Minimal Consequence: A Semantic Approach to Reasoning with Incomplete Information

Mary-Angela Papalaskari

Ph.D.

University of Edinburgh

1988

Declaration

The work described in this thesis arose through collaboration with Scott Weinstein of the University of Pennsylvania. I certify that my contribution to this joint work is highly significant and that this thesis was composed by me.

Acknowledgements

I would like to thank my supervisor, Alan Bundy for his uninterrupted help, encouragement, and support in writing this thesis. The work of this thesis grew out of my initial efforts in finding ways to guide the application of circumscription. Alan significantly contributed to that work and provided me with direction in pursuing subsequent research. I would also like to thank Alan Smaill, who came to be my second supervisor in the fall of 1986, but who nevertheless quickly came to play an active role in advising me.

This thesis describes joint work with Scott Weinstein of the University of Pennsylvania. His contribution to this work is significant and is gratefully acknowledged. Scott's influence radically changed my way of thinking about problems in Artificial Intelligence and my collaboration with him has greatly increased my knowledge of logic and recursion theory.

The University of Edinburgh provided financial support through a Postgraduate Studentship. I would also like to thank the University of Pennsylvania for a Reasearch Assistantship at the Department of Computer and Information Science (supported in part by DARPA grants NOOO14-85-K-0018, NSF-CER grant MCS-8219196 and U.S. Army grants DAA29-84-K-0061, DAA29-84-9-0027) arranged by Aravind Joshi and Bonnie Webber.

I would like to thank Len Schubert, who first introduced me to AI and the area of non-monotonic reasoning. Many thanks to all those who have subsequently discussed or commented on previous drafts of this thesis: Naoki Abe, David Etherington, John Jones, Aravind Joshi, Chris Mellish, Bhagirath Narahari, Allan Ramsey, Ethel Schuster, Sunil Shende, and Bonnie Webber.

Finally, I would like to thank all my friends.

Abstract

Minimal consequence is embodied in many approaches to non-monotonic reasoning. In this thesis we define minimal consequence in sentential logic and present a number of results of a model theoretic and recursion theoretic character about this newly introduced non-monotonic consequence relation. We show that the minimal consequence relation is not compact and is Π_2^0 and not Σ_2^0. We also connect this relation to questions about the completion of theories by "negation as failure." We give a complete characterization of the class of theories in sentential logic which can be consistently completed by "negation as failure" using the newly introduced notion of a subconditional theory. We show that the class of theories consistently completable by negation as failure is Π_2^0 and not Σ_2^0.

In first order logic minimal consequence is the semantic notion underlying circumscription formalisms. We study domain, predicate, and formula minimal consequence, which are obtained by varying the type of minimization involved and correspond to domain, predicate, and formula circumscription, respectively. The results, again, are model theoretic and serve to clarify properties of minimal consequence in first order logic. Relationships between domain, predicate, and formula minimal consequence are established. We show that every satisfiable theory in a finite language can be finitely expanded to a minimally satisfiable theory in an extended language which has the same logical consequences in the original language as the original theory. We also show that minimal satisfiability is not compact for any of the types of minimality under consideration.

Table of Contents

Chapter 1

Introduction

When looking at the world through a small "information window" our imagination and ingenuity are responsible for making up the remainder. Most of the time we are forced to try our competence in this respect, and, fortunately, most of the time the world is very lenient: it is usually good enough to consider a small number of possibilities. As we experience the world, through ever widening information windows, we develop and improve these capabilities; every time that the window widens, we find out a little more about the world; more importantly, we also find out something about our choices and the methods that led us to these choices.

It is very difficult to study a process like this, because certainly by the time we are old enough to discuss it, it is already routine; from then on we are happy to simply use it in dealing with other pressing concerns, like learning to speak. By the time that we are ready to be interviewed and psychologically tested we are too competent and too familiar with it. The best we can do is to isolate a list of "features" that appear to work.

Suppose now that one of us crosses to the other side of the window and w play a game. You are on the outside and I am on the inside and we now have a telescope. Now the whole world is in view for you, but I can only see a piece of it through the telescope, which you control. You may zoom or point the telescope in whichever direction you choose. Your objective is to give me as much information as possible about the world. You are also free to talk to me and explain which part of the world is visible, so as to help me make a better judgement about the rest, but it is impossible for you to communicate all the information about the world, either through the telescope or through words or with the two together. There is a crucial factor which you would like to take into account, and that is my method of going about making conjectures. Unfortunately, all you can do is assume that I am like you. Moreover, although you are very competent at applying your method, you know little about its working. Perhaps you discuss this with me. You may also look through the telescope and imagine what you would think the world is like from what you see, and decide whether that is indeed close to what it is really like, from what you know. This knowledge is somewhat of a burden, however: once you know what the world is like, it is very difficult to imagine anything else. So, you try something else: you get me to look through the telescope and tell you what I think the world is like. This seems to work a little better. Soon you find out which angles are good and which are misleading. Now, suppose that you were to generalize this, *i.e.*, suppose that new worlds went by and for each one you had to point the telescope for me. Could you develop a good methodology for doing this?

Note that you may develop a good methodology for this game in two ways: 1) you learn my approach to making conjectures; or 2) we come to an understanding about what type of information you will be showing me and what type of conjectures I should be making.

This is the game of logic programming. The methodology, we can say, is fair and it has been developed in a hybrid way: you (the designer of a logic programming paradigm) take into account some features of my (the programmer's) approach to making conjectures, but we have also come to an understanding about the sorts of conjectures that I should make (we have fixed a logic programming language and I have learned its interpretation).

1 Objectives

The hypothesis of this thesis is that the features of our approach to making conjectures, both in everyday life and in a contrived situation such as logic programming, should coincide. Although a good methodology for winning the logic programming game could be developed in either of the two ways mentioned above, we maintain that the first way is superior because it sheds some light into human reasoning. Furthermore, it reflects *how* we win this game — with style and ease; by providing a robust, elegant, and user friendly logic programming environment.

This thesis explores the use of one feature common among approaches to logic programming and to human reasoning with incomplete information — that of "smallness." We study minimal consequence relations, expressing that "B is true in all minimal models of A." Minimal consequence is a semantic notion implicit in past attempts to deal with the problem of incomplete information, both in the context of logic programming and in the context of modelling human reasoning. From the point of view of logic programming, these involve the notion of "negation as failure," *i.e.*, roughly, the assumption that information not derivable can be assumed to be false. From the point of view of modelling human reasoning, past attempts in this line involve the notion of "circumscription," *i.e.*, formalisms that generate conjectures of the sort "the objects that can be shown to exist (or, can be shown to have a certain property P) by reasoning from a set of facts A, are all the objects there are (or, that have the property P)."

Our objectives are twofold:

1. to formulate minimal consequence for sentential logic[1] and provide a complete and detailed study of its model theoretic and recursion theoretic properties that bear on practical applications.

---

[1] Also called propositional logic.

2. to investigate properties of minimal consequence for first order languages where these have a direct impact on the applicability and power of circumscription formalisms.

Previous work on reasoning with incomplete information has focussed primarily on first order logic and has largely been occupied with syntactic as opposed to semantic questions. This is true, for example, of McCarthy's original work on circumscription [McCarthy 80], as well as Clark's work on negation as failure [Clark 78]. Subsequent developments in each of these traditions have accorded some attention to semantic issues. A significant example of this trend is Schlipf's extended study of the recursion theoretic complexity of various semantically defined first order consequence relations developed as partial explications of circumscription formalisms. None of this work has, however, addressed the question of reasoning with incomplete information in the context of sentential logic. This question is of great significance to the development of automated systems for reasoning with incomplete information. In semantics of logic programming, one is often led to consider the set of ground instances of a (finite) database; in general, this is an infinite sentential theory. Hence, the formulation and study of minimal consequence in sentential logic underlies the applicability of negation as failure rules in logic programming. Thus, our first objective is to formulate minimal consequence in sentential logic and to settle questions that naturally arise regarding this consequence relation.

