neg \Box A$ and hence for any $\Phi \in \mathcal{L}_{NT}$. It follows by Proposition 5.3 that each $X \subseteq \mathcal{L}_{NT}$ has a unique preferred model.

The languages mentioned in the above proposition are quite unrealistic. They are either only propositional or nearly trivial (NT). However, by Proposition 5.3 we have that we can combine these languages, i.e.

Corollary 5.3. Any $X \subseteq \mathcal{L}_{NT} \cup \mathcal{L}$ has at most one preferred model.

We shall now describe a language \mathcal{L}_{D} such that Φ has a unique preferred model if and only if Φ is equivalent to some sentence of \mathcal{L}_{D} . To establish this result we need an additional concept.

Definition 5.5 (M-literal, Modal Horn). Let $\Phi \in \mathcal{L}_M$ be a formula. Φ is said to be an *M*-literal if and only if Φ has the form $\Box A$ or $\neg \Box A$ where A is any nonmodal formula

An M-literal is said to be *positive* if and only if it has the form $\Box A$. A *negative* M-literal is an M-literal of the form $\neg \Box A$.

 Φ is said to be *modal Horn* if and only if Φ has the form

$$\Phi_1 \lor \Phi_2 \lor \ldots \lor \Phi_r$$

where each Φ_i is an M-literal and at most one Φ_i is a positive M-literal.

We can now define the language $\mathcal{L}_{\rm D}$, which is a sublanguage of the language $\mathcal{L}_{\rm MCNF}$, i.e. of the set of all formulas in *Modal Conjunctive Normal Form*. A formula $\Phi = \Phi_1 \wedge \ldots \wedge \Phi_n$ is in $\mathcal{L}_{\rm MCNF}$ if and only if each Φ_i is a disjunction such that each disjunct is either (a) an M-literal or (b) a formula from \mathcal{L}_0 . It is known that for each modal formula Φ there is $\Phi_{\rm MCNF}$ such that $\Phi \leftrightarrow \Phi_{\rm MCNF}$ is an S5-tautology. Even though, $\mathcal{L}_{\rm MCNF}$ restrict the syntactical appearance of a formula, it does not restrict the expressibility.

Define $\mathcal{L}_{D} \subset \mathcal{L}_{MCNF}$ as follows: \mathcal{L}_{D} is the set of all formulas $\Phi = \Phi_{1} \land \ldots \land \Phi_{n}$ where each Φ_{i}

- 1. contains at most one positive M-literal and
- 2. if Φ_i contains a positive M-literal, then no disjunct from Φ_i is an ordinary propositional formula.

The following formulas are in $\mathcal{L}_{\rm D}$

$$(\neg \Box A \lor B) \land ((\neg \Box C \land D) \lor E \to F)$$
$$(\Box A \lor \neg \Box B \lor \neg \Box C)$$

The following formulas are *not* in \mathcal{L}_{D}

$$(\neg \Box A \lor B) \land ((\Box C \land D) \lor E \to F)$$
$$(\Box A \lor \Box B \lor \neg \Box C)$$

Proposition 5.5. Let $\Phi \in \mathcal{L}_M$ be an arbitrary modal formula. The class of models for Φ is finite.

Proof. Let Σ_n be the set of all variables occurring in Φ . There are only finitely many propositional interpretation functions $I : \Sigma_n \to \{t, f\}$. Let \mathfrak{I} be the class of all these functions. Clearly, \mathfrak{I} is finite and hence, the power set $2^{\mathfrak{I}}$ is finite. Thus, there are only finitely many Kripke-structures over Σ_n . It follows that there are only finitely many Kripke-models for Φ and only finitely many preferred Kripke-models for Φ .

The following lemma is similar to the well-known result that any non-HORNclause has two minimal models.

Lemma 5.3. Let Φ be a formula with preferred models $\mathfrak{M}_1, \ldots, \mathfrak{M}_n$ (we know by Proposition 5.5 that this class is finite) such that $\mathfrak{M}_i \not\cong \mathfrak{M}_j$, $i \neq j$, i.e. the models are pairwise not isomorphic. There is a formula

$$\varphi = \Box A_1 \lor \Box A_2 \lor \ldots \lor \Box A_n$$

such that

- 1. $\Phi \mid_{\mathsf{MK}} \varphi$ 2. $\Phi \mid_{\mathsf{MK}} \Box A_{\pi_1} \lor \ldots \lor \Box A_{\pi_{n-1}}$ such that $\Box A_{\pi_i} = \Box A_j$, for some disjunct $\Box A_j$ of φ .
- 3. $A_1 \not\equiv A_2 \not\equiv \ldots \not\equiv A_n$

Proof. Consider an arbitrary model \mathfrak{M}_i of Φ . Assume to the contrary that for any nonmodal A_i with $\mathfrak{M}_i \models_{\mathsf{MK}} A_i$ there is some \mathfrak{M}_j , $\mathfrak{M}_i \not\cong \mathfrak{M}_j$ such that $\mathfrak{M}_j \models_{\mathsf{MK}} A_i$. Since \mathfrak{M}_i and \mathfrak{M}_j are not isomorphic there is at least one $\alpha \in M_j$ such that for all $\beta \in M_i$ it holds that $V_{j_{\alpha}} \neq V_{i_{\beta}}$. Define

$$\begin{split} M'_i &=_{\scriptscriptstyle def} M_i \cup \{\alpha\} \\ V'_i &=_{\scriptscriptstyle def} V_i \cup \{V_{j_\alpha}\} \\ \mathfrak{M}'_i &=_{\scriptscriptstyle def} (M'_i, V'_i) \end{split}$$

For all nonmodal A we have

$$\mathfrak{M}_i \models_{\overline{\mathsf{M}}\mathsf{K}} A \text{ implies } \mathfrak{M}'_i \models_{\overline{\mathsf{M}}\mathsf{K}} A$$

and hence, for all \varPsi

$$\mathfrak{M}_i \models_{\mathcal{M}_{\mathcal{K}}} \Psi \text{ implies } \mathfrak{M}'_i \models_{\mathcal{M}_{\mathcal{K}}} \Psi.$$

Thus, \mathfrak{M}_i cannot be preferred. Hence, there must be a nonmodal formula A_i such that

$$\mathfrak{M}_i \models_{\mathsf{MK}} A_i \text{ and for all } \mathfrak{M}_j \colon \mathfrak{M}_j \not\models_{\mathsf{MK}} A_i, \quad i \neq j.$$

Thus,

$$\mathfrak{M}_i \models_{\mathsf{MK}} \Box A_i \text{ and for all } \mathfrak{M}_j \colon \mathfrak{M}_j \not\models_{\mathsf{MK}} \Box A_i, \quad i \neq j$$

Therefore there are nonmodal formulas A_1, \ldots, A_n , which are not semantically equivalent, such that $\mathfrak{M}_i \models_{\mathsf{MK}} \Box A_i, \mathfrak{M}_i \not\models_{\mathsf{MK}} \Box A_j, j \neq i$ Since these $\mathfrak{M}_1, \ldots, \mathfrak{M}_n$ are the only preferred models of Φ we also have that

$$\Phi \Vdash_{\overline{\mathsf{M}}\mathsf{K}} \Box A_1 \lor \ldots \lor \Box A_n$$

and

$$\Phi \not\Vdash_{\mathsf{MK}} \Box A_{\pi_1} \lor \ldots \lor \Box A_{\pi_{n-1}}$$

with $\Box A_{\pi_i} = \Box A_j$, for some disjunct A_j of $\Box A_1 \lor \ldots \lor \Box A_n$.

Theorem 5.4. Φ has a unique preferred model if and only if there is $\Phi_D \in \mathcal{L}_D$ such that $\Phi \leftrightarrow \Phi_D$ is an S5-tautology.

Proof.

Part 1 'If' By Lemma 5.3 we have that if Φ has more than one preferred model, then there is a formula $\varphi = \Box A_1 \lor \ldots \lor \Box A_n$ such that

$$\begin{split} \varPhi \Vdash_{\widetilde{\mathsf{M}}\mathsf{K}} \Box A_1 \lor \ldots \lor \Box A_n \\ \varPhi \nvDash_{\widetilde{\mathsf{M}}\mathsf{K}} \Box A_{\pi_1} \lor \ldots \Box A_{\pi_{n-1}} \text{ where } A_{\pi_i} = A_j \text{ for some disjunct } \Box A_j \text{ of } \varphi. \end{split}$$

There is, however, no formula $\Phi_D \in \mathcal{L}_D$ with these properties (if there were such a formula, then there would be a formula from \mathcal{L}_D which is semantically equivalent to φ).

Part 2 'Only if' We have to show that if there is no syntactically equivalent formula $\Phi_D \in \mathcal{L}_D$, then Φ has at least two preferred models. Assume without loss of generality that Φ is in MCNF. If $\Phi \notin \mathcal{L}_D$, then Φ has some conjunct Ψ_i such that

$$\Psi_i = \varphi_1 \vee \ldots \vee \varphi_n$$

and

$$\varphi_i, \varphi_j \text{ are both positive M-literals}, i \neq j \text{ or}$$
(1)

 φ_i is a positive M-literal and $\varphi_j \in \mathcal{L}_0, i \neq j$ (2)

If every MCNF-formula Φ' which is semantically equivalent to Φ contains some disjunct in which two positive M-literals $\Box A$, $\Box B$ appear, then there are preferred models $\mathfrak{M}, \mathfrak{N}$ of Φ such that $\mathfrak{M} \models_{MK} \Box A \land \Diamond B \land \Diamond \neg B$ and $\mathfrak{N} \models_{MK} \Box B \land$ $\Diamond A \land \Diamond \neg A$. Clearly $\mathfrak{M} \ncong \mathfrak{N}$. Hence, Φ has at least two preferred models. This proves the case (1).

For the the case (2) the argumentation is nearly identical, because any S5structure which satisfies A also satisfies $\Box A$ (by soundness of the Rule of Necessitation). Hence, $A \vee \Box B$ has two preferred models because $\Box A \vee \Box B$ has two preferred models. Thus, if every MCNF formula Φ' which equivalent to Φ contains $A \vee \Box B$ in some conjunct, then Φ has at least two preferred models.

It might be worth noting that even though $\Phi = \neg \Box A \land (\Box A \lor B) \notin \mathcal{L}_{D}$, there is an \mathcal{L}_{D} formula which is syntactically equivalent to Φ , namely $\neg \Box \Phi \land B$.

Theorem 5.4 identifies a sublanguage \mathcal{L}_{D} whose subclasses always have at most one preferred model. This implies that as long as a database language is identical to (or a subset of) \mathcal{L}_{D} , then the entailment relation $\parallel_{\overline{\mathsf{MK}}}$ appropriate w.r.t. the analysis of the meaning of unknown, i.e. does it hold for all nonmodal A that

 $X \Vdash_{MK} A$ and $X \Vdash_{MK} \neg A$ implies $X \Vdash_{MK} \Diamond A$ and $X \Vdash_{MK} \Diamond \neg A$?

How appropriate is the language \mathcal{L}_{D} w.r.t. our aim of reasoning about unknown information? Does it allow us to express all the things we would like to express? Just think of an expression like 'if A is unknown then B'. This can be translated as $(\Diamond A \land \Diamond \neg A) \rightarrow B$ which equivalent to the MCNF formula $\Box \neg A \lor \Box A \lor B$. Since this is the only information in the database, the database cannot be transformed to some \mathcal{L}_{D} -formula and by Theorem 5.4 has no unique preferred model.

Even if we transcribed 'if A is unknown then B' as $\Diamond \neg A \rightarrow B$ we would get $\Box A \lor B$ as MCNF formula. Again, it is easy to see that there is no syntactically equivalent formula in \mathcal{L}_{D} . This is somewhat disappointing, especially if one has practical applications in mind for which such a reasoning pattern might be quite useful. A solution is to apply the following filtration method.

5.3.2 Treating Implicational Formulas as Rules

Consider the database $X = \{ \Diamond A \to A \}$. This set has two preferred models, yielding $\Diamond A \notin \operatorname{Cn}_{\mathsf{MK}}(X)$ and $\Diamond \neg A \notin \operatorname{Cn}_{\mathsf{MK}}(X)$. This is a consequence of treating \rightarrow as material implication. Nevertheless, sometimes formulas like $\Diamond A \to A$ are used to express "If A is consistent with your formulas, then assert A" (cf. [Emde, 1991]). The point is that even though \rightarrow is given the semantics of material implication, as in [Emde, 1991], [Morik and Wrobel, 1933], it is *used* or to be read as an inference rule.

Lukaszewicz described a cunning trick to rule out $\mathfrak{A}_1 = (\{1\}, \{V_1(A) = f\})$ as a preferred model for X. Thus, the only preferred model which remains is $\mathfrak{A}_2 = (\{1\}, \{V_1(A) = t\})$ (cf. [Lukaszewicz, 1990]). We will adapt this trick for our needs. In order to make it work, we have to ensure that each formula in our database has a certain normal form.

Definition 5.6 (Ordered MCNF). A formula $\Phi_1 \land \Phi_2 \land \ldots \land \Phi_n$ is said to be in *ordered conjunctive normal form* if and only if each Φ_i is of the form

$$\neg \Box B \lor \Box C_1 \lor \ldots \lor \Box C_k \lor A$$

where $B, C_1, \ldots, C_k, A \in \mathcal{L}_0$.

It is known that each modal formula can be reduced in S5 to ordered MCNF (cf. [Hughes and Cresswell, 1968]). Please, note that we can write each conjunct Φ_i as $(\Box B \land \neg \Box C_1 \land \ldots \land \neg \Box C_k) \to A$.

Definition 5.7 (Applicable [Lukaszewicz, 1990]). Let $\Phi \in \mathcal{L}$ in ordered MCNF, $T = Th_{S5}(T)$ a set of formulas. A conjunct $\Phi_i = (\Box B \land \neg \Box C_1 \land \ldots \land \neg \Box C_n) \rightarrow A$ of Φ is said to be applicable w.r.t. T if and only if $B \in T$, $C_1, \ldots, C_n \notin T$; otherwise, Φ_i is said to be inapplicable w.r.t T.

Definition 5.8 (Strongly preferred model). Let $X \subset \mathcal{L}$ be a set of formulas, \mathfrak{A} a Kripke structure and $T_{\mathfrak{A}} =_{def} \{ \Phi \mid \mathfrak{A} \models_{MK} \Phi \}$. Define

$$X' =_{def} X \setminus \{ \Phi \mid \Phi \text{ is inapplicable w.r.t } T_{\mathfrak{A}} \}$$

 \mathfrak{A} is a *strongly preferred model* for X if and only if

$$\mathfrak{A} \models_{\overline{\mathsf{M}}\mathsf{K}} X \qquad \text{Cond. A} \\ \mathfrak{A} \models_{\overline{\mathsf{M}}\mathsf{K}} X' \qquad \text{Cond. B}$$

We say that $X \Vdash_{\mathfrak{F}} \Phi$ if and only if every strongly preferred model of X is a model of Φ . Clearly, $X \Vdash_{\widetilde{\mathsf{MK}}} \Phi$ implies $X \Vdash_{\mathfrak{F}} \Phi$.

Example 5.3. Let $\Sigma = \{A\}$, $X = \{\Diamond A \to A\}$. There are two preferred models $\mathfrak{M}_1, \mathfrak{M}_2$ for X but only one them is strongly preferred. Let $\mathfrak{M}_1 = (\{1\}, \{I_1(A) = t\})$ and $\mathfrak{M}_2 = (\{1\}, \{I_1(A) = f\})$. $\Diamond A \to A$ is applicable w.r.t $T_{\mathfrak{M}_1}$ but not w.r.t $T_{\mathfrak{M}_2}$. Thus we have to check whether Cond. B holds.

1. $\mathfrak{M}_1 \not\models_{\overline{\mathsf{M}}^{\mathsf{K}}} X \setminus \varnothing$ (there are no inapplicable formulas w.r.t $T_{\mathfrak{M}_1}$) and 2. $\mathfrak{M}_2 \not\models X \setminus X$ (every formula of X is inapplicable w.r.t $T_{\mathfrak{M}_2}$).

Hence, $X \Vdash_{\mathbb{S}} A$.

The strong consequence operator $Cn_S(X) =_{def} \{ \Phi \mid X \Vdash_{\mathbb{S}} \Phi \}$ inherits all closure properties of Cn_{MK} . It does, however, not always yield a unique preferred model. Consider the following example:

Example 5.4. Consider the following set $X = \{ \Diamond A \to B, \Diamond \neg B \to \neg A \}$. We have the following preferred models for X:

$$\mathfrak{A}_{1}: I_{1}(A) = t | I_{2}(A) = f \qquad \mathfrak{A}_{2}: I_{1}(A) = f | I_{2}(A) = f I_{1}(B) = t | I_{2}(B) = t \qquad \mathfrak{A}_{2}: I_{1}(A) = f | I_{2}(A) = f I_{1}(B) = f | I_{2}(B) = t \qquad \mathfrak{A}_{2}: I_{1}(A) = f | I_{2}(A) = f \\ I_{2}(B) = f | I_{2}(B) = t \qquad \mathfrak{A}_{2}: I_{1}(A) = f | I_{2}(A) = f | I_{2}(B) = f | I$$

Theorem 5.5. Cn_S is not compact.

Proof. Consider the following formula:

$$\Phi_n =_{def} \bigwedge_{i=0}^n A_i \land (\Diamond \neg A_{n+1} \to B)$$

and define

$$X = \{ \Phi_n \mid n \le \omega \}$$

The set X has the following preferred model:

$$\mathfrak{A}: I_1(A_i) = t | I_2(A_i) = t \quad i \le \omega$$
$$I_1(B) = f | I_2(B) = t$$

Moreover, this model is strongly preferred and hence, we have $\Diamond \neg B \in Cn_S(X)$. However, for every finite $X_{\text{fin}} \subset X$ we have $B \in Cn_S(X_{\text{fin}})$: let $k = max\{n \mid \Phi_n \in X_{\text{fin}}\}$. X_{fin} has the following two preferred models:

$$\begin{array}{l|l} \mathfrak{A}_{1}: \ I_{1}(A_{i}) = t \\ I_{1}(A_{k+1}) = t \\ I_{1}(B) = t \end{array} \begin{vmatrix} I_{2}(A_{i}) = t \\ I_{2}(A_{k+1}) = f \\ I_{2}(B) = t \end{vmatrix} \begin{array}{l} \mathfrak{A}_{2}: \ I_{1}(A) = f \\ I_{1}(A_{k+1}) = t \\ I_{1}(B) = f \end{vmatrix} \begin{vmatrix} I_{2}(A) = f \\ I_{2}(A_{k+1}) = t \\ I_{2}(B) = t \end{vmatrix} \begin{vmatrix} I_{2}(A) = f \\ I_{2}(A_{k+1}) = t \\ I_{2}(B) = t \end{vmatrix}$$

The formula $\Diamond \neg A_{k+1} \rightarrow B$ is applicable w.r.t. $T_{\mathfrak{A}_1}$ but not w.r.t. $T_{\mathfrak{A}_2}$ because $A_2 \in T_{\mathfrak{A}_2}$. Moreover, \mathfrak{A}_2 is no strongly preferred model of $X \{ \Diamond \neg A_2 \rightarrow B \}$. Hence, \mathfrak{A}_1 is the only strongly preferred model of X_{fin} and we have $\Diamond \neg B \notin Cn_S(X_{\text{fin}})$, for every finite $X_{\text{fin}} \subseteq X$.

5.4 Relationship to Nonmonotonic Modal Logics

We shall now relate our logic to some nonmonotonic modal formalisms. One of the earliest attempts to attack the problem of nonmonotonic reasoning was by means of modal logic. In 1982 McDermott introduced some nonmonotonic versions of well-known monotonic modal logics such as S4 and S5. But unfortunately, it turned out that the most promising formalisation of nonmonotonic reasoning based on modal S5 collapses to monotonic S5. Then, in 1983, Moore decided to develop a new formalism called autoepistemic logic, which is a reconstruction of an earlier proposal by Doyle and McDermott. Autoepistemic logics are nowadays the most prominent nonmonotonic modal formalisms.

From a syntactical point of view, autoepistemic logic can be regarded to be a weaker⁶ system than a nonmonotonic formalism based on modal S5, because

⁶ Please note that the terms stronger (containing more axioms) and weaker (containing less axioms) are somewhat meaningless within nonmonotonic logics, because the addition of axioms does not guarantee that we get more theorems.

autoepistemic logic is based on a system which is called *weak S5* (i.e. K45). The main result of this Section is that our logic lies between nonmonotonic K45 and S5, that is, it can be seen as an extension of autoepistemic logic towards McDermott's ideas. Thus, we will first compare our logic with Moore's autoepistemic logic and then with McDermott's Nonmonotonic Logic II (NML-2).

5.4.1 Extensional Entailment

Many nonmonotonic formalisms for example, Default Logics and Autoepistemic Logics do not discuss consequence operators. They rather make use of the notion of an *extension*. The main difference between the set of (nonmonotonic) consequences and an extension is that if we are given a set X of sentences, then the set of (nonmonotonic) consequences is unique while X may have different extensions.

In order to provide a basis for a comparison between logics based on the notion of extension and our logic, we have to say what corresponds to an extension in our logic. According to Stalnaker, an extension can be regarded as a final belief set; thus, containing a maximal set of beliefs⁷. One property which reflects the intuition about final beliefs sets is that they should be *stable* (cf. [Stalnaker, 1993]).

Definition 5.9 (Stable). A set $X \in \mathcal{L}$ of sentences is stable if and only if it meets the following requirements:

- 1. X is closed under classical (nonmodal) propositional consequence Cn_{cl}
- 2. if $\Phi \in X$ then $\neg \Diamond \neg \Phi \in X$
- 3. if $\neg \Phi \notin X$ then $\Diamond \Phi \in X$

We propose to regard every preferred model of X as a final belief state. The set of formulas which are valid in a preferred model of X is a final belief set. This motivates the following definition.

Definition 5.10 (Extensional entailment). Let $X, Y \subseteq \mathcal{L}$; we say X extensionally entails Y (denoted by $X \Vdash_{\mathbb{E}} Y$ if and only if there is a structure \mathfrak{M} such that $\mathfrak{M} \models_{\mathsf{MK}} X$ and $Y = \{ \varPhi \mid \mathfrak{M} \models_{\mathsf{MK}} \varPhi \}$. Y is said to be an S5-based extension of X (because the validity relation $\models_{\mathsf{MK}} \mathfrak{M}$ is that of modal S5).

⁷ Stalnaker talks about autoepistemic extensions, but I think his interpretation does also apply to every other nonmonotonic formalism which makes use of the term 'extension'

Please note that the relation \Vdash_{E} is only defined between sets of formulas. Further, $\operatorname{Cn}_{MK}(X) = \bigcap \{T \mid X \Vdash_{E} T \}.$

It has already been observed by Konolige, Moore and Fitting (cf. [Konolige, 1988]) that if we have the validity relation of modal S5, then the set of all sentences valid in an S5 structure is stable.

Proposition 5.6 (Konolige, Moore, Fitting). Let \mathfrak{M} be a structure. Then the set $T_{\mathfrak{M}} =_{def} \{ \Phi \mid \mathfrak{M} \models_{\overline{\mathsf{M}}\mathsf{K}} \Phi \}$ is stable.

Extensions can be characterised by self-referential equations. We shall now give such an equation for S5-based extensions.

Theorem 5.6. $X \Vdash_{\mathbb{E}} T$ if and only if $T = \operatorname{Th}_{S5}(X \cup \{\Diamond A \mid \neg A \in \mathcal{L} \setminus T\})$.

Proof. Let RHS denote the right hand side (i.e $Th_{S5}(X \cup \{\Diamond A \mid \neg A \in \mathcal{L} \setminus T\})$) of the above equation.

- '⇒'
- $RHS \subseteq T$ $X \Vdash_{\mathbb{E}} T$ implies that T is stable and consistent. From stability we can conclude that $\neg A \notin T$ implies $\Diamond A \in T$, thus T contains all formulas which are added to the right hand side (RHS) of the above equation via the condition $\{\Diamond A \mid A \in \mathcal{L} \text{ and } \neg A \notin T\}$. Moreover, T is closed under Th_{S5} .
- $T \subseteq RHS$ It holds that $Th_{S5}(X) \subseteq RHS$ and since $X \subseteq T$ and T is closed under Th_{S5} we have $Th_{S5}(X) \subseteq T$. It remains to show that $T \setminus Th_{S5}(X) \subset$ RHS. But this can be reduced to showing that each formula of the form $\Diamond \Phi \in T \setminus Th_{S5}(X)$ is in RHS. The proof can be carried out by induction on the degree of Φ ; we have to distinguish the following cases:
 - $\Phi \in \mathcal{L} \text{ In this case } \Diamond \Phi \text{ is added to the RHS via the condition } \{ \Diamond A \mid A \in \mathcal{L} \text{ and } \neg A \notin T \}$
 - $\Phi \equiv \Diamond \Phi'$ In this case we must ensure that $\Diamond \Diamond \Phi'$ is added to the RHS. By induction hypothesis we have $\Diamond \Phi' \in \text{RHS}$, thus by $\Diamond \Psi \to \Box \Diamond \Psi$ (Axiom 5) we have that $\Box \Diamond \Phi' \in \text{RHS}$. Since RHS is closed under Th_{S5} we have $\Diamond \Diamond \Phi' \in \text{RHS}$.
 - $\Phi \equiv \Box \Phi'$ We have to show that $\Diamond \Box \Phi' \in \text{RHS}$. By induction hypothesis, $\Box \Phi' \in \text{RHS}$ and, again, since $\Psi \to \Diamond \Psi$ is a theorem of S5 and RHS is closed under Th_{S5} , we have $\Diamond \Box \Phi' \in \text{RHS}$.

'⇐' We have to show that there is \mathfrak{M} such that $\mathfrak{M} \models_{\overline{\mathsf{M}}\mathsf{K}} X$ and $T = \{ \varPhi \mid \mathfrak{A} \models_{\overline{\mathsf{M}}\mathsf{K}} \varPhi \}$. It is sufficient to prove that $Th_{S5}(T) \neq \mathcal{L}_M$ because then such an \mathfrak{M} exists. It is clear that for any T we have: $\Diamond A \in T$ iff $\neg A \notin \mathcal{L} \setminus T$. Thus T is consistent or it does not exist. \Box The above theorem gives a syntactical characterisation of the sentences valid in a preferred model of X. It can be easily modified to characterise those sentences which are valid in a strongly preferred model of X (cf. Definition 5.8). Remember that a preferred model \mathfrak{A} is said to be a strongly preferred model of X if and only if \mathfrak{A} is a preferred model of all applicable formulas in \mathfrak{A} .

Theorem 5.7. Let X be in ordered MCNF. $T_{\mathfrak{A}}$ is the set of all formulas valid in some strongly preferred model \mathfrak{A} of X if and only if

$$T_{\mathfrak{A}} = \mathrm{Th}_{\mathrm{S5}}(X' \cup \{\Diamond A \mid \neg A \in \mathcal{L} \setminus T_{\mathfrak{A}}\}) = \mathrm{Th}_{\mathrm{S5}}(X \cup \{\Diamond A \mid \neg A \in \mathcal{L} \setminus T_{\mathfrak{A}}\})$$

where $X' = X \setminus \{ \Phi \mid \Phi \text{ is inapplicable w.r.t. } T_{\mathfrak{A}} \}$

Proof. By Definition 5.8 we have $\mathfrak{A} \models_{MK} X \Leftrightarrow \mathfrak{A} \models_{MK} X'$ and by Theorem 5.6 this yields

$$\mathfrak{A} \models_{\mathsf{MK}} X \Leftrightarrow \mathfrak{A} \models_{\mathsf{MK}} X \cup Th_{S5}(X' \cup \cdots) \Leftrightarrow \mathfrak{A} \models_{\mathsf{MK}} X' \Leftrightarrow \mathfrak{A} \models_{\mathsf{MK}} X' \cup Th_{S5}(X' \cup \cdots)$$

With $X \subseteq Th_{S5}(X \cup \cdots)$ and $X' \subseteq Th_{S5}(X' \cup \cdots)$ this yields

$$\mathfrak{A} \models_{\mathsf{MK}} Th_{S5}(X \cup \cdots) \Leftrightarrow \mathfrak{A} \models_{\mathsf{MK}} Th_{S5}(X' \cup \cdots)$$

and hence, $Th_{S5}(X \cup \cdots) = Th_{S5}(X' \cup \cdots) = T_{\mathfrak{A}}$.