The earliest semantic study of a minimal consequence relation related to circumscription appears in [Davis 80]. Davis gave the natural definition of a minimal model which underlies McCarthy's domain circumscription and showed that there are satisfiable first order theories that have no minimal models. Subsequent research defined analogous notions of a minimal model that capture other forms of circumscription and showed that the same result holds for these also [Etherington 86]. None of this work, however, considered language related issues, *i.e.*, the effect of the particular symbols available in the non-logical vocabulary of a first order language to the minimal models of a theory in that language. As we will see, language related issues play a crucial role in the existence of minimal models. Previous syntactic research on circumscription attempted also to relate the various circumscription formalisms, but produced no tangible results (see *e.g.*, [Etherington & Mercer 86]). Our second objective is thus to study the well known semantic formulations of circumscription, settling questions regarding the existence of minimal models and drawing precise relationships among the different notions of minimal model and minimal consequence.

2 Outline

Chapter 2 gives a broad introduction to the consequence relations studied in logic and traces the development of minimal consequence as a method of reasoning with incomplete information. This chapter reviews the background material in logic and recursion theory that will be used in the remainder of the thesis and surveys other work on reasoning with incomplete information that is important in motivating this research.

Chapter 3 introduces minimal consequence in sentential logic and gives examples which illustrate some of its more interesting features. We show that every satisfiable sentential theory has a minimal model, but that the consequence relation is not compact. We also connect this relation to questions about the completion of theories by negation as failure. We introduce the notion of a subconditional theory (which subsumes propositional Horn theories) and give a complete characterization of the class of theories in sentential logic which can be consistently completed by negation as failure. Lastly, this chapter examines questions of complexity. We show that both the minimal consequence relation and the class of theories consistently completable by negation as failure (*i.e.*, theories equivalent to subconditional theories) are Π^0_2 and not Σ^0_2 in sentential logic, *i.e.*, that deciding these is as difficult as deciding whether a given program (in a Turing equivalent programming language) halts on every input.

Chapter 4 studies minimal consequence relations in first order logic. In first order logic minimal consequence is the semantic notion underlying circumscription formalisms. Domain, predicate and formula minimal consequence are obtained by varying the type of minimization involved and correspond to domain, predicate, and formula circumscription, respectively. The results, again, are model theoretic and serve to clarify properties of minimal consequence in first order logic which have an impact on the usefulness and applicability of circumscription formalisms. Each type of minimal consequence relation is illustrated by several examples that clarify important issues and some concerns expressed in the literature. Perhaps the most prominent of these concerns is that, unlike sentential theories, there are satisfiable theories in first order logic which have no minimal models. Unusual phenomena which arise when considering the minimal consequences of a theory in an extended language are discussed and exploited in the solution that we offer for this problem: we show that every satisfiable theory in a finite first order language can be finitely extended to a theory (in a finitely extended language) which has minimal models and makes true the same sentences in the original language. We show this fact for all three notions of minimality that we study; in particular, for domain minimal consequence, the extension to the language suffices. Next we establish precise relationships between domain, predicate, and formula minimal consequence. We also show that minimal satisfiability is not compact for any of the types of minimality under consideration in this chapter.

Chapter 5 discusses related work in AI and Logic Programming and compares it to the results obtained in this thesis.

Chapter 6 contains ideas for further research on minimal consequence and reasoning with incomplete information. Here we discuss how the methods developed in this thesis can be employed in investigating alternative directions. We also connect reasoning with incomplete information to learning and propose that learnability considerations should play an important role in future research on this subject.

Chapter 7 gives a summary of the main results of this thesis and conclusions.

Chapter 2

Consequence Relations and Incomplete Information

1 Introduction

The power of human reasoning can, to a great extent, be attributed to people's ability to reason from incomplete information. People seem able to "jump to conclusions" and make plausible conjectures which fill out the information they are given. Such leaps are not only very common, but also appear in all forms of reasoning, from linguistic inferences to problem solving. This raises some very interesting questions about whether such behaviour can be reproduced and be advantageously used in a computer program. We need both a good understanding of this behaviour and a rigorous framework for its study, in order to exploit it in a computational setting. This chapter gives the background to this study, by means of an introduction to a logical treatment of consequence relations and reasoning with incomplete information.

Consequence relations are relations holding between theories (sets of sentences) and sentences. Of interest to logic are consequence relations expressing the fact that the truth of a sentence follows from the truth of a theory. This chapter gives a broad introduction to the consequence relations studied in logic and traces the development of minimal consequence as a method of reasoning with incomplete information. In Section 3 we begin with a discussion of the character and scope of consequence relations and their connection to proof and provability; next we give a brief introduction to the two most prominent logical languages, the language of propositions (or sentence letters) and the first order language of relation and function symbols with quantifiers and equality; lastly, we outline the main properties of semantic consequence relations in classical logic that relate to computational considerations. Section 3 reviews the background material in logic and recursion theory, which will be used in subsequent chapters, so as to introduce the notation used and the spirit of the research. Section 4 below discusses the problem of reasoning with incomplete information, from a logical perspective, as it arises in modelling common sense reasoning, scientific discovery, and logic programming. It introduces a first, very general definition of minimal consequence and discusses past attempts to capture various special cases of it in computational systems.

2 General Notational conventions

We favour English sentences for definitions, theorems, proofs, *etc*, wherever this is possible, but in some cases, to ease readability and avoid ambiguous or lengthy statements, we will use a limited number of (metalinguistic) logical symbols. The standard abbreviation "iff" for "if and only if" will sometimes be denoted "$\Leftrightarrow$," with "$\Rightarrow$" and "$\Leftarrow$" used to denote the "if ... then..." and "only if" parts respectively. In addition, we will (very occasionally) use "$\exists$," "$\forall$," and "&" to stand for "there exists," "for every," and "and" respectively, wherever clarity is improved by the use of metalinguistic quantifiers and conjunction.

The symbols "$\in$," "$\subseteq$," "$\cup$," "$\cap$," and "$\mathcal{P}$" will be used for the usual set-theoretic relations and operations denoting membership, subset, set union, set intersection, and powerset, respectively; "$\subset$" is used for proper subset; "$\emptyset$" is used to denote the empty set. The symbol ω will be used to denote the set of finite ordinals, *i.e.*, natural numbers.[1]

The equality symbol, "$=$," will be used to denote equality among all types of objects. Note that the equality symbol and the quantifiers of the metalanguage will *not* be distinguished from their formal counterparts in first order languages. Wherever this may result in some confusion we refrain from using such symbols in the metalanguage and revert to English.

---

[1]ω is the least infinite ordinal and the smallest infinite cardinal, since we identify cardinals with *initial ordinals*.

3 Semantic Consequence Relations

Semantic consequence, or *logical consequence*, or simply *consequence*, as it is often called, forms the basis of the study of "correct reasoning" in the classical logic tradition. Although defined and studied with the utmost rigour, it expresses a simple to state and clear intuitive relation among sentences in a given language:

$$\text{sentence } \phi \text{ is true whenever sentences } \{\psi_1, \ldots, \psi_n\} \text{ are true.} \qquad (2.1)$$

Related to semantic consequence, are the syntactic notions of *proof* and *provability*. A proof is simply an argument that derives a sentence ϕ from a set of sentences $\{\psi_1, \ldots, \psi_n\}$. Clearly, for the notion of proof to bear any relation to normal English usage of the same word, we must place some constraints on the form of acceptable arguments, ruling out unsound or vague arguments. The most crucial point to note is that the soundness of an argument or method of arguing will ultimately depend on how well proofs mirror consequence, *i.e.*, whether proofs only yield true conclusions from true premises. The notions of proof and provability therefore are closely connected with the notions of truth and consequence.

Our study of minimal consequence is primarily motivated by practical considerations. This suggests that we value proofs and syntactic formalisms in general very highly. Although, as argued earlier, the semantic notions of consequence are fundamental, our study of these will concentrate on properties that have an impact on the existence of proof systems and syntactic characterizations.

3.1 Logical Consequence

The intuitive notion of logical consequence given in (2.1) above is made precise in logic by specifying what is meant by "sentence," "true" and "whenever." Paraphrasing (2.1) simplifies this task:

$$\text{In every situation where sentences } \{\psi_1 \ldots \psi_n\} \text{ are true, sentence } \phi \text{ is also true.} \qquad (2.2)$$

The approach in logic is to first fix a language, so that what constitutes a sentence is well understood. Then, with respect to this language it is possible to define set-theoretic structures that represent each of the possible situations that may arise. Finally, a definition of what is meant by being true in such a stucture is given.