The notion of extensional entailment as well as the syntactical characterisation of extensions provide the basis for a comparison between our logic and two other nonmonotonic modal formalisms.

5.4.2 Autoepistemic Logics

Autoepistemic logics (AEL) can be considered as a semantical approach to common sense reasoning (contrary to syntactical approaches like Default Logic or the logics by Doyle and McDermott).

As already mentioned, autoepistemic logic uses the concept of extension to characterise the possible belief states. A set of sentences $T \subset \mathcal{L}_M$ is an *AE*-extension of a set (of initial premises) X if and only if

$$T = Cn_{cl}(X \cup \{\Box \Phi \mid \Phi \in T\} \cup \{\neg \Box \Phi \mid \Phi \notin T\})$$

 Cn_{cl} denotes the consequence operator of classical propositional logic.

AE-extensions can be characterised by stable sets which have the additional property of being *grounded*.

Definition 5.11 (Grounded). A set T is grounded in X if and only if

$$T \subseteq Cn_{cl}(X \cup \{\Box \Phi \mid \Phi \in T\} \cup \{\neg \Box \Phi \mid \Phi \notin T\}).$$

Groundedness guarantees that the only beliefs an agent has are those from his initial premises and those required by the stability conditions. The following proposition can be found in [Lukaszewicz, 1990].

Proposition 5.7. T is an AE-extension of X if and only if

1. T is grounded in X 2. T is stable 3. $X \subseteq T$

Proof. see [Lukaszewicz, 1990], Theorem 4.62.

The main result of this subsection is that if T is an autoepistemic extension of X and T contains all instances of the modal axiom scheme \mathbf{T} , then $X \models T$ (Theorem 5.8). We already know by Proposition 5.6 that the set of all sentences which are valid in a structure \mathfrak{A} is stable. Note, that the converse of Proposition 5.6 does not hold because there are stable sets $T \subset \mathcal{L}_M(\Sigma)$ (e.g. an AE-extension of $\{A, \Diamond \neg A\}$) which do not have a model from the set of all S5-structures STRUCT(Σ). However, if T contains all instances of the modal axiom scheme \mathbf{T} and is closed under the application of modus ponens and the rule of necessitation, then the converse of Proposition 5.6 does hold.

Definition 5.12 ($\mathbf{T}_{\mathfrak{A}}, \mathsf{T}_{\mathbf{X}}$). Let Σ be a signature, $\mathfrak{A} \in \mathrm{STRUCT}(\Sigma), X \subseteq \mathcal{L}_M(\Sigma)$. Define

$$T_{\mathfrak{A}} =_{_{def}} \{ \Phi \mid \mathfrak{A} \models_{_{MK}} \Phi \} \text{ and}$$
$$T_{X} =_{_{def}} \{ T_{\mathfrak{A}} \mid \mathfrak{A} \in \text{STRUCT}(\Sigma) \text{ and } \mathfrak{A} \models_{_{WK}} X \}$$

Lemma 5.4. Let Σ be a signature, $X \subset \mathcal{L}_M(\Sigma)$ and S_X be the set of all consistent stable theories T which contain X, every instance of the axiom scheme \mathbf{T} and which are closed under the rule of necessitation. Then, $\mathsf{T}_X = \mathsf{S}_X$.

Proof. $\mathsf{T}_{\mathsf{X}} \subseteq \mathsf{S}_X$ is immediately clear, since every $T_{\mathfrak{A}} \in \mathsf{T}_{\mathsf{X}}$ is a stable theory containing X. For the converse direction, assume that there is $T \in \mathsf{S}_X$ such that $T \notin \mathsf{T}_{\mathsf{X}}$. Thus, for all \mathfrak{A} such that $\mathfrak{A} \models_{\mathsf{M}\mathsf{K}} X$ we have $\mathfrak{A} \not\models_{\mathsf{M}\mathsf{K}} T$. This means that $T \cup X$ does not have a model. Hence, T cannot be consistent, because $X \subset T$ – a contradiction.

Theorem 5.8. Let $X, T \subset \mathcal{L}_M$. T is an AE-extension of X which contains all instances of the axiom scheme **T** and which is closed under the rule of necessitation if and only if $X \Vdash_{\mathbb{E}} T$.

Proof. ' \Rightarrow ': We assume that X is consistent (for inconsistent X we have immediately $X \Vdash T$).

- T is an AE-extension
- \Rightarrow T is stable (by Proposition 5.7)
- $\Rightarrow T \in S_X (S_X \text{ is defined in Lemma 5.4})$
- \Rightarrow $T \in \mathsf{T}_X$ by Lemma 5.4
- $\Rightarrow \quad \text{ex. } \mathfrak{A} \text{ such that } \mathfrak{A} \models_{\mathsf{M}_{\mathsf{K}}} X \text{ and } T = \{ \varPhi \mid \mathfrak{A} \models_{\mathsf{M}_{\mathsf{K}}} \varPhi \} \qquad \mathbf{0} .$

We have to show that there is also a preferred model with the above property (\bullet) . We assume the contrary, i.e. we assume that for every \mathfrak{A} for which (\bullet) holds there is $\mathfrak{A}', \mathfrak{A} \prec \mathfrak{A}'$ and $\mathfrak{A}' \models_{\mathsf{MK}} X$. Thus,

- $\Rightarrow \quad \text{ex. } \varPhi \in \mathcal{L} \text{ such that } \mathfrak{A} \models_{_{\overline{\mathsf{M}}\mathsf{K}}} \Box \varPhi \text{ and } \mathfrak{A}' \models_{_{\overline{\mathsf{M}}\mathsf{K}}} \neg \Box \varPhi$
- $\Rightarrow \Phi, \Box \Phi \in T \text{ and } X \cup \{\neg \Phi\} \text{ is consistent}$
- \Rightarrow T cannot be grounded in X
- \Rightarrow T is no AE-extension of X (by Proposition 5.7) a contradiction.

Hence, there is a preferred model of X in which T is valid. Since T is stable it follows that T is the set of all formulas being valid in this preferred model. Therefore $X \Vdash_{E} T$.

'⇐': We have $X \Vdash_{\mathbb{E}} T$. Thus T is stable by Proposition 5.6. It remains to show that T is grounded in X, i.e. we have to show

$$T \subseteq Cn_{cl}(X \cup \{\Box \Phi \mid \Phi \in T\} \cup \{\neg \Box \Phi \mid \Phi \notin T\}).$$

By the prerequisite we know that T contains all instances of **T**. Hence we have $T \subseteq Cn_{cl}(X \cup ...)$. Therefore, T is an AE-extension of X.

Theorem 5.8 relates AEL to our logic via semantical terms like extensional entailment. Let us now look at the syntactical aspects which become clearer when looking at a syntactical characterisation (like the one of Theorem 5.6) of autoepistemic logic. Konolige showed that there is a close correspondence between AEL and the modal system K45. Let Th_{K45} be the syntactical consequence operator of modal K45 (cf. Proposition 5.2). We say that Φ ($\Phi \in \mathcal{L}_M$)
is strongly K45-provable from a set $X \subseteq \mathcal{L}_M$ if and only if there are formulas $\Phi_1, \ldots, \Phi_n \in X$ such that $\Phi_1 \wedge \ldots \wedge \Phi_n \to \Phi \in Th_{K45}(X)$; denote the set of all sentences which are strongly K45-provable from X by $Th_{K45}^S(X)$. T is an autoepistemic extension of X if and only if

$$T = Th_{K \downarrow 5}^{S}(X \cup \{ \Diamond A \mid \neg A \in \mathcal{L} \setminus T \} \cup \{ \Box A \mid A \in T \cap \mathcal{L} \})$$

If we compare the above characterisation with Theorem 5.6 we can see that the main difference between our logic and autoepistemic logic is the absence of the axiom scheme **T**. Interestingly, both logics add only all *nonmodal* sentences which are consistent with the initial set X.

5.4.3 Nonmonotonic Logic II

McDermott's notion of an NML2-S5-Extension can be written as

Definition 5.13 (NML2-S5-Extension). Let $X, T \subset \mathcal{L}_M$. T is an NML2-S5extension of X if and only if

$$T = Th_{S5}(X \cup \{ \Diamond \Phi \mid \Phi \in \mathcal{L}_M \text{ and } \neg \Phi \notin T \})$$

In the above definition of an extension, \Diamond plays the role of something like 'it is S5-consistent, that ...' where Φ is S5-consistent with X means that $\neg \Phi \notin Th_{S5}(X)$. This is exactly the drawback we discussed when trying to find a definition of the entailment relation in Section 5.1.2. This deficiency leads to the collapse of nonmonotonic S5 to monotonic S5:

 $\bigcap \{S \mid S \text{ is an NML2-S5-extension of } X\} = Th_{S5}(X),$

If we compare the syntactical characterisation of NML2-S5-extensions of X with syntactical characterisation by Theorem 5.6, then we see that the only difference is that McDermott forces *all* sentences which are consistent with an extension to enter the extension, i.e. $\Phi \in \mathcal{L}_M$ and $\neg \Phi \notin T$, whereas we do only require that T contains all *nonmodal* sentences which are consistent with T.

Theorem 5.9. Let $X, T \subset \mathcal{L}_M$. If $X \Vdash T$ then T is an NML2-S5-extension.

Proof. Follows from Theorem 5.6 and by the fact that $\mathcal{L} \subset \mathcal{L}_M$.

5.5 Historical Remarks and Related Work

The idea of maximising S5 structures to obtain a logic of minimal knowledge (or maximal ignorance) originates in the work of Halpern and Moses (1984). They restrict themselves to sets of formulas with exactly one preferred model. These sets are called *honest*. To see why they are said to be honest consider the set $X = \{\Box A \lor \Box B\}$, i.e. an agent claims that he knows A or that he knows B. But then we would expect at least A or B to be in the agent's inferential closure. Since $A, B \notin Th_{S5}$ and $A, B \notin Cn_{MK}$ we say that the agent is dishonest because he claims to know A or to know B but actually he does not. Halpern and Moses therefore chose $\parallel_{\rm E}$ as entailment relation.

In 1991 Lifschitz picked up the ideas of Halpern and Moses and related full MK to Logic Programming and Default Logic. He showed that logic programs can be seen as MK-theories if we extend the language by a negation-as-failure operator.

Again several years later, in 1996, del Cerro, Delgado and Herzig developed a theory to talk about consistency , [del Cerro et al., 1996] . The basic aim of their logic is quite similar to that of MK. But in order to guarantee that the unknown-operator works correctly they restrict the inference relation to be relation among propositional formulas and modal formulas. As we have seen in Proposition 5.4 this yields that there is always a unique preferred model. Hence, their logic is a special case of MK.

5.6 Conclusion

We showed that the unknown-operator, as used for example in practical AI systems, can be given a semantics which is based on a subset $|\underline{\bullet}_{MK}$ of S5's validity relation $|\underline{\bullet}_{MK}$. The resulting logic MK is a generalisation of [Halpern and Moses, 1984]. We solved the problem of multiple preferred models by showing that there is a modal language such that X has a unique preferred model if and only if X is semantically equivalent to some subset of this language. Moreover we proved that Cn_{MK} is a nonmonotonic preclosure operator, for which several weaker versions of the compactness theorem holds.

The relation to Moore's Autoepistemic Logic (AEL) and McDermott's NML-2 is given by showing that there is a syntactical characterisation of Cn_{MK} which is located exactly between NML-2 and AEL. We thus have reached our goal to provide a reasonable formal basis for a semantics of reasoning systems which supply an unknown-operator.

CHAPTER 6

Combining K_3 and MK

I shall now combine the paraconsistent logic K_3 with the logic MK of minimal knowledge. The resulting logic MK_3 should serve as a tool for reasoning with paraconsistent and unknown information. The main results of this chapter are

- $-MK_3$ is a faithful reformulation of Belnap's ideas of how a computer should answer questions.
- Contrary to ${\bf L4}$ the logic $M\!K_3$ handles implications by means of the material conditional.
- If the input contains only literals, then the answers generated by MK_3 are identical to the ones generated by L4. However, in general L4 and MK_3 are incomparable.

This chapter is organised as follows: we shall first investigate the semantical entailment relation for MK_3 . Next we shall discuss properties for MK_3 . We shall pay a special attention to normal forms and to the notion of a stable set. We conclude with relating MK_3 to Belnap's **L4** and to the problem of logical omniscience.

6.1 Semantical Entailment

The basic idea is to replace the two-valued interpretations in the logic MK by three-valued ones. There are, however a few points which deserve special attention. First, since every subset of our propositional language \mathcal{L} has a three-valued model we cannot simply say that A is unknown w.r.t. the database X if A is satisfiable in some K₃ model of X. That is we have to reformulate our concept of 'unknown' for the paraconsistent case. Second, we have also to think about the concept of knowledge in the presence of contradicting information. For example, when given $\{A, \neg A\}$ is it reasonable to accept $\Box A$ and $\Box \neg A$, i.e. can we say that we know A or $\neg A$? In other words, do we accept the Rule of Necessitation as a legal pattern of reasoning in the presence of contradictions?

We claimed in Chapter 2 that the truth-status of contradicting information is extremely vague. Therefore, when given $X = \{A, \neg A\}$ the reasoner should admit that he or she does not really know much about A, because she cannot say anything definite about A's truth-status after a revision process. We therefore intend that $\Box A$ should be entailed by X if and only if A is true in every preferred model of X. Hence, 'knowing A' ($\Box A$) means A holds and that it cannot be suspected to be paraconsistent. This ensures that A will still be known after a revision process.

As for 'unknown', we simply could have taken the opposite of 'known' which would yield that A is unknown w.r.t. to the above X, because A is not consistently known. This means that $\Diamond A$ holds in the very moment where we can imagine a state where A holds, no matter whether A holds consistently in this state.

It is important to note that even though this semantics for \Diamond might be debatable¹, it has two advantages. First, the operator \Box and \Diamond are interdefinable as in classical modal logic and second, if X contains no contradiction then the above will still yield the same consequences as Cn_{MK} .