The languages that have been studied in logic are very simple, as compared to human languages. But even for these simple languages the major obstacle has been the definition of truth — the definition of truth in first order logic was not given in a rigorous manner until this century [Tarski 35]. We will restrict attention to the two most studied types of languages in logic: the languages of sentence letters and connectives (sentential logic) and the first order languages of relation, function, and constant symbols with logical connectives, quantifiers, and equality (first order logic). Each of these is outlined in a section below.

Structures interpret the primitive elements of a language. The truth of sentences in a structure $\mathcal{M}$ is determined by the interpretation in $\mathcal{M}$ of the primitive elements that comprise them. For each language, there are usually many structures that interpret it. For a language $\mathcal{L}$ we will refer to the structures that interpret it as $\mathcal{L}$-structures; $\mathcal{L}$-sentences and $\mathcal{L}$-theories will be used to refer to sentences and theories (sets of sentences) of $\mathcal{L}$. When an $\mathcal{L}$-sentence ϕ is true in an $\mathcal{L}$-structure $\mathcal{M}$ we say that $\mathcal{M}$ *satisfies* ϕ (or that ϕ *holds* in $\mathcal{M}$, or that $\mathcal{M}$ is a *model* of ϕ) and write $\mathcal{M} \models \phi$. An $\mathcal{L}$-structure $\mathcal{M}$ satisfies an $\mathcal{L}$-theory Γ if and only if it satisfies each sentence of Γ. The set of models of a theory Γ are denoted $Mod_{\mathcal{L}}(\Gamma)$; $Mod_{\mathcal{L}}(\emptyset)$ (the models of the empty theory) thus denotes the set of all $\mathcal{L}$-structures, since, trivially, all $\mathcal{L}$-structures satisfy every sentence in $\emptyset$. Leaving the issue of languages and truth aside, we are now ready for a definition of logical consequence:

Definition 3.1 Let Γ be an $\mathcal{L}$-theory and ϕ an $\mathcal{L}$-sentence. $\Gamma \models \phi$ (ϕ is a *logical consequence* of Γ) iff for every $\mathcal{M}$ such that $\mathcal{M} \models \Gamma$, $\mathcal{M} \models \phi$.

Note that the symbol "$\models$" is used to denote both satisfaction and logical consequence. This is standard practice in the logic literature and no confusion should arise since, in the case of satisfaction, it denotes a relation between a model and a sentence (or a set of sentences), whereas, in the case of consequence, it denotes a relation between a theory and a sentence. In the special case where the theory Γ is the empty theory, in the above definition, the sentence ϕ is said to be *valid.*

Observe that we neither require that theories be closed under consequence nor that they be finite (as suggested by (2.2)). A theory is any set of sentences of the language and need not include the consequences of those sentences.

Definition 3.2 An $\mathcal{L}$-theory Γ is *complete* iff for every $\mathcal{L}$-sentence ϕ, either $\Gamma \models \phi$ or $\Gamma \models \neg\phi$.

Note that we will be using the words "complete" and "completeness" in several different (standard) senses in this thesis. Completeness of a theory, as we see here, means that the theory decides every sentence of its language. Below, we will see the term applied to a deductive calculus and also to a set in a complexity class — with different, though related, meanings.

So far we have intentionally avoided introducing language related issues. These are addressed in Sections 3.2 and 3.3 below. In Sections 3.4 and 3.5 we return to a general discussion of consequence relations.

3.2 Sentential logic

A language $\mathcal{L}$ of sentential logic is a set of sentence letters and sentences built up from them using the sentential connectives "$\neg$" and "$\wedge$." Given a set S of sentence letters, we can define a language for sentential logic as the smallest set $\mathcal{L}$ such that $S \subseteq \mathcal{L}$; if $\phi \in \mathcal{L}$, then $(\neg\phi) \in \mathcal{L}$; if $\phi \in \mathcal{L}$ and $\psi \in \mathcal{L}$, then $(\phi \wedge \psi) \in \mathcal{L}$. We will call the objects of the language $\mathcal{L}$, *$\mathcal{L}$-sentences* (or simply *sentences* when the language is clear from the context). Among these, we will use the term *atom* to refer to sentence letters and the term *basic formula* to refer to atoms and their negations. A set of $\mathcal{L}$-sentences is called a *theory.*

For convenience, as is usual, the symbols "$\vee$," "$\rightarrow$," and "$\leftrightarrow$" are introduced as abbreviations. Note that the language $\mathcal{L}$ is uniquely determined by the set S, and is of the same cardinality as S (if S is infinite), so from now on a language will be given as its set of atoms. The symbols p, q, r, s (possibly subscripted) will be used to denote distinct atoms of S. For a countable language, the atoms will typically be denoted by $p_1, p_2, \ldots$, although in some cases, for reasons of clarity of exposition, a number of additional symbols will be used, again with the assumption that distinct symbols denote distinct atoms of the language. Boldfaced versions of these symbols ($\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}, \boldsymbol{s}$) will be used where it is necessary to have a symbol ranging over the atoms of the language.

A structure for a language $\mathcal{L}$ is defined as a subset of S, so the set of structures for $\mathcal{L}$ is of cardinality 2^S; thus, the set of structures of any countably infinite language has the cardinal number of the continuum.

In a language of sentential logic the definition of truth is very simple — it is only necessary to specify how to interpret sentence letters and the two connectives, $\neg$ and $\wedge$:

Definition 3.3 Let $\mathcal{M}$ be an $\mathcal{L}$-structure and ϕ an $\mathcal{L}$-sentence. $\mathcal{M} \models \phi$ ($\mathcal{M}$ *satisfies a sentence* ϕ) iff

(i) ϕ is an atom and $\phi \in \mathcal{M}$; or

(ii) $\phi = \psi_1 \wedge \psi_2$ and $\mathcal{M} \models \psi_1$ and $\mathcal{M} \models \psi_2$; or

(iii) $\phi = \neg\psi$ and $\mathcal{M} \not\models \psi$

3.3 First-order logic

A first order language $\mathcal{L}$ is determined by a set of relation, function, and constant symbols, the equality symbol "=,"[2] and the first order connectives "$\neg$," "$\wedge$," and

[2]The equality symbol is sometimes considered optional for first order languages (see, for example, [Enderton 72], page 68). In this thesis we assume that it is present in all first order languages that we study.

"$\forall$." As such, a first order language provides a much greater richness of expression than the simple sentential languages which we encountered in the previous section. The definitions of what constitutes a sentence and truth in a structure also become more challenging.

We will use uppercase letters P, $R, \ldots$ to denote relation symbols, a, b, $c, \ldots$ for constant symbols, f, g, $h, \ldots$ for function symbols, and x, y, $z, \ldots$ for variables — all these possibly subscripted.

Definition 3.4 The set of *terms* of a first order language $\mathcal{L}$ is the least set T such that T contains all constant symbols and all variables of $\mathcal{L}$ and, whenever f is an n-placed function symbol and $t_1, \ldots, t_n \in T$, then $f(t_1, \ldots, t_n) \in T$.

Definition 3.5 The set of *atomic formulas* of a first order language $\mathcal{L}$ are strings of the form given below:

(i) $t_1 = t_2$ is an atomic formula, where t_1 and t_2 are terms of $\mathcal{L}$.

(ii) If R is an n-placed relation symbol and $t_1, \ldots, t_n$ are terms, then $R(t_1, \ldots, t_n)$ is an atomic formula.

Definition 3.6 The set of *formulas* of a first order language $\mathcal{L}$ is the least set Φ such that every atomic formula belongs to Φ and, whenever $\phi, \psi \in \Phi$, then $(\phi \wedge \psi)$, $\neg\phi$, and $\forall x(\phi)$, for each variable symbol x of the language, all belong to Φ.