Definition 6.1 (Three-valued Kripke Structure, Validity). A three-valued Kripke structure \mathfrak{A} is a tuple (M, V) where M is a nonempty index set and V is a family of three-valued valuation functions such that for each $\alpha \in M$ there is some $I_{\alpha} \in V$ with $I_{\alpha} : \operatorname{ATOM}(\Sigma) \to \{t, f, \top\}$.

We say that $\mathfrak{A} \models_{\overline{\mathsf{M}}_{\mathsf{K}}} \Phi$ if and only if $\mathfrak{A} \models_{\overline{\mathsf{M}}_{\mathsf{K}_{\alpha}}} \Phi$ for all $\alpha \in M$.

The connectives \rightarrow , \lor are introduced by defining

¹ An alternative would be to require that $\Diamond \Phi$ holds, if there is at least one state in which Φ holds consistently. Another one would be to require that Φ does in every state take a value from t or f. In both cases, \Diamond and \Box are no longer interdefinable.

Remark 6.1 (Interdefinability). For any Φ , $\mathfrak{A} \models_{\overline{\mathsf{M}} \underline{\mathsf{h}}} \Box \Phi$ if and only if $\mathfrak{A} \models_{\overline{\mathsf{M}} \underline{\mathsf{h}}} \neg \Diamond \neg \Phi$.

The definition of a preferred three-valued Kripke model is a little bit more complicated in this case. In a first step, we consider only three-valued Kripke models which are maximal with respect to the S5-substructure relation \preccurlyeq (note that Definition 5.2 defines a substructure relation between three-valued Kripke-structures). Consider for example $X = \{A, A \rightarrow B\}$ and $\Sigma = \{A, B, C\}$. The following is a \preccurlyeq -maximal model for X: $\mathfrak{A} = (\{1, \ldots, 12\}, \{I_1, \ldots, I_{12}\})$ with

$$\begin{array}{c|c} I_1(A) = t & |I_2(A) = t & |I_3(A) = t & |I_4(A) = \top & |I_5(A) = \top & |I_6(A) = \top \\ I_1(B) = t & |I_2(B) = t & |I_3(B) = t & |I_4(B) = f & |I_5(B) = f & |I_6(B) = f \\ I_1(C) = f & |I_2(C) = t & |I_3(C) = \top & |I_4(C) = f & |I_5(C) = t & |I_6(C) = \top \\ \end{array}$$

$$\begin{split} &I_{7}(A) = \top \begin{vmatrix} I_{8}(A) = \top & I_{9}(A) = \top \\ &I_{7}(B) = t \\ &I_{7}(C) = f \end{vmatrix} \begin{vmatrix} I_{8}(B) = t & I_{9}(B) = t \\ &I_{9}(C) = \top \end{vmatrix} \begin{vmatrix} I_{10}(A) = \top & I_{11}(A) = \top \\ &I_{10}(B) = \top \\ &I_{10}(C) = f \end{vmatrix} \begin{vmatrix} I_{11}(B) = \top & I_{12}(A) = \top \\ &I_{12}(B) = \top \\ &I_{12}(C) = \top \end{vmatrix}$$

In order to select a preferred three-valued Kripke model, we take all \preccurlyeq -maximal substructures \mathfrak{A}' from \mathfrak{A} such that no world in \mathfrak{A}' overinterprets a sentence in X. That is $\mathfrak{A}' = (M', V')$ where $M \subseteq M'$ and

 $V' = \{I_{\alpha} \mid \text{ there is no } I_{\beta} \in V \text{ such that } I_{\beta} \sqsubset I_{\alpha}\}$

In the above example this yields that $\mathfrak{A}' = (\{1,2\},\{V_1,V_2\})$ is the only preferred model of X.

Let us formulate the above ideas more precisely. Consider the following relation \leq among Kripke-structures.

Definition 6.2 (\leq). Let $\mathfrak{M} = (M, V)$ and $\mathfrak{N} = (N, W)$ and define

 $\mathfrak{M} \leq \mathfrak{N} =_{def} M \subseteq N$ and for every $I \in N$ there is some $J \in W$ such that $I \sqsubseteq J$.

As usual we write $\mathfrak{M} < \mathfrak{N}$ if and only if $\mathfrak{M} \leq \mathfrak{N}$ and NOT $\mathfrak{N} \leq \mathfrak{M}$.

Clearly, \leq is a partial ordering, i.e. reflexive, transitive and anti-symmetric. Consider again the above example. Here we have that $\mathfrak{A}' < \mathfrak{A}$. Moreover, \mathfrak{A}' is <-minimal.

We shall now define a partial ordering relation among three-valued Kripke structures:

Definition 6.3 (\subseteq). Let \mathfrak{M} , \mathfrak{N} be two models of a given set X. We say that \mathfrak{M} is preferred over \mathfrak{N} , denoted by $\mathfrak{M} \subseteq \mathfrak{N}$, if and only if

- 1. there is a \preccurlyeq -maximal model \mathfrak{A} of X such that $\mathfrak{N} \preccurlyeq \mathfrak{A}$ and $\mathfrak{M} < \mathfrak{A}$ is <-minimal, or
- 2. \mathfrak{M} is isomorphic to \mathfrak{N} .

The relation \in is a partial ordering.

We say that \mathfrak{M} is a *preferred three-valued Kripke model* of X, denoted by $\mathfrak{M}|_{\overline{\mathfrak{M}}_{\mathfrak{H}}} X$ if and only if $\mathfrak{M}|_{\overline{\mathfrak{M}}_{\mathfrak{H}}} X$ and there is no three-valued Kripke model \mathfrak{N} of X such that $\mathfrak{N} \in \mathfrak{M}$.

Note, that the partial ordering \in is very sparse: if for any two MK₃ models $\mathfrak{M}, \mathfrak{N}$ of X we have $\mathfrak{M} \in \mathfrak{N}$, then \mathfrak{M} is isomorphic to \mathfrak{N} or \mathfrak{M} is a preferred model. Any preferred model of X can be seen as containing only the \square -minimal states of \mathfrak{A} , where \mathfrak{A} is the \preccurlyeq -maximal model of Definition 6.3.

The following lemma is the three-valued Kripke analogue of Lemma 5.1 and for the records.

Lemma 6.1. Let
$$\mathfrak{A} \models X, \mathfrak{A} \models \Phi$$
. Then $\mathfrak{A} \models X \cup \{\Phi\}$.

Proof. Analogous to Lemma 5.1.

As usual we build up the entailment relation by means of preferred models.

Definition 6.4 (MK₃-Entailment). We say that X MK₃-entails Φ , denoted by $X \models_{\mathsf{M}_{\mathfrak{H}}} \Phi$, if and only if $\mathfrak{A} \models_{\mathsf{M}_{\mathfrak{H}}} X$ implies $\mathfrak{A} \models_{\mathsf{M}_{\mathfrak{H}}} \Phi$. Further, $\operatorname{Cn}_{\mathsf{M}_{\mathfrak{H}}}(X) =_{def} \{\Phi \mid X \models_{\mathsf{M}_{\mathfrak{H}}} \Phi\}$.

As another example consider

Example 6.1. Let $X = \{(\Diamond A \lor \Diamond B) \land \neg(\Diamond A \land \Diamond B)\}$, i.e. either $\Diamond A$ holds or $\Diamond B$ holds. This set has two preferred MK₃-models:

$\mathfrak{A}_1: I_1(A) = f$	$\mathfrak{A}_2: I_1(A) = f$
$I_1(B) = f$	$I_1(B) = f$
$I_2(A) = f$	$I_2(A) = t$
$I_2(B) = t$	$I_2(B) = f$

The following is a more complex example.

Example 6.2.

$$X = \{A, A \to B, A \to \neg B\}$$

The set X has a unique preferred MK_3 -model $\mathfrak{A} = (\{1, 2, 3\}, V)$ with

$$V = \{ I_1(A) = \top | I_2(A) = \top | I_3(A) = t \\ I_1(B) = t | I_2(B) = f | I_3(B) = \top \}$$

Hence, $(\Diamond A \land \Diamond \neg A)$ and $(\Diamond B \land \Diamond \neg B)$ is a consequence of X. This is reasonable, because both A and B can be suspected to be paraconsistent.

Now consider $Y = X \cup \{ \Diamond \neg A \to C \}$. The set Y has two preferred models: $\mathfrak{A}_1 = (\{1, 2, 3\}, V_1)$ and $\mathfrak{A}_2 = (\{1, 2\}, V_2)$ where

$$V_{1} = \{ \begin{array}{c} I_{1}(A) = \top & | I_{2}(A) = \top & | I_{3}(A) = t \\ I_{1}(B) = t & | I_{2}(B) = f & | I_{3}(B) = \top \\ I_{1}(C) = t & | I_{2}(C) = t & | I_{3}(C) = t \} \\ \end{array} \right| V_{2} = \{ \begin{array}{c} I_{1}(A) = t & | I_{2}(A) = t \\ I_{1}(B) = \top & | I_{2}(B) = \top \\ I_{1}(C) = t & | I_{2}(C) = f \\ \end{array} \right)$$

Here we have that $\mathfrak{A}_2 \not\models_{\mathsf{M}_s} \Diamond \neg A$ which means that we cannot conclude $\Diamond A \land \Diamond \neg A$ from Y, i.e. we cannot conclude that A is unknown.

Whether one considers the above result to be intuitive or not, depends on how we wish the additional formula $\Diamond \neg A \rightarrow C$ to be read. One possibility is to resolve material implication and to read it as a disjunction. This yields that we claim: *C* holds or *A* is consistently known ($\Diamond \neg A \rightarrow C = \neg \Diamond \neg A \lor C = \Box A \lor C$). If we choose this reading, then Cn_{MK} operates correctly.

If, however, we want to read it like an inference rule, then $\Diamond \neg A \notin \operatorname{Cn}_{\mathsf{M}\mathsf{g}}(Y)$ is not intuitive, because Y still contains the contradiction between $A, A \to B$ and $A \to \neg B$. The solution is then to consider a strongly preferred model like we did in Chapter 5.

I have the impression that reading \rightarrow as an inference rule is not very convincing in the presence of inconsistency. For example, consider $X = \{A, \Diamond B \rightarrow \neg A\}$. The set X has exactly one preferred model $\mathfrak{A} = (\{1\}, V_1)$ and $V_1(A) = t$, $V_1(B) = f$. Hence, $X \Vdash_{\mathsf{Mg}} \Diamond B$. If we read $\Diamond B \rightarrow \neg A$ as an inference rule we would unnecessarily add a new contradiction. Hence, we prefer to read the implication as a disjunction which, in the above case says: $\neg B$ is consistently known or $\neg A$ holds $(\Box \neg B) \lor \neg A$.

6.2 Properties

6.2.1 Basic Closure Properties

Proposition 6.1. $\operatorname{Cn}_{\mathsf{M}\mathsf{K}}$ is a cumulative and nonmonotonic pre-closure operator which satisfies the AND-property, i.e. $\Phi \wedge \Psi \in \operatorname{Cn}_{\mathsf{M}\mathsf{K}}(X)$ if and only if $\Phi, \Psi \in \operatorname{Cn}_{\mathsf{M}\mathsf{K}}(X)$. **Theorem 6.1 (Preferred Model Existence).** If X has an MK_3 -model then X has a preferred MK_3 -model.

Proof. The proof is not very complicated but requires two subproofs by structural induction, which are similar to Theorem 5.1. Let $\mathfrak{M}|_{\overline{\mathfrak{M}}_{\mathsf{K}}} X$. We first have to show that there is a \preccurlyeq -maximal model \mathfrak{A} of X. Consider the chain

$$\mathfrak{M} \preccurlyeq \mathfrak{N}_1 \preccurlyeq \mathfrak{N}_2 \preccurlyeq \dots$$

and let $\mathfrak{N}_i = (N_i, W_i)$. Define $N =_{def} \bigcup N_i$ and $W =_{def} \bigcup W_i$ and let $\mathfrak{N} = (N, V)$. Clearly, $\mathfrak{N}_i \preccurlyeq \mathfrak{N}$. We have to show that $\mathfrak{N} \models_{\overline{\mathsf{M}} \mathsf{K}} X$. Let $\Phi \in X$. If Φ is atomic, we are done. For the inductive step we have to distinguish several cases:

- $\Phi = \Diamond \varphi$. For every \mathfrak{N}_i there must be a state $\alpha \in N_i$ such that $\mathfrak{N}_i \models_{\mathfrak{M}_{\mathfrak{N}\alpha}} \varphi$. By construction of \mathfrak{M} and the induction hypothesis we have $\mathfrak{N}|_{\overline{\mathbb{M}}_{k_{\alpha}}}\varphi$ and hence,
 $$\begin{split} \mathfrak{N} & \underset{\mathsf{M} \not\models}{\underset{\mathsf{W} \not\models}{\underset{\mathsf{V}}{\models}}} \Diamond \varphi. \\ \varPhi &= \varphi_1 \lor \varphi_2 \quad \text{By induction hypothesis.} \end{split}$$
- $\Phi = \neg \varphi$ For every \mathfrak{N}_i and every $\alpha \in N_i$ we have $\mathfrak{N}_i \not\models_{\mathsf{M}_{\mathfrak{N}_{\alpha}}} \varphi$. Suppose that there is some world $\alpha \in N$ such that $\mathfrak{N} \models_{\overline{\mathsf{Ms}}_{\alpha}} \varphi$. By the construction of N and the induction hypothesis we have that there must be some model \mathfrak{N}_i of the above chain such that $\mathfrak{N}_i \not\models_{\mathfrak{M}_{\mathfrak{N}_{\mathfrak{N}}}} \varphi$ and hence \mathfrak{N}_i cannot be a model for φ – a contradiction.

Hence, by Zorn's Lemma there is a \preccurlyeq -maximal model \mathfrak{A} of X. Next, we have to show that there is some \mathfrak{C} such that $\mathfrak{C} < \mathfrak{A}$ is <-minimal.