For convenience, as is usual, the symbols "$\vee$," "$\rightarrow$," "$\leftrightarrow$," and "$\exists$" are introduced as abbreviations. An occurrence of a variable in a formula is said to be *bound* iff it is in the scope of one of the quantifiers governing it; otherwise it is said to be *free*. A *closed term* is a term with no variables. Similarly, a *closed formula* or a *sentence* is a formula with no free variables. Again, Greek letters will be used to denote sentences and formulas.

First order structures give meanings to terms, formulas, and sentences by interpreting the non-logical symbols of the language[3] — the relation, function, and constant symbols. We will again use $\mathcal{M}$, $\mathcal{N}$, ...to denote structures; M will denote the domain of $\mathcal{M}$; $R^{\mathcal{M}}$ will denote the interpretation of the relation symbol R in the structure $\mathcal{M}$ and similarly $c^{\mathcal{M}}$ and $f^{\mathcal{M}}$ for constant and function symbols. Thus we can write a structure as a tuple $\langle M, R_1^{\mathcal{M}}, \ldots, R_n^{\mathcal{M}}, f_1^{\mathcal{M}}, \ldots, f_m^{\mathcal{M}}, c_1^{\mathcal{M}}, \ldots c_q^{\mathcal{M}}\rangle$. Where it is necessary to refer to elements in the domain of a structure we will use boldfaced letters **a**, **b**, ..., possibly subscripted.

Let $\mathcal{M}$ be an $\mathcal{L}$-structure. Let α be an assignment to the variables of $\mathcal{L}$, *i.e.*, a function from the set of variables of $\mathcal{L}$ into M (the domain of $\mathcal{L}$-structure $\mathcal{M}$). Let $\bar{\alpha} : T \rightarrow M$. $\bar{\alpha}$ associates with each term of $\mathcal{L}$ a domain element which is to be understood as its interpretation. $\bar{\alpha}$ is defined from α as follows:

(i) $\bar{\alpha}(x) = \alpha(x)$, if x is a variable;

(ii) $\bar{\alpha}(c) = c^{\mathcal{M}}$, if c is a constant;

(iii) $\bar{\alpha}(f(t_1, \ldots, t_n)) = f^{\mathcal{M}}(\bar{\alpha}(t_1), \ldots, \bar{\alpha}(t_n))$ otherwise.

Definition 3.7 Let ϕ be an $\mathcal{L}$-formula. $\mathcal{M} \models \phi[\alpha]$ ($\mathcal{M}$ *satisfies* ϕ *under the assignment* α) iff one of the following holds:

(i) ϕ is an atomic formula $P(t_1, \ldots, t_n)$ and $P^{\mathcal{M}}(\bar{\alpha}(t_1), \ldots, \bar{\alpha}(t_n))$; or

(ii) ϕ is $t_1 = t_2$ and $\bar{\alpha}(t_1) = \bar{\alpha}(t_2)$;[4] or

(iii) ϕ is $\psi_1 \wedge \psi_2$ and $\mathcal{M} \models \psi_1[\alpha]$ and $\mathcal{M} \models \psi_2[\alpha]$; or

---

[3]The first order connectives, the equality symbol, parentheses, and the variable symbols are called the *logical symbols* of a first order language (their interpretation is fixed), whereas the relation, function, and constant symbols comprise the *non-logical symbols* of the language (the symbols that are open to interpretation).

[4]Note that we are using the symbol "=" both as a formal and a metalinguistic symbol.

(iv) ϕ is $\neg\psi$ and $\mathcal{M} \not\models \psi[\alpha]$; or

(v) ϕ is $\forall x\, \psi$ and for every assignment β, such that $\beta(y) = \alpha(y)$ for all $y \neq x$, $\mathcal{M} \models \psi[\beta]$.

It should be clear from this definition that if α and β are two assignments such that $\alpha(x) = \beta(x)$ whenever x is free in ϕ, then $\mathcal{M} \models \phi[\alpha]$ iff $\mathcal{M} \models \phi[\beta]$. We are now ready for the definition of satisfaction:

Definition 3.8 Let ϕ be an $\mathcal{L}$-sentence and $\mathcal{M}$ an $\mathcal{L}$-structure. $\mathcal{M} \models \phi$ ($\mathcal{M}$ *satisfies* ϕ) iff for some assignment (equivalently, for every assignment) α, $\mathcal{M} \models \phi[\alpha]$.

We now define some notions that can be used to compare first order structures.

Definition 3.9 $\mathcal{L}$-stuctures $\mathcal{M}$ and $\mathcal{N}$ are *elementarily equivalent* iff for any $\mathcal{L}$-sentence ϕ, $\mathcal{M} \models \phi \iff \mathcal{N} \models \phi$.

Definition 3.10 $\mathcal{L}$-stuctures $\mathcal{M}$ and $\mathcal{N}$ are *isomorphic* iff there is a one-one map $\mathcal{G}$ of M onto N satisfying:

(i) For each n-place relation $R^{\mathcal{M}}$ of $\mathcal{M}$ and the corresponding relation $R^{\mathcal{N}}$ of $\mathcal{N}$,

$$R^{\mathcal{M}}(\mathbf{a_1}, \ldots, \mathbf{a_n}) \text{ iff } R^{\mathcal{N}}(\mathcal{G}(\mathbf{a_1}), \ldots, \mathcal{G}(\mathbf{a_n}))$$

for all $\mathbf{a_1}, \ldots, \mathbf{a_n}$ in M;

(ii) For each n-place function $f^{\mathcal{M}}$ of $\mathcal{M}$ and the corresponding function $f^{\mathcal{N}}$ of $\mathcal{N}$,

$$\mathcal{G}(\mathbf{f}^{\mathcal{M}}(\mathbf{a_1}, \ldots, \mathbf{a_n})) = f^{\mathcal{N}}(\mathcal{G}(\mathbf{a_1}), \ldots, \mathcal{G}(\mathbf{a_n}))$$

for all $\mathbf{a_1}, \ldots, \mathbf{a_n}$ in M; and

(iii) For each constant $c^{\mathcal{M}}$ of $\mathcal{M}$ and the corresponding constant $c^{\mathcal{N}}$ of $\mathcal{N}$,

$$\mathcal{G}(c^{\mathcal{M}}) = c^{\mathcal{N}}$$

Definition 3.11 Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathcal{L}$-structures. $\mathcal{M}$ is a *substructure* of $\mathcal{N}$ iff

(i) $M \subseteq N$;

(ii) $R^{\mathcal{M}}$ is the restriction of $R^{\mathcal{N}}$ to M, for each relation symbol $R \in \mathcal{L}$;

(iii) $f^{\mathcal{M}}$ is the restriction of $f^{\mathcal{N}}$ to M, for each function symbol $f \in \mathcal{L}$; and

(iv) $c^{\mathcal{M}} = c^{\mathcal{N}}$, for each constant symbol $c \in \mathcal{L}$.

We say that $\mathcal{N}$ is an *extension* of $\mathcal{M}$ iff $\mathcal{M}$ is a substructure of $\mathcal{N}$.

In some contexts it is useful to consider the addition of some new symbols to a first order language and the enlargement of structures of that language to accomodate the new symbols, which we will call *expansions*. We emphasise the distinction between *extensions*, which increase the domain of a structure, and *expansions*, which increase the number of symbols interpreted by the structure.

3.4 A Generalization of Logical Consequence

As we stated in the beginning of Section 3.1, the definition of logical consequence aims to capture an intuitive notion, given in (2.2):

In every situation where sentences $\{\psi_1 \ldots \psi_n\}$ are true, sentence ϕ is also true.

In order to make this notion precise, we stated that it is necessary to define structures that represent all the possible situations that may arise (see Section 3.1). The class of all structures for a given language contains enough structures to represent all situations that can be (in set-theoretic terms) *expressed* for that language, but in the definition of logical consequence these are implicitly identified with all *possible* situations. This identification is not always justifiable, since the class of all structures, even for a sentential language, usually contains a great many structures which represent situations that the common person would deem unfathomable; a definition of consequence that requires that all these be taken into account in order to determine whether a sentence follows from a theory may be too restrictive in some cases. Without entering into a discussion of what might be meant by "possible situations," we may proceed by assuming that if these cannot be identified with the class of all the situations expressible for the language, they are at least included therein[5].

In the style of (2.2) above, we reinterpret a consequence relation between an $\mathcal{L}$-theory and an $\mathcal{L}$-sentence, relativizing it to a class of structures $\mathcal{C}$:

In every $\mathcal{L}$-structure $\mathcal{M}$ such that $\mathcal{M} \in \mathcal{C}$ and $\mathcal{L}$-theory Γ is true in $\mathcal{M}$, $\mathcal{L}$-sentence ϕ is also true. (2.3)

Logical consequence is then seen as a special case of consequence as defined in (2.3), where $\mathcal{C}$ is the class of all $\mathcal{L}$-structures.

We will not discuss at this point questions of definability of $\mathcal{C}$ (in the language of Γ or otherwise). There is little one can say about the structure or contents of $\mathcal{C}$ without delving into philosophical questions concerning the meaning of "possible situations," except that it is very likely that they depend on the contents of the theory Γ. This observation suggests a further refinement of (2.3), yielding the following definition:

Let $\mathcal{F} : \mathcal{P}(\mathcal{L}) \rightarrow \mathcal{P}(Mod_{\mathcal{L}}(\emptyset))$. An $\mathcal{L}$-sentence ϕ is an $\mathcal{F}$*-consequence* of an $\mathcal{L}$-theory Γ iff for every $\mathcal{M} \in \mathcal{F}(\Gamma)$ such that $\mathcal{M} \models \Gamma$, $\mathcal{M} \models \phi$.

Instead of fixing a "relevant" class of structures, $\mathcal{C}$, we now have a function from the set of $\mathcal{L}$-theories into the "powerset" of $\mathcal{L}$-structures, which picks a "relevant" class of $\mathcal{L}$-structures depending on the $\mathcal{L}$-theory. Without loss of generality, however, we may restrict $\mathcal{F}$ to be of type $\mathcal{F} : \mathcal{P}(\mathcal{L}) \rightarrow \mathcal{P}(Mod_{\mathcal{L}}(\Gamma))$ (since $Mod_{\mathcal{L}}(\Gamma) \subseteq Mod_{\mathcal{L}}(\emptyset)$

[5] A complementary concern is that the class of all (set-theoretic) structures is too narrow for assessing the consequence relation, *i.e.*, some possible situations are not representable by set-theoretic structures and hence not expressible by a language whose semantics is set-theoretic[Kreisel 67]. We do not engage this topic here.

and in the definition we are only interested in those elements of $\mathcal{F}(\Gamma)$ that are models of Γ). As with "$\mathcal{M} \models \Gamma$" ($\mathcal{M}$ is a model of Γ), we let "$\mathcal{M} \models_{\mathcal{F}} \Gamma$" stand for "$\mathcal{M} \in \mathcal{F}(\Gamma)$" (intuitively, $\mathcal{M}$ is a relevant model of Γ), arriving at our final definition of consequence relative to $\mathcal{F} : \mathcal{P}(\mathcal{L}) \rightarrow \mathcal{P}(Mod_{\mathcal{L}}(\Gamma))$:

Definition 3.12 An $\mathcal{L}$-sentence ϕ is an *$\mathcal{F}$-consequence* of an $\mathcal{L}$-theory Γ (written $\Gamma \models_{\mathcal{F}} \phi$) iff for every $\mathcal{L}$-structure $\mathcal{M}$, if $\mathcal{M} \models_{\mathcal{F}} \Gamma$, then $\mathcal{M} \models \phi$.

For logical consequence $\mathcal{F}(\Gamma) = Mod_{\mathcal{L}}(\Gamma)$, for every $\mathcal{L}$-theory Γ. As we will see, minimal consequence relations will be consequence relations where $\mathcal{F}(\Gamma)$ is identified with the set of all minimal models of Γ.

We say that Γ is *$\mathcal{F}$-satisfiable* iff $\mathcal{F}(\Gamma) \neq \emptyset$. For a consequence relation $\models_{\mathcal{F}}$, let $Cn_{\mathcal{F}}(\Gamma) = \{\phi | \Gamma \models_{\mathcal{F}} \phi\}$.

Definition 3.13 A consequence relation $\models_{\mathcal{F}}$ is *monotonic* iff $\forall\Gamma\forall\Delta(\Gamma \subseteq \Delta \Rightarrow Cn_{\mathcal{F}}(\Gamma) \subseteq Cn_{\mathcal{F}}(\Delta))$.

The consequence relations studied in logic are typically monotonic, but those that may be of interest in approaches to reasoning with incomplete information are most often non-monotonic. Thus, any type of defeasible reasoning is often referred to as "non-monotonic reasoning," and it includes minimal consequence relations, as we will see.

The next definition generalizes the notion of completeness (of a theory) to $\mathcal{F}$-completeness, also relativizing it to a set of sentences.

Definition 3.14 Let Θ be a set of $\mathcal{L}$-sentences. An $\mathcal{L}$-theory Γ is *$\mathcal{F}$-complete for* Θ iff for every ϕ in Θ, either $\Gamma \models_{\mathcal{F}} \phi$ or $\Gamma \models_{\mathcal{F}} \neg\phi$.

An $\mathcal{L}$-theory is $\mathcal{F}$-complete iff it is $\mathcal{F}$- complete for the set of all $\mathcal{L}$-sentences. We will loosely refer to $\mathcal{L}$-theories that are complete ($\mathcal{F}$-complete) for some set $\Theta \subset \mathcal{L}$ as *partially complete (partially $\mathcal{F}$-complete).*

3.5 Logical and Computational Considerations

There are a number of important model theoretic properties of languages and consequence relations which underlie their applicability in a variety of ways. The first and foremost, at least for computer scientists, is the existence of a proof system that captures the consequence relation. By this is meant a method of deriving consequences of theories that derives all (completeness) and only (soundness) the consequences of a theory. In addition, it is required that the proofs generated by such a system bear a strong resemblance to what, say, a mathematician would accept as a proof, *i.e.*, we would like to rule out vague proofs, or proofs that cannot be given in a finite amount of time or space. If there is to be a sound and complete proof system, the latter of these requirements promptly translates to a condition that proofs be finite, and therefore (as discussed below), that the consequence relation be compact; the former must remain an intuitive concern, because, as we will see, it translates to a requirement that the consequences of a theory be semi-decidable. The compactness and semi-decidability of a consequence relation are necessary and sufficient conditions for the existence of a sound and complete proof system.

Compactness

Perhaps the most basic property of a first order language from a model theoretic point of view is the compactness of the first order consequence relation.

Let S be a set and R a relation, $R \subseteq \mathcal{P}(S) \times S$.

Definition 3.15 R is *compact* iff for every $X \subseteq S$ and $a \in S$, if $R(X, a)$ then there is a finite $Y \subseteq X$ such that $R(Y, a)$.

For a consequence relation (*i.e.*, a relation between theories and sentences), the compactness property requires that, if a sentence is a consequence of a theory, then it is a consequence of a finite subset of that theory. If a deductive calculus can be constructed that will be capable of proving all the consequences of any theory of the languages under consideration (*e.g.*, first order languages), *i.e.*, a complete proof system exists, then it must *a priori* be the case that each of the consequences of a theory is also a consequence of a finite subset of the sentences of that theory (*i.e.*, a finite set of hypotheses). The compactness theorems for sentential and first order logic ensure us that this will not be an obstacle. On the other hand, from the fact that a sound and complete formalization exists, in which proofs are finite, it immediately follows that the consequence relation formalized is compact. For this reason, the compactness theorem for sentential and first order logic is often stated as a corollary to the completeness theorem.

The two (equivalent) formulations of compactness in classical logic are the following:

(i) If $\Gamma \models \phi$ then there exists a finite $\Delta \subseteq \Gamma$ such that $\Delta \models \phi$. (*Compactness of logical consequence*)

(ii) If every finite subset of Γ is satisfiable, then Γ is satisfiable. (*Compactness of satisfiability*)

These are equivalent, since $\Gamma \models \phi$ iff $\Gamma \cup \{\neg\phi\}$ is unsatisfiable. Moreover, the equivalence follows from the monotonicity of consequence. As we will see, minimal consequence is non-monotonic, so it will be important to distinguish between the two types of compactness defined above. In fact, for sentential logic, minimal satisfiability is compact, whereas minimal consequence is not (see Chapter 3, Section 3).