We have to show that every chain

$$\mathfrak{A} > \mathfrak{B}_0 > \mathfrak{B}_1 > \mathfrak{B}_2 > \ldots$$

has a lower bound. Let $\mathfrak{B}_i = (M_i, V_i)$. Define $M =_{def} \bigcap M_i$ and $V =_{def} \bigcap V_i$ and let $\mathfrak{B} = (M, V)$. Clearly, $\mathfrak{B} < \mathfrak{B}_i$. It remains to show that $\mathfrak{B} \models_{\overline{\mathsf{MS}}} X$. Let $\Phi \in X$. The base case where Φ is atomic is again immediate. For the inductive step we distinguish again the following cases:

- $\Phi = \Diamond \varphi$. For every \mathfrak{B}_i there must be a state $\alpha \in M_i$ such that $\mathfrak{B}_i \models_{\overline{\mathsf{MS}}_{\alpha}} \varphi$. By construction of \mathfrak{B} and the induction hypothesis we have $\mathfrak{B} \models_{\mathfrak{M}_{\mathfrak{H}_{\alpha}}} \varphi$ and hence, $\mathfrak{B} \models_{\mathsf{M}\mathsf{K}} \Diamond \varphi.$ $\varPhi = \varphi_1 \lor \varphi_2$ By induction hypothesis.
- $\Phi = \neg \varphi$ For every \mathfrak{B}_i and every $\alpha \in M_i$ we have $\mathfrak{B}_i \not\models_{\mathfrak{M}_{\mathfrak{S}_{\alpha}}} \varphi$. By the construction of M and the induction hypothesis we have that $\mathfrak{B} \not\models_{\mathsf{M}_{\mathfrak{H}_{\alpha}}} \varphi$. Hence, $\mathfrak{B} \models_{\mathsf{M}_{\mathfrak{H}_{\alpha}}} \neg \varphi$.

We conclude by Zorn's Lemma that there is a model \mathfrak{C} of X such that $\mathfrak{C} < \mathfrak{A}$ is <-minimal. П

The following proposition relates MK_3 to properties of System C (cf. 2).

Proposition 6.2. Let X be a set of formulas.

- $\begin{array}{ll} 1. \ If \ \Phi, \Upsilon \Vdash_{\overline{\mathsf{M}} \S} \Phi \ and \ \Phi \Vdash_{\overline{\mathsf{M}} \S} \Upsilon \ then \ \Phi \Vdash_{\overline{\mathsf{M}} \$} \Psi \ (Cutty). \\ 2. \ If \ \Phi \leftrightarrow \Upsilon \ is \ an \ \mathsf{MK}_3 \text{-tautology and } \Phi \Vdash_{\overline{\mathsf{M}} \$} \Psi \ then \ \Upsilon \Vdash_{\overline{\mathsf{M}} \$} \Psi \ (Left \ Logical \ Equivalence). \end{array}$ alence).

Proof.

Ad 1) Let $\mathfrak{A} \models \mathfrak{A} \longrightarrow \mathfrak{A}$. We then know by $\varPhi \models \mathfrak{A} \longrightarrow \mathfrak{A}$ that $\mathfrak{A} \models \mathfrak{A} \wedge \Upsilon$ and since every preferred model of $\Phi \wedge \Upsilon$ is preferred model of Ψ we have that $\mathfrak{A} \models_{\overline{\mathsf{M}}\mathsf{b}} \Psi$. Thus, $\Phi \to \Psi$.

Ad 2) Immediately.

MK₃-Tautologies 6.2.2

A formula Φ is an MK₃-tautology if and only if Φ holds in all MK₃-structures. The set of MK₃-tautologies does, however, not coincide with the set of S5 tautologies. As a counter-example consider the single world structure $\mathfrak{A} = (\{1\}, \{I_1(A)\})$ \top }). \mathfrak{A} is a model of $\{A, \neg A\}$. Clearly $A \to A$ is valid in \mathfrak{A} but $\Box(A \to A)$ is not. Even though this matches our intuition about the \Box -operator, it invalidates the Rule of Necessitation (NEC).

The invalidity of NEC implies that there is a difference between the set of MK_3 -tautologies and the MK_3 -consequences of the empty set. The following proposition shows how $\operatorname{Cn}_{MK}(\emptyset)$, MK_3 -tautologies and S5-tautologies relate to each other.

Proposition 6.3. Let T_{S5} be set of S5-tautologies, T_{MK_3} the set of MK₃-tautologies. Then,

- 1. $\operatorname{Cn}_{MK}(\emptyset) = T_{S5}$
- 2. $T_{MK_3} \subseteq T_{S5}$
- 3. $T_{S5} = E$ where E is the smallest set which contains T_{MK_3} and which is closed under NEC.

Proof. The first two points are quite easy. To show (1) simply note that since the empty set is consistent we have that the structure \mathfrak{A} which contains a maximal set of states and no atomic variable takes the value \top in any state of \mathfrak{A} is

the \preccurlyeq -maximal S5 model for the empty set. In other words, \mathfrak{A} is equivalent to the maximal two-valued S5-Kripke-structure. With (1) it follows easily that (2) $T_{\mathsf{MK}_3} \subseteq T_{S5}$.

As for (3), consider the syntactical generation of all S5 tautologies:

$$\Gamma_{0} =_{_{def}} A_{x_{55}}$$

$$\Gamma_{i+1} =_{_{def}} \Gamma_{i} \cup \{ \Psi \mid \Phi, \Phi \to \Psi \in \Gamma_{i} \} \cup \qquad (*)$$

$$\{ \Box \Phi \mid \Phi \in \Gamma_{i} \} \qquad (**)$$

$$T_{S5} =_{def} \bigcup \Gamma_i$$

We proof by induction on the stage i of the construction of T_{S5} that $\Gamma_i \subseteq E$ for every i. It easy to check that the set Ax_{S5} of S5 axioms is valid in every MK_3 -structure. Hence, $\Gamma_0 \subseteq E$. Assume that the proposition holds for every i < n. To see that $\Gamma_n \subseteq E$, note that first, MP is a sound rule of proof and second, E is closed under NEC. Hence, $\Phi, \Phi \to \Psi \in \Gamma_{n-1}$ implies $\Psi \in E$ and $\Phi \in \Gamma_{n-1}$ implies $\Box \Phi \in E$. Thus, $T_{S5} \subseteq E$. The inclusion $E \subseteq T_{S5}$ follows from $E \subseteq Cn_{M5}(\emptyset)$. Hence, $T_{S5} = E$.

6.2.3 Normal forms

We know that every formula $\Phi \in \mathcal{L}_M$ is semantically S5-equivalent to some formula in ordered MCNF. We shall now show that a similar result holds for MK_3 . The proof is similar to the S5-case; see also [Hughes and Cresswell, 1968], p.50ff.

Definition 6.5. Two formulas Φ, Φ' are said to be semantically MK_3 -equivalent, if and only if for every structure \mathfrak{A} and state α we have: $\mathfrak{A} \models_{\overline{M}\mathfrak{H}_{\alpha}} \Phi \Leftrightarrow \mathfrak{A} \models_{\overline{M}\mathfrak{H}_{\alpha}} \Phi'$.

Corollary 6.1. Φ, Φ' are semantically MK₃-equivalent if and only if $\Phi \leftrightarrow \Phi'$ is an MK₃-tautology.

The condition that $\Phi \leftrightarrow \Phi'$ is a tautology is sometimes referred to as *syntactical equivalence*.

Proposition 6.4. For each formula $\Phi \in \mathcal{L}_M$ there is some formula Φ' such that Φ' is semantically MK₃-equivalent to Φ and Φ' is in ordered MCNF.

Proof. We have to show that for every $\Phi \in \mathcal{L}_M$ there is a formula Φ' having the form $\Phi' = D_1 \wedge \ldots \wedge D_n$ where each D_i has the form

$$\alpha \vee \Box \beta_1 \vee \ldots \vee \Box \beta_k \vee \Diamond \gamma_1 \vee \ldots \vee \Diamond \gamma_l$$

where $\alpha, \beta_1, \ldots, \beta_k, \gamma_1, \ldots, \gamma_l$ are nonmodal.

We first show that any modal formula φ is semantically equivalent to a socalled *first-degree* formula. A first-degree formula is a formula which no modal operator is in the scope of another modal operator. Once we have shown that every formula can be reduced to an equivalent formula having modal degree at most 1, it is easy to show that each formula can be transformed in MCNF.

Formally, the modal degree $m(\Phi)$ of a formula $\Phi \in \mathcal{L}_M$ is defined as follows:

$$\begin{split} m(\varPhi) &=_{\scriptscriptstyle def} 0, \text{ for } \varPhi \in \mathcal{L} \\ m(\Box \varPhi) &=_{\scriptscriptstyle def} m(\varPhi) + 1 \\ m(\varPhi \land \Psi) &=_{\scriptscriptstyle def} max(m(\varPhi), m(\Psi)) \\ m(\varPhi \lor \Psi) &=_{\scriptscriptstyle def} max(m(\varPhi), m(\Psi)) \\ m(\varPhi \to \Psi) &=_{\scriptscriptstyle def} max(m(\varPhi), m(\Psi)) \\ m(\neg \varPhi) &=_{\scriptscriptstyle def} m(\varPhi) \end{split}$$

To reduce every formula to formula of degree at most 1 consider the following $\ensuremath{\mathsf{MK}_3}\xspace$ -tautologies:

- 1. $\Box(\Phi \lor \Psi) \to (\Box \Phi \lor \Diamond \Psi)$ 2. $\Box(\Phi \lor \Box \Psi) \leftrightarrow (\Box \Phi \lor \Box \Psi)$
- 3. $\Box(\Phi \lor \Diamond \Psi) \leftrightarrow (\Box \Phi \lor \Diamond \Psi)$
- 4. $\Diamond(\Phi \land \Diamond \Psi) \leftrightarrow (\Diamond \Phi \land \Diamond \Psi)$ $\leftrightarrow(\Diamond \Phi \land \Diamond \Psi)$
- 5. $\Diamond (\Phi \land \Box \Psi) \leftrightarrow (\Diamond \Phi \land \Box \Psi)$

We show that 1) is valid. The proofs for 2) – 5) are analogous. Consider $\mathfrak{A} \models_{\overline{\mathsf{M}} \mathsf{k}} \Box(\varPhi \lor \Psi)$. Thus, for all α we have $\mathfrak{A} \models_{\overline{\mathsf{M}} \mathsf{k}_{\alpha}} \Box(\varPhi \lor \Psi)$. This means that we have two cases:

$$\begin{array}{l} \mathfrak{A} \models_{\overline{\mathsf{M}} \, \mathbb{k}_{\alpha}} \varPhi \ , \, \text{for all } \alpha, \, \text{or} \\ \mathfrak{A} \models_{\overline{\mathsf{M}} \, \mathbb{k}_{\beta}} \varPhi \ , \, \text{for all } \beta \, \text{with} \, \mathfrak{A} \models_{\overline{\mathsf{M}} \, \mathbb{k}_{\beta}} \varPhi \end{array}$$

Given the above tautologies it is easy to show by induction on the modal degree $m(\varphi)$ that we can reduce Φ to an equivalent formula of degree at most 1. The rest follows from [Hughes and Cresswell, 1968]:

- 1. If Φ contains no modal subformulas, then Φ is already in MCNF.
- 2. If φ is a first-degree formula which is equivalent to Φ , we take each formula of the form $\Box\beta$ or $\Diamond\alpha$ as an indivisible unit and reduce this formula to CNF by methods of classical propositional logic. Finally, we replace each occurrence of $\neg \Box\beta$ by $\Diamond \neg\beta$ and $\neg \Diamond\beta$ by $\Box \neg\beta$. The resulting formula is in MCNF.

6.2.4 P-Stable Sets

Let us now see how we can adapt the concept of a stable set, used in Chapter 5, to contradicting information. One requirement was that a stable set should be closed under classical consequence. We can replace classical consequence by K_3 consequence for the paraconsistent case. The next requirement was that if A is contained in a stable set, then $\Box A$ must also be a member of this set. Since $\Box A$ means in the context of MK₃ that A is consistently known, we have to guarantee we only add $\Box A$ if we are absolutely sure that A is not involved in any contradiction. Particularly, we do not accept $\Box A$ when we are given $X = \{A, A \to B, A \to \neg B\}$. Here we cannot guarantee that A is consistent because we can think of a state α in a preferred model of X in which A takes the value \top .

Definition 6.6 (P-Stable). A set X is said to be paraconsistently stable, or short p-stable, if

- 1. $Ax_{S5} \subseteq X$
- 2. $\Phi, \Phi \to \Psi \in X$ and $\Diamond \neg \Phi \notin X$ implies $\Psi \in X$
- 3. $\Phi, \Phi \to \Psi \in X$ and $\Phi \to \Psi$ is a tautology which satisfies NC implies $\Psi \in X$
- 4. $\Phi \in X$ and $\Diamond \neg \Phi \notin X$ implies $\Box \Phi \in X$
- 5. $\Box \neg \Phi \notin X$ implies $\Diamond \Phi \in X$
- 6. For no formula Φ , both $\Box \Phi$ and $\neg \Box \Phi$ are a member of X.

Point 2) and 3) are nothing but the Cautious Modus Ponens whereas 4) corresponds to a cautious version of the Rule of Necessitation. The last point ensures that each p-stable set has an MK_3 -model.

Proposition 6.5. Let X be a set of formulas, \mathfrak{A} a preferred model of X. Then $T_{\mathfrak{A}} =_{def} \{ \Phi \mid \mathfrak{A} \mid_{\overline{\mathbf{k}}_{2}} \Phi \}$ is stable.

Proof. Easy. Simply check that each condition in Definition 6.6 corresponds to a sound rule of inference. \Box

Like ordinary stable sets, p-stable sets are completely characterised by the nonmodal formulas they contain.

Proposition 6.6. Let X, Y be two stable sets which contain the same nonmodal formulas, i.e. $X \cap \mathcal{L} = Y \cap \mathcal{L}$. Then X = Y.

Proof. The proof is similar to the one given in [Meyer and Van der Hoek, 1995] for stable sets. We shall show by induction on the modal degree $m(\Phi)$ of a formula Φ that

$$\Phi \in X \Leftrightarrow \Phi \in Y$$

The modal degree $m(\Phi)$ is defined as in the proof of Proposition 6.4. For nonmodal formulas, the proposition follows immediately. Assume that the induction hypotheses holds for all Φ with $m(\Phi) < n$ and suppose we are given Φ with $m(\Phi) = n$. For every modal Φ there is an equivalent formula Φ' in ordered MCNF (cf. Definition 5.6) which is of modal degree at most $m(\Phi)$ (see[Hughes and Cresswell, 1968], p.117f. for details).

Assume that Φ is in ordered MCNF , i.e. Φ is a conjunction $D_1 \wedge \ldots \wedge D_n$ where each D_i has the form

$$\neg \Box \Psi \lor \Box \varphi_1 \lor \Box \varphi_2 \lor \Box \varphi_k \lor P$$

where Ψ, φ_i, P are nonmodal. Hence for every disjunct δ in D_i we have $m(\delta) < 1$ and therefore

$$\delta \in X \Leftrightarrow \delta \in Y$$

It is easy to see that for any stable set X we have

$$\Box \Phi \lor \Psi \in X \Leftrightarrow \Phi \in X \text{ or } \Psi \in X$$
(i)

$$\neg \Box \varPhi \lor \Psi \in X \Leftrightarrow \varPhi \notin X \text{ or } \Psi \in X, \text{ or } (ii)$$
$$\varPhi, \neg \varPhi \in X \text{ or } \Psi \in X$$

With the help of (i) and (ii) and the induction hypotheses we have $C_i \in X \Leftrightarrow C_i \in Y$, for all C_i and hence $\Phi \in X \Leftrightarrow \Phi \in Y$.

6.3 Relationship to Belnap's Approach

How do MK_3 and Belnap's L4 relate? L4 can be axiomatised by the following sequent style calculus LL4 which has been taken from [Wagner, 1994]. Axioms: All sequents of the form $\Phi \vdash \Phi$.

Rules:

$$\begin{aligned} \frac{X \vdash \Phi \quad X \vdash \Psi}{X \vdash \Phi \land \Psi} (\land -S) & \frac{X, \Phi, \Psi \vdash Y}{X, \Phi \land \Psi \vdash Y} (\land -A) \\ \frac{X \vdash \sim \Phi}{X \vdash \sim (\Phi \land \Psi)} (\sim \land -S) & \frac{X, \sim \Phi \vdash Y \quad X, \sim \Psi \vdash Y}{X, \sim (\Phi \land \Psi)} (\sim \land -A) \\ \frac{X \vdash \sim \Psi}{X \vdash \sim (\Phi \land \Psi)} (\sim \land -A) & \frac{X \vdash \sim \Phi}{X, \sim (\Phi \land \Psi) \vdash Y} (\sim \land -A) \\ \frac{X \vdash \Phi}{X \vdash \sim \sim \Phi} (\sim \sim -I) & \frac{X \vdash \sim \sim \Phi}{X \vdash \Phi} (\sim \sim -E) \\ \frac{X \vdash \Phi}{X \vdash \Psi} \text{ CUT-like} & \frac{X \vdash \Phi}{X, \Psi \vdash \Phi} \text{ Monotonicity} \end{aligned}$$

We can verify that each rule except for Monotonicity is valid for MK_3 . Define the *nonmonotonic fragment* of LL4 to be all rules of the above calculus except Monotonicity.

Observation 6.1. The nonmonotonic fragment of LL4 is contained in MK_3 .

Actually, Belnap does not consider the logic L4 to be final. He develops his ideas further and we shall see how this development relates to MK_3 . After having defined the notion of entailment for L4 Belnap begins to investigate how unknown information can be handled appropriately. As a starting point, he shows that there is a sentence which is not faithfully represented in any set-up s. For example, if 'either the Pirates or the Orioles won' is marked True in any set-up then either 'the Pirates won' or 'the Orioles won' is marked True. Thus, any set-up would lead the computer to answer YES either to question, Did the Orioles win?, or, Did the Pirates win?

Belnap's solution is to use a *collection of set-ups* to represent the sentence 'either P or O'. Thus, the computer's state of mind when representing the above sentence consists of two set-ups:

$$s_1(P) = t \quad s_2(P) = u \quad s_3(P) = t \quad s_4(P) = f s_1(O) = u \quad s_2(O) = t \quad s_3(O) = f \quad s_4(O) = t$$

In order to answer questions like, Did the Pirates win?, Belnap computes the greatest lower bound of $s_1(P) \dots s_4(P)$ with respect to the following lattice:



A collection of set-ups is called *epistemic state*². If E is an epistemic state, then the truth-value of a formula Φ in this epistemic state is defined as

$$E(\Phi) = \sqcap \{ s(\Phi) \mid s \in E \}$$

where \sqcap denotes the meet in the lattice **A4**. If we take $E = \{s_1, \ldots, s_4\}$ we have that E(P) = E(O) = u, i.e. if we put the query, Did the Pirates win?, then the computer answers, I don't know.

The above shows a great similarity to MK_3 . First note, that if we restrict the lattice A4 to t, f, b the resulting semi-lattice is identical to one on which K_3 's preference relation is based (cf. Chapter 2). Moreover, there is great similarity between the way Belnap defines the truth-value in an epistemic state E and the way how K_3 -entailment is defined. If $E(\Phi) = f$ then this means that Φ is at least false in every set-up s of E. This relates nicely to the observation that if $\neg \Phi \in Cn_3(X)$ then Φ is at least false in every preferred three-valued model of X. Whether we try to minimise the information given in X by considering preferred models or by forming $\sqcap \{\ldots\}$ is all the same.

Does this mean that whenever an epistemic state representing X judges a formula Φ to be unknown then MK₃ judges Φ to be unknown? The answer is 'No, not in the general case'. This is because any set-up s determines the truth-value of a compound formula according to Belnap's truth-tables. Since these tables differ slightly from Kleene's, both systems are different. For example, we have that Belnap's computer says B is unknown when given $A, \neg A \lor B$, since the following set-ups determine the computer epistemic state:

$$s_1(A) = t \ s_2(A) = b \ s_3(A) = b$$

 $s_1(B) = t \ s_2(B) = t \ s_3(B) = u$

If we take $E = \{s_1, s_2, s_3\}$ we have E(A) = t and E(B) = u. But, $B \in Cn_{M\S}(\{A, \neg A \lor B\})$. However, if we feed the computer only with literal input, then both logics are equivalent:

² Please do not confuse an epistemic state in Belnap's sense with a state or possible world of a Kripke-structure. Possible worlds are sometimes also referred to as epistemic states. An epistemic state in Belnap's sense, however, is more similar to Kripke structure, since it is a collection of interpretation functions.

Proposition 6.7. Let X be a set of literals, E the epistemic state representing X. Then

$$E(L) = f \text{ implies } \neg L \in \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X); L \notin \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X)$$
$$E(L) = t \text{ implies } L \in \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X); \neg L \notin \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X)$$
$$E(L) = u \text{ implies } \Diamond \neg L \land \Diamond L \in \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X)$$
$$E(L) = b \text{ implies } \neg L, L \in \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X); L, \neg L \notin \operatorname{Cn}_{\mathsf{M}\mathsf{k}}(X)$$

Proof. It is easy to verify that E(L) = f if $\neg L \in X$ (remember that X is a set of literals). Since $\operatorname{Cn}_{M\S}$ is inclusive we have $L \in \operatorname{Cn}_{M\S}(X)$. The other cases are similar.

Thus, Belnap's logic and MK_3 have different opinions on what the truth-value of a compound statement should be. If there aren't any compound statements, Cn_{MK} can be seen as a conservative extension of Belnap's logic.

6.4 Related Work

I shall now relate the logic MK_3 to two other systems: Wagner's *Vivid Logic* and Lakemeyer's and Levesque's *Knowledge Representation Service*. As we shall see, Vivid Logic³ shares many aspects with our idea of reasoning about paraconsistent and unknown information; both logics base on Belnap's logic L4. It turns out that MK_3 has almost all desirable properties of a vivid logic; if it lacks a property then on purpose.

Whereas the language of vivid reasoning is the propositional, non-modal language \mathcal{L} plus an additional connective – which expresses another form of negation, Lakemeyer's and Levesque's system is at least from a formal point of view very close to ours. They use a modal language and a Kripke-style semantics with three-valued interpretation functions. The underlying ideas and intentions do however strongly differ from ours.

6.4.1 Vivid Logic

Wagner characterises a vivid logic as a formal system which has the following properties:

³ Wagner stresses that Vivid Logic is not a fixed system; for convenience, however, I shall use the terms vivid reasoning and Vivid Logic interchangeably.

- Two Kinds of Negation These two kinds are referred to as *weak* and *strong* negation. A form of weak negation is *negation-as-failure* which means that $\sim A$ is accepted if A cannot be proved. Strong negation means that the falsity of A must be directly established; it is therefore also called *constructible falsity*.
- Non-Explosiveness This is a stronger form of paraconsistency. Whereas paraconsistency simply means that there is a sentence Φ such that $\{A, \land A\} \not\models \Phi$, non-explosiveness means that if Φ is any non-tautology, then for every X there is a variant Φ' of Φ such that $X \not\models \Phi'$. For example, Johansson's Minimal Logic (cf. [Johansson, 1937]) is paraconsistent but not explosive, because we have $\{A, \neg A\} \vdash \neg \Phi$, for all Φ but in general $\{A, \neg A\} \not\vdash \Phi$.
- Constructivity Let X be a set of atomic and negated atomic formulas. A consequence relation \vdash is said to be *constructive* if the following holds:

$X \vdash \varPhi \lor \Psi$ implies $X \vdash \varPhi$ or $X \vdash \Psi$	(Constructible Truth)
$X \vdash \neg(\varPhi \land \Psi)$ implies $X \vdash \neg \varPhi$ or $X \vdash \neg \Psi$	(Constructible Falsity)

Wagner observed that neither Classical Logic nor Johansson's Minimal Logic nor Heyting's Intuitionistic Logic is constructive.

Restricted Reflexivity Reflexivity is restricted to consistent formulas, i.e. $X \cup \{\Phi\} \vdash \Phi \text{ if } \Phi \text{ is consistent w.r.t. } X$. Wagner remarks that this does of course require an appropriate notion of consistency. In our case, a sensible notion of consistency is to say: if Φ takes the value t in every preferred K₃-model of X, or $\Box \Phi$ hold in every preferred MK₃-model of X.

Whether the above principles hold for MK_3 depends on the type (i.e. modal or non-modal) of a formula. For example, since $\Box \Phi \land \neg \Box \Phi$ has no model, we have that MK_3 is in general not paraconsistent.

Remark 6.2. Cn_{MK} is not paraconsistent.

This coincides perfectly with our intuition about the modal operators: the formula $\Box \Phi \in \operatorname{Cn}_{MS}(X)$ means that Φ is consistently provable from X. Whereas $\{A, \neg A\}$, A is nonmodal, means that both A and $\neg A$ have been told we cannot have that for example A is consistently provable from X and not consistently provable from X. This is similar to saying A has been told and has not been told. Thus, if we are talking about knowledge or modal operators it makes sense that the tertium non datur as well as EFQ holds.

On the other hand, the sublogic K_3 is not only paraconsistent but also non-explosive.

Proposition 6.8 (Non-Explosiveness of Cn_3). Let Φ be non-tautological, $X \subset \mathcal{L}$ in INF such that there is some $A \in \Sigma$ with $A \notin X$ or $\neg A \notin X$. There is a variant Φ' of Φ such that $\Phi' \notin Cn_3(X)$.

Proof. By induction on the degree of Φ . The base case follows easily from the prerequisite that there is at least one $A \in \Sigma$ such that $A \notin Cn_3(X)$ or $\neg A \notin Cn_3(X)$. Assume that the proposition holds for all Φ with degree at most n. We distinguish two cases:

 Φ has the form $\neg \varphi$: Assume to the contrary that for all $\varphi \in \mathcal{L}$ we have $\neg \varphi \in Cn_3(X)$. Hence, $\neg A, \neg \neg A \in Cn_3(X)$, for every $A \in \Sigma$ – a contradiction. Φ has the form $\varphi_1 \land \varphi_2$: Immediately by the induction hypotheses.

The requirement that there is at least one A such that either A or $\neg A$ is a member of X excludes the case of degenerated X (for example $X = \mathcal{L}$ or $X = \{A, \neg A \mid A \in \Sigma\}.$

Corollary 6.2. Let X be a set of non-modal sentences. Further assume that all prerequisites if Proposition 6.8 hold. Then there is a variant Φ' of Φ such that $\Phi' \notin \operatorname{Cn}_{\mathsf{MS}}(X)$.

The restriction that X contains only propositional sentences is also made by Wagner: he requires that the database does not contain weak negation (p.20). Since weak negation is in fact a modality, the above restriction does not affect the comparison between MK_3 and Vivid Reasoning.

In the light of Remark 6.2 and Proposition 6.8 it could seem that whereas K_3 satisfies an important principle of vivid reasoning the addition of modal operators made vivid reasoning impossible. This is, however, not true; it depends on whether we look at modal or nonmodal formulas. A principle which holds for modal but not for nonmodal formulas is Constructive Truth:

Remark 6.3. If $\Box \Phi \lor \Box \Psi \in \operatorname{Cn}_{\mathsf{M}\mathfrak{h}}(X)$ then $\Box \Phi \in \operatorname{Cn}_{\mathsf{M}\mathfrak{h}}(X)$ or $\Box \Psi \in \operatorname{Cn}_{\mathsf{M}\mathfrak{h}}(X)$.

Again this coincides with our intuition about the \Box -operator; if we claim that Φ is consistently provable or Ψ is consistently provable, then we must of course be able to prove Φ or to prove Ψ .

The following table extends the table for vividness criteria given in [Wagner, 1994].

0				T.
Classical				
Heyting	\checkmark			
Johansson	\checkmark		\checkmark	
Kleene	\checkmark	\checkmark		
Belnap	\checkmark	\checkmark	\checkmark	\checkmark
Nelson	\checkmark	\checkmark	\checkmark	\checkmark
K ₃		\checkmark	\checkmark	\checkmark
MK_3^4	$\sqrt{5}$	$\sqrt{5}$	$\sqrt{6}$	$\sqrt{6}$

Logic Constr. Truth Constr. Falsity Paracons. Non-Explos.

Here, 'Belnap' denotes Belnap's **L4**; 'Kleene' denotes Kleene's strong-three valued logic *without* a second designated truth-value. Having only one designated truth-value implies that the 'Kleene' logic has no tautologies. This yields that we do not have for every X that $X \vdash A \lor \neg A$ and moreover, Constructible Truth is satisfied.