Semi-decidability

The existence of a sound and complete deductive calculus that captures a consequence relation by means of a reasonable notion of proof depends on yet another factor: the semi-decidability of the consequence relation between a theory and a sentence. Reasonable proofs, as discussed, must be finite. In addition, we would like to rule out proofs that are vague, *i.e.*, contain gaps or claims that require further proof.[6] Thus we require that a proof p of a sentence ϕ from a theory Γ be a finite syntactic object that can be checked by someone with access to Γ and ϕ, but with no special talents except clerical competence, *i.e.*, that there is an *effective procedure* that, given Γ, ϕ, and p, always terminates and outputs "yes" iff p is a proof of ϕ from Γ and "no" otherwise — in short, we require that the proof relation be *decidable.*

If there is to be a sound and complete deductive calculus where the proof relation is decidable, then the consequence relation must be semi-decidable. To see this, note that from an effective procedure that decides the proof relation, we can construct an effective procedure that, given Γ and ϕ, will output "yes" iff ϕ is a consequence of Γ. For example, this procedure could systematically examine initial segments of Γ, constructing ever longer proofs from sentences in each segment, so that all proofs are constructed in the limit; at each stage, the procedure examines the proofs constructed so far to determine if one of them is a proof of ϕ and, if so, it outputs "yes" and stops. Thus, if the deductive calculus is sound and complete, this procedure will output "yes" iff $\Gamma \models_{\mathcal{F}} \phi$.

Conversely, if the consequence relation is semi-decidable, then there exists a complete deductive calculus that incorporates a decidable proof relation. To see this, note that, given Γ and ϕ, the trace of a terminating computation of the semi-decision procedure that outputs "yes" iff $\Gamma \models_{\mathcal{F}} \phi$, can be viewed as a proof of ϕ from Γ (*i.e.*, when the trace is finite and the output is "yes"). Now, given Γ, ϕ, and p, it is possible to check that p is such a trace, so the proof relation is decidable. In addition, since there will be a terminating computation that outputs "yes" iff $\Gamma \models_{\mathcal{F}} \phi$, we have that the deductive calculus is complete.

At first glance, it may seem obscure how an intuitively conceived mechanical procedure of this sort could accept an input for Γ, which is not restricted to be

[6]Clearly, mathematicians do not place such stringent requirements on proofs. Mathematical rigour does, however, require at least the *existence* of such detailed proofs, to support the more informal arguments that are commonplace in mathematics.

finite. This can be visualized as follows: we assume that our idealized machine is connected to an infinite memory device containing Γ (this may be conceived as an *oracle* for Γ) and that any (converging) computation involves making only finitely many calls to this oracle; since the computation must be finite, only a finite amount of input will actually be used.

Formal characterizations of the notion of an effective procedure are numerous, and are all shown to be equivalent; this general agreement supports both the belief that the notion of an effective procedure is natural and useful and the claim that the proposed formalizations indeed define exactly the class of (intuitively conceived) effective procedures. The claim that the proposed formalizations include all effective procedures is called *Church's Thesis* (see [Rogers 67] for further discussion). The (partial) functions computed by any of the proposed formalizations of the notion of an effective procedure are called *partial recursive funtions.* The formal notion corresponding to decidability is that of recursiveness: a set (or relation) is *recursive* iff it has a (total) recursive characteristic function. A set (or relation) is *semi-recursive* iff it is the domain of a partial recursive function.

As noted above, exploring the computational properties of consequence relations requires formulating the notion of semi-decidability for "higher type" relations. That is, the consequence relation is a relation between a (possibly) infinite set of sentences and a sentence. The standard approach invokes the arithmetization of syntax (where sentences are encoded as natural numbers and sets of sentences by functions on natural numbers) and defines partial recursive functionals (corresponding to formalizations of effective procedures which accept functions as input), following the intuititive conception of an oracle machine given above, to arrive at a definition of semi-recursiveness that can be used to describe a consequence relation.

Note that the usual treatment of soundness and completeness for the propositional and predicate calculus engages neither formal notions of effectiveness nor complexity of relations of higher types. As we saw in the case of compactness, (classical) logical consequence enjoys certain properties that greatly simplify matters, specifically, by collapsing compactness of consequence and compactness of satisfiability. Similarly, the requirement of semi-decidability for logical consequence reduces to requirements for compactness of satisfiability and effective enumerability of validities (*i.e.*, the existence of an effective procedure that enumerates the valid sentences of the language). Moreover, since in the case of logical consequence these requirements are met, it is not necessary to bring in formal notions of effectiveness. [7]

Remarks

The foregoing remarks serve to motivate our consideration of compactness and complexity of consequence relations and to describe, in broad terms, the approach taken in this thesis. As we will see, the minimal consequence relations that we study are neither compact nor semi-recursive (see Chapter 3, Section 5 and [Schlipf 87]), so, by Church's Thesis it then follows that they are not semi-decidable and, therefore, there cannot be a sound and complete deductive calculus that captures them via a reasonable notion of proof. Indeed, we show that minimal consequence, even in the case of sentential logic, is a rather complex relation — it is as hard as deciding whether a given program (in a Turing equivalent language) terminates on every input.

This section has given an introduction to elements of logic, model theory, and recursion theory that will be used in subsequent chapters. Much of the material on logic and model theory is drawn from [Chang & Keisler 73] and [Enderton 72]; the material on recursion theory is drawn from [Rogers 67] and [Hinman 78].

[7]It should be clear that a formalization of the notion of effective procedure is only necessary in establishing negative results.

4 Incomplete Information

By incomplete information we mean that the collection of facts that are given — a *theory* — does not contain enough information to decide every question on these facts — *i.e.*, it is not the case that, for every sentence of the language of the theory, either it or its negation is a consequence of the theory. Clearly, there is nothing particularly odd about a theory that is not complete, from a logical or theorem proving point of view, but rather, the problem arises when it is desirable to treat it as a complete theory. There are two ways in which this can be the case.

First, it is possible that a theory is complete, but not feasibly axiomatizable, so that it is not usable by a theorem prover or a database system based on a theorem prover. One famous example of this is the airline reservation problem, where, even if we take the language to be finite (in this case, the set of pairs of cities and times), it is still very impractical to encode complete information. The usual solution is to encode all and only positive information (*e.g.*, source-destination pairs served and times) so that negative information is implicitly assumed when positive information is not present. The idea is that instead of using the intended (complete) theory, we can use an incomplete theory together with a systematic method of completing the theory.

The second way in which it would be desirable to have a complete theory, is in areas where the objective is to model the process of characterizing, through approximations, a single model. The goal is to analyse the behaviour of a person observing certain facts, whose assumption is that these facts partially describe a model of the world which he or she is trying to discover. For example, a scientist believes that there is some explanation for his or her observations of the world, be that the world of physical objects or an abstract world, such as that of arithmetic, *i.e.*, that there is a way in which the world is. In other words, the scientist believes that, among the possible models of the world consistent with the facts observed, there is one "real" model, which s/he wants to determine so as to be able to make predictions. The scientist will first choose a language so as to be able to keep track of observations and to characterize this model. To this end, since the information available will typically be incomplete, s/he will proceed by making reasonable conjectures that at least explain the set of facts observed and hypothesise other facts which s/he subsequently may test by performing experiments. Whether the scientist decides to conjecture a model or a set of possible models, the only means available for doing so will be by conjecturing a theory that characterizes it, which is more complete (has fewer models) than the theory consisting only of the facts observed. Though, in general, it will not be possible to characterize a single model up to elementary equivalence by using a discovery procedure of this sort, in a variety of interesting cases, single structures may even be characterizable up to isomorphism (among countable structures) by such techniques. Recently, a framework for studying scientific inference from this point of view has been developed in the context of theoretical Computer Science (see [Osherson & Weinstein 88]).

A person faced with "common sense" information about the world will behave in a similar manner, assuming that, even though s/he is not aware of the answer to some question, there is an answer. Using whatever means of intuition, common sense and experience that s/he is capable of, s/he can conjecture an answer that seems reasonable. To mention the inevitable bird example, suppose that you know that Tweety is a bird and are wondering whether Tweety can fly. If you are type A, you may assume that it can; if you are type B, you may assume that it cannot fly. Either way, the point is this: you will conjecture a theory which has fewer models than your observations, in an attempt to get at the ultimate truth about your bird problem.

4.1 Minimal Consequence

For each of the cases described above, and in many more, concerning incomplete information, whether this arises due to practical limitations or due to the nature of the problem, the objective is to formulate a way in which an incomplete theory can be completed, or partially completed. Generally speaking, this has been viewed as the problem of defining a new consequence relation which yields more conclusions than the original theory, or, alternatively, extending the theory with new consistent facts, with some ideology justifying their choice. Another way to view it, however, is as the problem of choosing, among the models of a theory, one model or a set of models, again with some intuitively motivated justification for that choice. Minimal consequence encapsulates a seemingly natural choice: the set of minimal models of a theory. The justification for this choice, on grounds of its applicability to common sense reasoning, learning or scientific discovery, must await a precise definition and study of its properties and is beyond the scope of this thesis.

Definition 4.1 Let $\subseteq_x$ be a partial order on the structures of a language $\mathcal{L}$. A structure is *x-minimal* iff it is minimal with respect to $\subseteq_x$.

Definition 4.2 Let Γ be an $\mathcal{L}$-theory and ϕ an $\mathcal{L}$-sentence. $\Gamma \models_x \phi$ (ϕ is an *x-minimal consequence* of Γ) iff ϕ is true in every x-minimal model of Γ.

From now on, we restrict attention to minimal consequence relations, so we will find it convenient to subscript consequence relations in accordance with the minimality criteria that give rise to them, rather than the function $\mathcal{F}$, as we have done so far; similarly for satisfiability and the set of consequences. Thus, instead of writing "$\models_\mathcal{F}$," for $\mathcal{F}(\Gamma) = \{\mathcal{M} \mid \mathcal{M}$ is an x-minimal model of $\Gamma\}$, we simply write "$\models_x$"; similarly we say that a theory is x-satisfiable when it has an x-minimal model and write "$Cn_x(\Gamma)$" for the set of x-minimal consequences of a theory Γ. This abuse of notation would make it difficult or obscure to discuss consequence relations that cannot be articulated as minimal consequence, but it is preferable, for our purposes, to avoid a proliferation of symbols.

Several choices of $\subseteq_x$ arise very naturally. In sentential logic we define minimal models with respect to the subset relation (see Chapter 3). In first order logic perhaps the most natural definition is in terms of the substructure relation, proposed originally by McCarthy in [McCarthy 77], where it is called *minimal entailment*, as a semantic counterpart of his domain circumscription formalism (discussed in Section 4.3 below). Other partial orders can easily be defined for structures with the same domain, focusing on containment relations among the extensions of some of the relations in the structures. We will call the earlier form of minimal consequence (in terms of substructures) *domain minimal consequence* (or *d-minimal consequence*) to distinguish it from the other types of first order minimal consequence that we study in this thesis: *predicate minimal consequence* and *formula minimal consequence* (*p-minimal and f-minimal consequence*, respectively). The ideas for p-minimal and f-minimal consequence can also be traced back to McCarthy, to his formalisms for predicate circumscription [McCarthy 80] and formula circumscription [McCarthy 84], although a rigorous semantic definition of f-minimal consequence first appeared in [Etherington 86]. p-minimal consequence is based on minimization of the extensions of a set of relations in the language — for structures with the same domain, interpreting all other symbols identically. f-minimal consequence is also based on minimization of the extensions of a set of relations in the language — for structures with the same domain, and interpreting all other symbols identically except for another set of relation symbols of the language, whose extensions are allowed to vary.

4.2 Closed World Reasoning

One goal for Computer Science and AI has been to build into computer programs the ability to "jump to conclusions." This has led to the development of frames [Minsky 75], scripts [Schank 77], truth maintenance systems [McAllester 78], [McAllester 80], negation as failure rules [Kowalski 79], non-monotonic logics [McDermott & Doyle 80] and various forms of non-monotonic reasoning systems, such as default reasoning [Reiter 80b] and circumscription [McCarthy 80], [McCarthy 84].

One major theme shared by many approaches to common sense reasoning and others addressing purely computational concerns in logic programming (see, *e.g.*, [Clark 78], [Kowalski 79]) and database theory (see, *e.g.*, [Reiter 78], [Gallaire et al 84]) is the assumption that all relevant positive information is known (or represented). When faced with incomplete information, an agent operating under this assumption, popularly known as the *closed world assumption*, will conclude that anything about the domain of interest which is not known to be true is false. Minimal consequence, *i.e.*, truth in all minimal models, is the consequence relation expressed in this type of reasoning. The domain of interest may be the extension of a predicate (or predicates), the number of individuals in the world, *etc*

The following are specific instances of the closed world assumption that govern query evaluation in logic programming and most database systems:

1. The *domain closure assumption* — which states that there are no other individuals than those in the database. Thus, for a given first-order theory Γ, the domain of interpretation for all the models of Γ is restricted to the smallest set which contains the individuals mentioned in the theory [Reiter 80a].

2. The *Unique Names Assumption* — which states that individuals with different names are different. This assumption is invoked whenever it seems reasonable to assume that all equalities among names, and, more generally, terms, in a database are already known, *i.e.*, different names denote different individuals [Reiter 80a].

3. The *negation as failure assumption* — which states that facts not known to be true are assumed to be false, *i.e.*, any atomic sentence which is not a (logical) consequence of a theory is assumed to be false.

In this thesis we will use negation as failure to mean the addition to a theory of the negations of all atomic sentences that are not provable from the theory. In logic programming negation as failure has a slightly more specific meaning, namely the addition to the database of the negations of sentences whose proof results in a finite failure [Clark 78]. This amounts to a weakened form of the closed world assumption. This distinction, although of great practical import, will not be of concern in this thesis, since it is intricately connected with syntactic matters lying beyond the scope of this work. [8]

Each of the assumptions listed above concerns a specific aspect of a logic database that is being "closed off." The domain closure and unique names assumptions can be realized by developing semantics restricted to Herbrand models and interpreting equality as a non-logical symbol (*i.e.*, viewing it as just another relation symbol of the language). A Herbrand model is a model whose domain consists of the closed terms of the language; each term is interpreted in a Herbrand model as "itself," *i.e.*, its interpretation is the syntactic object itself. If the (real) equality symbol were present, a difficulty would arise when two terms were equated by a theory, because then they would have to be interpreted as the same object.[9] This is not a problem, however, if equality is treated as a non-logical symbol, *i.e.*, if we leave it open to interpretation, as with the other relation symbols of the language.

Given the restriction to Herbrand models, this approach to equality makes it convenient to then collapse the unique names assumption under negation as failure (since now sentences asserting equality of terms can be assumed false if unprovable). For a comprehensive survey of work in semantics of databases and logic programming see [Gallaire et al 84].

In order for the interpretation of negation as failure to be coherent, we must be sure that the operation $\Delta(\Gamma)$, of adding the negations of atomic sentences unprovable in Γ to the theories Γ with which we are concerned, is consistency preserving.

---

[8]Another point worth noting, concerning terminology, is that some authors identify the closed world assumption with what we here call the negation as failure assumption and do not view the domain closure and unique names assumptions as special cases of it.

[9]This problem is overcome for languages with equality by taking equivalence classes of terms under equality. Although this is straightforward, it means that the Herbrand base, *i.e.*, the set of equivalence classes of terms that form the domain of the Herbrand model, will not in general be unique.

For standard versions of logic programming — where programs are, in effect, sets of universal closures of Horn clauses — the fact that Δ preserves consistency follows from Herbrand's theorem and the existence of unique minimal models for Horn theories in sentential logic (see also [Van Emden & Kowalski 76]). The operation $\mathbf{\Delta}$, however, is consistency preserving for a wider collection of first order theories than universal Horn theories. In Chapter 3 we give a precise characterization of the theories (*i.e.*, *subconditional theories*) which have unique minimal models in sentential logic. In Chapter 6, we discuss the possibility of using this result to characterize the class of first order theories which possess unique minimal Herbrand models.