Contrary to the 'Kleene' logic, $A \vee \neg A$ is a tautology in K_3 and this is a desired property. Thus, it is not very surprising that K_3 does not have the property of Constructible Truth. It is, however easy to see that if X contains only literals and $\Phi \vee \Psi$ is non-tautological then $X \parallel \vdash \Phi \vee \Psi$ implies $X \parallel \vdash \Phi$ or $X \parallel \vdash \Psi$. That is K_3 satisfies a weakened (but not very restricted) form of Constructible Truth.

There are two other properties of vivid reasoning which have not been mentioned in the above table: two-kinds of negation and restricted reflexivity. First I must say that I find restricted reflexivity not very convincing in the absence of any modal operator expressing some notion of consistent knowledge. Consider the following example: Max adds A to the database X. Some hours later Moritz adds $\neg A$ to X which contains now both A and $\neg A$. Now Max, who is a very fearful person, wants to check the next day whether he really entered A and puts the query ?-A. Since the inference relation is not reflexive the answer is No. Max enters again A and checks whether the database has accepted this sentence by asking ?-A. Since X still contains A as well as $\neg A$ the answer is again No.

This episode shows that dropping reflexivity while simultaneously allowing paraconsistency is not very user-friendly when only a propositional language is given. If Max and Moritz' database system supported a \diamond -operator, then Max could ask ?- $\diamond A$ to see what's wrong with A.

The last point to be discussed are the two forms of negation. Wagner argues that a database needs two-kinds of negation: weak negation, which holds if A is not provable and strong negation which holds if $\neg A$ is provable. We can use the

⁴ Note, that if we had restricted the database X to contain only nonmodal sentences then MK_3 would have the same properties as K_3 . Thus, the price we have to pay to allow arbitrary databases $X \subseteq \mathcal{L}_M$ is that we loose some vividness properties.

 $^{^5}$ Only for modalised sentences as in Remark 6.3.

⁶ Provided that X contains only non-modal formulas.

modal operators \Diamond and \Box to express weak negation. For the logic MK this has already been observed in [Lifschitz, 1991]. In the case of MK we can define

$$\sim \Phi =_{\scriptscriptstyle def} \neg \Box \Phi$$

and in the case of MK_3 we can define

$$\sim \Phi =_{def} \neg \Box \Phi \land \neg \Diamond (\Phi \land \neg \Phi).$$

In order to make these operators work correctly, we can, for example, ensure that the database X has at most one preferred model.

Summarising we can say that MK_3 has many important properties of vivid reasoning. Only two properties do reasonably not hold or hold only in a weakened form: Constructive Truth and Restricted Reflexivity. This shows that, even though we made drastic changes to Belnap's original definitions, MK_3 retains almost all philosophical aspects of L4's vividness. This is especially important because Belnap's logic is not merely a system having aspects of vividness but the basis for vivid reasoning (cf. [Wagner, 1994], Chapter 2).

6.4.2 Logical Omniscience: Lakemeyer's and Levesque's Tractable Knowledge Representation Service

In [Lakemeyer and Levesque, 1988] the authors present a logic, which is at least from its formal basis very close to MK_3 . The main purpose for their logic was to express a notion of *only-knowing*. For example, if the system has been told a single fact 'John likes Mary or Sue'⁷ and we ask the system whether it believes that 'John likes Mary' we wish that the answer is NO, because the system *only* knows that John likes Mary or Sue. The problem sounds quite simple and a logic like *M*K or a trick like the closed-world assumption yields a solution. Any of these approaches has the drawback that the agent becomes *logically omniscient*. Logical omniscience means that the reasoner knows e.g. all S5 consequences of his set of initial beliefs. This point of view is regarded by many logicians to be too idealistic. As an example for this extremely idealistic behaviour consider for example the tautology $\Box \Phi \land \Box (\Phi \rightarrow \Psi) \rightarrow \Box \Psi$, which says that the agent's knowledge or his belief are closed under implication.

Lakemeyer and Levesque do not only consider a perfect reasoner to be too idealistic a model for a real reasoner but give another important argument for skipping logical omniscience: rejecting perfect reasoning can make the reasoner

⁷ The example is taken from [Lakemeyer and Levesque, 1988].

become tractable in the propositional case. This means that computing whether the agent believes Φ requires time polynomial in the size of the database.

As a consequence of giving up logical omniscience many classical modal and propositional tautologies are no longer valid in Lakemeyer's and Levesque's system:

LO 1 $A \lor \neg A$ is no tautology. This is basically a consequence of choosing a 4-valued valuation⁸ similar to Belnap's logic.

LO 2 $\neg \Box (A \land \neg A)$ is satisfiable.

LO 3 $\Box A \rightarrow \Box (A \lor B)$, i.e. disjunctive weakening does not hold.

Note that 2) is also invalid for MK_3 . In addition, there is another criterion of logical omniscience which is invalid for MK_3 :

LO 4 if Φ is a tautology then $\Box \Phi$ is valid. We have seen that there are structures in which, for example, $\Box(A \to A)$ is invalid.

In [Meyer and Van der Hoek, 1995] there are further criteria given:

- **LO 5** If $\Phi \to \Psi$ is a tautology then $\Box \Phi \to \Box \Psi$ is also a tautology (Closure under valid implication).
- **LO 6** If $\Phi \leftrightarrow \Psi$ then $\Box \Phi \leftrightarrow \Box \Psi$ is also a tautology (Belief of equivalent formulas).
- **LO** 7 $\Box \Phi \land \Box \Psi \rightarrow \Box (\Phi \land \Psi)$ is valid (Closure under conjunction)
- **LO 8** $\Box \Phi \rightarrow \neg \Box \neg \Phi$ is valid (Consistency of beliefs)

Van der Hoek and Meyer state that LO 1-9 are undesirable when 'modelling a (human or artificial) agent's belief'. And this is exactly the point where for example Lakemeyer's and Levesque's approach differs from ours. Instead of modelling a human or artificial agent we defined a formal system which is not fallible the way humans are. Thus, we wish to stay closer to Belnap's 'How a Computer Should Think' than to model agents.

LO 9 $\Box(\Box \Phi \to \Phi)$ is valid (Belief of having no false beliefs)

⁸ Instead of using truth-values, the authors use sets T and F which are similar to the set P_P^+ and P_P^- used in Chapter 3. Contrary to our requirement that $P_P^+ \cup P_P^- = D$ they do not require that $T \cup F = \Sigma$, i.e. there could be atomic formulas A which are neither contained in T nor in F. This is exactly the meaning of Belnap's truth-value u.

6.5 Conclusion

We have seen that MK_3 is – given the premises of Chapter 1 – an alternative to Belnap's L4. It has tautologies as desired, is able to express *if* ... *then* sentences and its inferential behaviour is still very close to Belnap's L4. Of course, MK_3 is in general no longer paraconsistent. If, however, we restricted the database to contain only nonmodal sentences (or any other suitable language restriction which forces the database to have an MK_3 -model) then we could ensure that there are no contradicting modal formulas. The contradicting propositional formulas can be handled as in the logic K_3 .

CHAPTER 7

Closure

Don't worry, I don't intend to tell you all the summaries which appeared at the end of each chapter again. Please, do consult these chapters if you are looking for a technical summary. At this point, I would like to discuss the pure practical aspects, or the usefulness as Belnap puts it, of the logic developed.

What has been achieved? We have defined a logic, MK_3 for reasoning about unknown and inconsistent information. We have shown that MK_3 is very close to Belnap's original ideas while also being able to represent implicational knowledge. An important point – at least one stressed by Belnap quite often – is the usefulness of his logic. What about the usefulness of MK_3 ? We have shown that the sublogic K_3 is extremely useful, because it does not only have many desired mathematical properties but its computational complexity is not worse than that of classical logic. Of course, the situation changes the very moment we are considering modal aspects. It is well-known that the PSPACE complete problem of Quantified Boolean Formulas (QBF) can be reduced to the question whether a modal formula Φ is an S5-tautology. We have seen that S5-tautologies coincide with the set $Cn_{M\xi}(\emptyset)$. Hence, entailment in MK_3 is PSPACE hard. This is the price we have to pay for having a logic which is able to express a notion of consistency.

Anyway, I think that MK_3 is useful (besides, I don't believe that 'useful' means 'computable in polynomial time'). I have several arguments defending this point. To me it seems that an important point of any Question/Answering systems is robustness, in the sense that the system never enters a state in which it becomes useless to the user. For any logical inference system this means that paraconsistency is a must. The second point is transparency. A system should not only have a clear well-founded (theoretical) basis, it should also be as close as possible to what the user is used to. In our case, we suspected the user to be used to classical logic. The discussion on the valid patterns of inference (i.e. the sequent-style system), the system's reasoning is that of classical logic except for that case where one of the formulas involved might be paraconsistent. The logic MK_3 is thus useful because it answers according to the rules of classical logic whenever possible.

Of course, MK_3 is not paraconsistent. That is, any input of the form $\Diamond A$, $\neg \Diamond A$ could bring the system to collapse. Whenever we allow the user to enter formulas involving modal operators, robustness cannot be guaranteed. As a practical solution I would suggest to allow the user to put queries involving modal operators but not to feed the database with such formulas. Only a small group of dedicated users should be allowed to enter formulas like e.g. $\Diamond A \to B$. Sounds strange? I don't think so. I mean, in any technical system we can *reduce* a risk of any kind but we cannot *exclude* it (hey, not even UNIX is fool-proved because the system operator can post your login names together with your password to a newsgroup).

The above approach of *access control* is a standard technique used by operating system or database management systems (DBMS). Ullman states: 'Access Control: The ability to limit access to data by unauthorized users, and the ability to check the validity of data' ([Ullman, 1988]). Other authors use the term *access level* to handle privileges in a database system (cf. [Elmasri and Navathe, 1989]).

I don't want to take the comparison between logic and deductive databases too far. This is mainly because no matter how useful a logic might be, at the moment it seems to be a too complex thing to be considered for a real world application. Anyway, logic is a funny thing to play around with and who knows whether any of the logics developed in Computing Science will ever find their way to a product, like ORACLE. All we can do is wait and see. For a quite a long time nobody could find any practical application of number theory. This branch of mathematics was also considered by mathematicians to be quite exotic. Then it turned out that number theory is a theoretical basis for cryptology. Attendons la fin de l'histoire; maybe this will happen to logic as well.

Index

k-Preferred SAT, 40 LK, 106 $LK_3, 108$ $T_{\mathfrak{A}}, 134$ α -rule, 97 ATOM, 7 Ax, 79, 83 $\parallel_{\rm Bolz},\,14$ β -rule, 97 $\overline{\mathbf{k}_{3}}$, 61 , 139 CMP, 91 $Cn_{Bolz}, 31, 34, 37$ $Cn_{\rm K}, \, 64$ $Cn_{MK}, 118$ $Cn_3, 17$ $Cn_3^+, 49$ Cred, 50 $\Delta_{k+1}^P, 52$ $\underbrace{\stackrel{\kappa_{\mp}}{\models}}_{M_{K}}^{n}, 118 \\ \underset{\overline{M}_{K}}{\models}, 141 \\ K_{3}\text{-model}, 13$ $\parallel_{\rm E}$, 131 CL, 106 \Box , 15 \sqsubseteq , 15 $MK_3, 138$ MOD(X), 13 $\models_{\overline{K}_3}, 61$ $\begin{matrix} \overset{}{\models}_{\mathsf{MK}}, 116 \\ & \overset{}{\models}_{\mathsf{MK}}, 139 \\ \mathsf{NC}(X), 89 \end{matrix}$ $\Pi_{k+1}^{P}, 52$ PMOD, 17 Cred, 48 Scep, 48 $\Sigma_{k+1}^{P}, 52$ Scep, 48, 51 \prec , 117 STRUCT, 115

 $T_X,\ 134$ $\mathsf{Taut}_{\mathsf{NC}}, 23$ $Th_{K_{4}5}, 135$ $Th_{S5}(X), 121$ ∖≡, 13 $\Vdash \, , \, 17$ $\parallel_{\overline{\kappa}_3}, 64$ II_{MK}, 118 $\underset{\mathcal{L}, 7}{\overset{\text{max}}{\Vdash}}, 141$ \mathcal{R} -provable, 85 4-axiom scheme, 121 5-axiom scheme, 121 K-axiom scheme, 121 L4, 3, 148 T-axiom scheme, 121 CA, 29 AND, 29 AND property, 5 D, 30 **OR** property, 5 S, 29 adaptive logics, 56AE-extension, see extension134 AEL, see autoepistemic logic131 autoepistemic logic, 131, 133 axiom system - for K₃, 79, 83 - for FOK_3 , 94 base revision, 52belief revision, 47, 51 Bolzano, 14 cautious modus ponens, 91 cautiousness, 10, 14 clausal normal form, 22 clause, 41 CNF, see conjunctive normal form21 $\operatorname{compactness}$

 $- K_3, 34$ - paraconsistent, 36 - failure of for Cn_S , 129 - nonmodal, 121 - of classical logic, 34 - of consistency -- for MK, 121 - of entailment, 34 - of satisfiability, 34 computational complexity, 39 - of Cn_3 , 42 - of belief revision, 53 - of finding a pref. K_3 -model, 40 - of full meet revision, 52 congruence, 31 conservativity, 10, 20 constructivity, 152 counterfactuals, 51 credulous consequence, 48 cumulativity, 26, 27, 33 cut, 26, 108 cut rule, 107 cut-elimination theorem, 107 cutty, 26, 27, 144 deduction theorem - for K_3 , 22 - for MK, 120 - for first-order logic, 70 - normalised, 23, 71 - sential, 71 - ultra normalised, 108

weak, 22
default logic, 47, 92, 131
degree (of a formula), 7
designated truth-value, 11
disjunctive syllogism, 55

efq, 2 EFQ, see ex-falso principle14 entailment - preferential, 14 - three-valued -- Bolzano, 14 -- monotonic, 31 -- preferential, 17 entailment, MK₃, 141 epistemic state, 150 ex-falso principle, 14 extension- autoepistemic, 134 fixpoint - semi-inductive, 92 formula - atomic, 7 full meet revision, 51 ground term, 61, 65 grounded sets, 134 Hintikka set, 104 honest theory, 137 implication, 4 - strict, 54 INF, see implicational normal form21 inference-frames, 31 intuitionistic logic, 152 K45, 135 Kleene's strong three-valued logic, 24 Kleene's truth-tables, 12 Kripke-model, 116 L4, 138 left logical equivalence, 26, 27, 144 literal, 7, 21 LP(m), 17, 112 M-literal, 124MCNF, see modal conjunctive normalform 124minimal logic, 152

modal conjunctive normal form, 124, 145

- ordered, 128

modal horn, 124 model - K_3 , 13 - preferred, 14, 24 -- MK, 118 -- existence of, 24 -- existence of MK, 119 -- existence of MK_3 , 143 -- strongly, 129 -- three-valued, 16 -- three-valued Kripke , 141 -- uniqueness, 122 - three-valued, 14 modus ponens, 23, 81 monotonic behaviour of K_3 , 33

NC, see normalisation condition23 NEC, see necessitation144 necessitation, 144 NML2-S5-Extension, 136 non-explosiveness, 152 non-redundant, 44 Nonmonotonic Logic II, 136 normalisation condition, 23, 81

omniscience, 155

paraconsistency, 9 paraconsistent deduction property, 23 paraconsistent extension, 92 paraconsistent logic, 2 polynomial time hierarchie, 52 preclosure, 10, 19 preferred subset, 42 pseudo relevance logics, 55

r.e. approximation, 77
Ramsey test, 51
reflexivity, 5, 26, 152
relevance logic, 47, 54
– undecidability of , 54
replacement theorem, 19
Rescher, 47, 48

restricted modus ponens, 83 right weakening, 26, 28 - normalised , 28 S5, 121, 136 SAT, 40, 52 - critical, 41 sceptical consequence, 48 semantical equivalence $-MK_3, 145$ sequent, 106 - validity, 108 sequent calculus, 106sequent calculus for K_3 , 108 set-up, 3 sets as signs, 101signature - first-order, 60 - propositional, 12 signed formula, 97 simple base revision, 52 simple base revision complexity, 52 stable set, 131, 147 strong consequence, 48 strongly relatively consistent, 84 $\operatorname{structure}$ Σ -algebraic, 60 - Kripke , 115- three-valued Kripke, 139 substructure, 63 supradeductivity, 36, 87, 90, 112 system C, 26, 28 system P, 26, 29 tableau, 97 - for K₃, 103 - closed, 99 - open, 99 - proof, 100 -- BF, 105 - rules -- for propositional logic, 98

- semantical, 97

tautologies - MK₃, 144 three-valued interpretation function, 13 turing reducibility, 52

unknown, 138 - operator, 115

validity - MK₃, 139

- of a K₃sequent, 108 of a nonmonotonic inference rule, 82
- of signed formulas, 100

- S5, 116

variable sharing property, 47, 55

vivid logic, 151

weak consequence, 48

Bibliography

- [Anderson and Belnap, 1975] Anderson, A. E. and Belnap, N. D. (1975). Entailment: The Logic Of Relevance And Necessity. Princeton University Press.
- [Avron, 1989] Avron, A. (1989). Relevance and paraconsistency a new approach. part iii: Cut free gentzen-type systems. Notre Dame Journal Of Formal Logic.
- [Avron, 1990a] Avron, A. (1990a). Relevance and paraconsistency a new approach. *Journal Of Symbolic Logic*.
- [Avron, 1990b] Avron, A. (1990b). Relevance and paraconsistency a new approach part ii: The formal systems. *Notre Dame Journal Of Formal Logic*.
- [Avron, 1991] Avron, A. (1991). Simple consequence relations. Information and Computation, 92:105–193.
- [Avron, 1993] Avron, A. (1993). Gentzen-type systems, resolution and tableaux. Journal of Automated Reasoning, 10:265-281.
- [Balcazar et al., 1988] Balcazar, J., Diaz, J., and Gabarro, J. (1988). Structural Complexity I. Springer, Berlin.
- [Batens, 1997] Batens, D. (1997). A survey of inconsistency adaptive logics. In Proc. of the World Congress on Paraconsistency. University of Ghent.
- [Bell and Weber, 1993a] Bell, S. and Weber, S. (1993a). On the close logical relationship between foil and the frameworks of Helft and Plotkin. In Muggleton, S., editor, *Third International Workshop on Inductive Logic Programming.* Extended Version Available as Technical Report 510/93, University of Dortmund.
- [Bell and Weber, 1993b] Bell, S. and Weber, S. (1993b). A three-valued logic for inductive logic programming. Technical Report 510/93, University of Dortmund.
- [Belnap, 1977] Belnap, N. D. (1977). A useful four-valued logic. In Dunn, J., editor, *Modern Uses of Multiple-Valued Logic*. Reidel, Dordrecht.
- [Benferhat et al., 1995] Benferhat, S., Dubois, D., and Prade, H. (1995). Some systactic approaches to the handling of inconsistent knowledge bases: A comparative study. Technical report, University of Toulouse, IRIT/94-55-R.
- [Bergmann and Noll, 1977] Bergmann, E. and Noll, H. (1977). Mathematische Logik. Springer.
- [Bloesch, 1993] Bloesch, A. (1993). A tableau style proof system for two paraconsistent logics. Notre Dame Journal Of Formal Logic.

- [Borgida, 1985] Borgida, A. (1985). Language feature for flexible handling of exceptions in information systems. ACM Trans. Database Systems, 10:536– 603.
- [Brewka, 1991] Brewka, G. (1991). Nonmonotonic Reasoning: Logical Foundations Of Commonsense. Cambridge University Press.
- [Chang and Keisler, 1977] Chang, C. C. and Keisler, H. J. (1977). Model Theory. North-Holland Publishing Co.
- [Chellas, 1980] Chellas, B. F. (1980). *Modal Logic: An Introduction*. Cambridge University Press.
- [de Swart, 1993] de Swart, H. (1993). Mathematics, Language, Computer Science and Philosophy. Peter Lang.
- [del Cerro et al., 1996] del Cerro, Delgado, and Herzig (1996). Talking 'bout consistency. In *FAPR 1996*. Springer, LNAI.
- [Dietrich, 1995] Dietrich, J. (1995). *Inferenzframes*. PhD thesis, Universität Leipzig.
- [Dunn, 1986] Dunn, J. (1986). Relevance logic and entailment. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic*. Reidel, Dordrecht.
- [Eiter and Gottlob, 1992] Eiter, T. and Gottlob, G. (1992). On the complexity of propositional knowledge base revision, updates and counterfactuals. *Artificial Intelligence*, 57.
- [Elmasri and Navathe, 1989] Elmasri and Navathe (1989). Fundamentals of Database Systems. Benjamin/Cummings Publishing Company, Menlo Park.
- [Emde, 1991] Emde, W. (1991). Modellbildung, Wissensrevision und Wissensrepräsentation im Maschinellen Lernen. Springer.
- [Fitting, 1990] Fitting, M. (1990). First-Order Logic and Automated Theorem Proving. Springer, New York.
- [Forbus, 1989] Forbus, K. (1989). Introducing actions into qualitative simulation. In *IJCAI*.
- [Gabbay, 1985] Gabbay, D. (1985). Theoretical foundations for nonmonotonic reasoning in expert systems. In *Proc. of the 1st Conference on Theoretical Aspects of Knowledge*. Morgan Kaufmann.
- [Garey and Johnson, 1979] Garey, M. and Johnson, D. (1979). Computers and Intractability. W.H. Freeman, New York.
- [Genesereth and Nilsson, 1987] Genesereth, M. and Nilsson, N. (1987). Logical Foundations of Artificial Intelligence. Morgan Kaufmann.
- [Gentzen, 1934] Gentzen, G. (1934). Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39:176–210 (Teil 1) and 405–431 (Teil 2).

- [Ginsberg, 1986] Ginsberg, M. (1986). Counterfactuals. Artificial Intelligence, 30:35-79.
- [Halpern and Moses, 1984] Halpern, J. and Moses, Y. (1984). Towards a theory of knowledge and ignorance. In Proc. of the AAAI Workshop on Nonmonotonic Reasoning.
- [Herre, 1990] Herre, H. (1990). Semantics of nonmonotonic reasoning in logic programming. In Nonclassical Logics and Information Processing. Springer.
- [Herre, 1995] Herre, H. (1995). Generalized compactness of nonmonotonic inference operations. Journal of Applied Nonclassical Logics, 5:121 – 135.
- [Hilbert and Bernays, 1934] Hilbert, D. and Bernays, P. (1934). Grundlagen der Mathematik I. Springer.
- [Hirsh, 1990] Hirsh, H. (1990). Overgenerality in EBG. In Brazdil, P. and Konolige, K., editors, *Machine Learning, Metareasoning and Logics*. Kluwer.
- [Hughes and Cresswell, 1968] Hughes, G. E. and Cresswell, M. J. (1968). An Introduction To Modal Logic. Methuen & Co.
- [Johansson, 1937] Johansson, I. (1937). Der Minimalkalkül Ein reduzierter intuitionistischer Kalkül. Composito Mathematica, 4:119–136.
- [Kifer and Lozinskii, 1992] Kifer, M. and Lozinskii, E. L. (1992). A logic for reasoning with inconsistency. *Journal Of Automated Reasoning*, 9:179–215.
- [Kleene, 1952] Kleene, S. (1952). Introduction to Metamathematics. Van Nostrad, Amsterdam.
- [Konolige, 1988] Konolige, K. (1988). On the relation between default logic and autoepistemic logic. *Artificial Intelligence*, 35.
- [Kraus et al., 1990] Kraus, S., Lehmann, D., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence, 44:167–207.
- [Lakemeyer and Levesque, 1988] Lakemeyer, G. and Levesque, H. (1988). A tractable knowledge representation service with full introspection. In Vardi, M., editor, *Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann.
- [Lifschitz, 1991] Lifschitz, V. (1991). Nonmonotonic databases and epistemic queries. In Proc. of the IJCAI'91. Morgan Kaufmann.
- [Lukaszewicz, 1990] Lukaszewicz, W. (1990). Non-Monotonic Reasoning: Formalization Of Commonsense Reasoning. Ellis Horwood Ltd.
- [Marek and Truszczyński, 1993] Marek, V. W. and Truszczyński, M. (1993). Nonmonotonic Logic: Context Dependent Reasoning. Springer-Verlag.
- [Meyer and Van der Hoek, 1995] Meyer, J. and Van der Hoek, W. (1995). Epistemic Logics for AI and Computer Science. Cambridge University Press.
- [Morik and Wrobel, 1933] Morik, K. and Wrobel, S. (1933). *Knowledge Acquisition and Machine Learning*. Academis Press.

- [Nebel, 1992] Nebel, B. (1992). Syntax based approaches to belief revision. In Gärdenfors, P., editor, *Belief Revision*. Cambridge University Press.
- [Norman and Sylvan, 1989] Norman, J. and Sylvan, R. (1989). Directions in Relevant Logic. Kluwer, Dordrecht.
- [Papadimitriou, 1994] Papadimitriou, C. (1994). Computational Complexity. Addison Wesley,.
- [Papadimitriou and Yannakakis, 1992] Papadimitriou, C. and Yannakakis, M. (1992). On finding extensions of default theory. In Proc. International Conference in Database Theory. Springer, LNCS.
- [Papalaskari and Weinstein, 1990] Papalaskari, M. and Weinstein, S. (1990). Minimal consequence in sential logic. Journal of Logic Programming, 9:19–32.
- [Priest, 1991] Priest, G. (1991). Minimally inconsistent LP. Studia Logica, 50 (2):321.
- [Rescher, 1964] Rescher, N. (1964). Hypothetical Reasoning. North-Holland Publishing Co.
- [Rescher, 1969] Rescher, N. (1969). Many Value Logic. McGraw-Hill Book Co.
- [Rosser and Turquette, 1952] Rosser, J. and Turquette, A. (1952). Many-Valued Logics. North-Holland, Amsterdam.
- [Satho, 1988] Satho, K. (1988). Nonmonotonic reasoning by minimal belief revision. In Proc. Intl. Conf. on Fifth Generation Systems, pages 455–462.
- [Shoham, 1988] Shoham, Y. (1988). Reasoning About Change: Time And Causation From The Standpoint Of Artificial Intelligence. MIT Press.
- [Stalnaker, 1993] Stalnaker, R. (1993). A note on nonmonotonic modal logic. Artificial Intelligence, 64.
- [Stockmeyer, 1976] Stockmeyer, L. (1976). The polynomial time hierarchie. *Theoretical Comp. Science*, 3:1–22.
- [Thiele, 1996] Thiele, H. (1996). Logische Systeme der Informatik. Technical report, University of Dortmund.
- [Ullman, 1988] Ullman, J. (1988). Principles of Database and Knowledge Base Systems. Computer Science Press, Rockville.
- [Urquhart, 1984] Urquhart, A. (1984). The undecidability of entailment and relevenat implication. *Journal of Symbolic Logic*, 49:1059 1073.
- [van Dalen, 1980] van Dalen, D. (1980). Logic and Structure. Springer.
- [Wagner, 1994] Wagner, G. (1994). Vivid Logic: Knowledge Based Reasonong With Two Kinds of Negation. Springer.
- [Weber, 1996a] Weber, S. (1996a). On the semantics of the unknown. In *Prac*tical Reasoning – Proc. on Formal and Applied Practical Reasoning. Springer, Berlin, New York.
- [Weber, 1996b] Weber, S. (1996b). Proof systems for paraconsistent preferential entailment. In Proc. of NMELP '96, Dix, Perreira and Prymuszinski (Eds.).

- [Weber, 1997a] Weber, S. (1997a). Mathematical properties of Priest's version of Kleene's SK₃. In Batens, D. and Priest, G., editors, *Proceedings of the Congress on Paraconsistency*.
- [Weber, 1997b] Weber, S. (1997b). Proof systems for minimal LP. In Batens, D. and Priest, G., editors, Proceedings of the Congress on Paraconsistency.
- [Weber and Bell, 1994] Weber, S. and Bell, S. (1994). A note on paraconsistent entailment. Technical report, University of Dortmund.
- [Winslett, 1990] Winslett, M. (1990). Updating Logical Databases. Cambridge University Press.