It should be clear that the first order minimal consequence relations that are the subject of this thesis (d-, p-, and f-minimal consequence) are not defined in terms of minimal Herbrand models and represent a more general approach to closed world reasoning. The domain closure assumption is formulated as d-minimality, a minimality property defined in terms of arbitrary first order theories, not just universal Horn theories (see Chapter 4, Section 2). Similarly, p-minimality and f-minimality of structures can be used to enforce (generalizations) of negation as failure and the unique names assumption for first order theories (see Chapter 4, Sections 3 and 4). Indeed, a large part of the logic programming approach to the closed world assumption can be accomodated via our treatment of minimality for sentential theories via the reduction of satisfiability questions for skolemized first order sentences to questions of satisfiability of sentential theories achieved by Herbrand's Theorem.

The following sections give brief expositions of circumscription formalisms; these aim to formalize the first order minimal consequence relations which we discuss here, by introducing schemata or second order sentences that generate domain closure or negation as failure assumptions for (arbitrary) finitely axiomatizable first order theories. Although subsequent chapters do not engage their discussion further, since they concentrate on the purely semantical issues relating to minimal consequence, circumscription formalisms inspired much of the work reported in this thesis.

4.3 Domain Circumscription

Domain circumscription provides a syntactic way of conjecturing domain closure axioms. The domain circumscription of a sentence A is the schema:

$$[(\exists x\Phi(x) \wedge Axiom(\Phi) \wedge A^{\Phi}) \rightarrow \forall x\Phi(x)]^{\circ} \tag{2.4}$$

where Φ is any formula with at least one free variable, $[\Psi]^{\circ}$ denotes the closure of a formula Ψ (*i.e.*, the sentence obtained by prefixing Ψ by universal quantifiers with respect to all of its free variables), $Axiom(\Phi)$ is the conjunction of $\Phi(a)$ for each constant symbol a and $\forall x_1, \ldots, x_n(\Phi(x_1) \wedge \cdots \wedge \Phi(x_n)) \rightarrow \Phi(f(x_1, \ldots, x_n))$ for each n-ary function symbol f, and A^{Φ} is the result of rewriting A replacing each universal or existential quantifier "$\forall x$" or "$\exists x$" in A by "$\forall x\Phi(x) \rightarrow$" or "$\exists x\Phi(x) \wedge$," respectively [McCarthy 77,McCarthy 80].

[Davis 80] shows that domain circumscription is sound, *i.e.*, every instance of (2.4) is true in all minimal models of A. [10] He also shows that the converse (*i.e.*, completeness) is false and gives a partial completeness result for some classes of theories.

4.4 Predicate and Formula Circumscription

Predicate circumscription and *formula circumscription* are rules of conjecture developed by McCarthy [McCarthy 80,McCarthy 84] that provide partial characterizations of minimal consequence with respect to the extensions of some set of predicates. In predicate circumscription the extension of this set of predicates is minimized while the extensions of all other predicates are held fixed. In formula

[10]The proof in [Davis 80] omitted the case of theories with no constant symbols or existential sentences and in fact did not work for the original formulation of domain circumscription proposed in [McCarthy 77] which omitted the "$\exists x\Phi(x)$" conjunct from (2.4); (2.4) was proposed by Etherington in [Etherington 86] who also completed the soundness proof in [Davis 80].

circumscription there is an additional set of predicates that is allowed to vary while the first set of predicates is minimized. Predicate and formula circumscription are intended to model reasoning under the closed world assumption in the following sense: we assume that the objects which can be shown to have a certain property P, by reasoning from certain facts A, are all the objects that satisfy P.

Definition 4.3 The predicate circumscription of an n-ary predicate P in a sentence $A(P)$ is the sentence schema:

$$A(\Phi) \land \forall\vec{x}(\Phi(\vec{x}) \rightarrow P(\vec{x})) \rightarrow \forall\vec{x}(P(\vec{x}) \rightarrow \Phi(\vec{x})) \tag{2.5}$$

$A(P)$ is a sentence involving the predicate P and expresses that which is known so far by the reasoning agent. The predicate P is singled out as that with respect to which circumscription is carried out. The assumption is that all relevant facts about P are stated in $A(P)$ so circumscription enables a jump to the conclusion that what does not follow from $A(P)$ is actually false. Φ is any formula of the language of A on n free variables.

A generalization of (2.5) allows circumscribing several predicates jointly; jointly circumscribing P and Q in $A(P,Q)$ is given by the schema:

$$(A(\Phi,\Psi) \land \forall\vec{x}(\Phi(\vec{x}) \rightarrow P(\vec{x})) \land \forall\vec{y}(\Psi(\vec{y}) \rightarrow Q(\vec{y})))$$
$$\rightarrow (\forall\vec{x}(P(\vec{x}) \rightarrow \Phi(\vec{x})) \land \forall\vec{y}(Q(\vec{y}) \rightarrow \Psi(\vec{y})))$$

[McCarthy 80] shows that predicate circumscription is a sound formalization of p-minimal consequence, by showing that every instance of the circumscription schema is true in every p-minimal model of the sentence $A(P)$. This generalizes to the joint circumscription of multiple predicates. An argument given in [Davis 80] for domain circumscription can be adapted to show that the converse, *i.e.*, completeness of predicate circumscription as a formalization of p-minimal consequence, does not hold. [Perlis & Minker 76] give a number of partial completeness results in the case of various classes of theories.

Formula circumscription is a generalization of predicate circumscription that provides for the minimization of arbitrary formulas, rather than just predicates

[McCarthy 84]. Let $T(\mathbf{R})$ be a second order sentence and $\mathbf{R} = \{P_1, \ldots, P_n\}$, a set of predicate symbols that occur free in $T(\mathbf{R})$. Let $E(\mathbf{R}, \vec{x})$ be a second order formula in which the predicates in $\mathbf{R}$ and the variables in $\vec{x} = \{x_1, \ldots, x_m\}$ occur free. The formula circumscription of the formula $E(\mathbf{R}, \vec{x})$ in the sentence $T(\mathbf{R})$ is the second order sentence:

$$T(\mathbf{R}) \wedge \forall\vec{\Phi}((T(\vec{\Phi}) \wedge \forall\vec{x}(E(\vec{\Phi}, \vec{x}) \rightarrow E(\mathbf{R}, \vec{x}))) \rightarrow (\forall\vec{x}E(\mathbf{R}, \vec{x}) \rightarrow E(\vec{\Phi}, \vec{x}))) \quad (2.6)$$

where $E(\vec{\Phi}, \vec{x})$ is the result of replacing every occurrence of the predicate letters P_i in $E(\mathbf{R}, \vec{x})$ with predicate variables, Φ_i, of the same arity.

Etherington introduces the notion of f-minimal consequence as the semantic notion underlying formula circumscription and proves soundness [Etherington 86]. He also observes that minimizing the extensions of arbitrary formulas is equivalent to minimizing the extensions of some predicates, provided predicates other than those being minimized are allowed to vary. [Perlis & Minker 76] give a finitary completeness result for formula circumscription for a special class of theories.

5 Summary

As mentioned in the introduction, the starting point of this chapter is the problem of formulating a reasonable and computationally viable means of reasoning with incomplete information. The object of this chapter was to furnish the necessary background so that we are in a position to define and tackle the part of this problem that is the subject of this thesis — minimal consequence relations. To this end, we introduced a rigorous framework for the study of reasoning with incomplete information, based on a generalization of logical consequence. We then surveyed work on reasoning with incomplete information that is related to minimal consequence, aiming to provide a better understanding of the problem and to further motivate the type of study undertaken in this thesis.

Chapter 3

Minimal Consequence in Sentential Logic[1]

1 Introduction

This chapter presents a study of minimal consequence in sentential logic. The simplicity of sentential logic makes for a transparent exposition of many interesting aspects of the model theory and formalization of minimal consequence. This is of great significance to the development of automated systems for reasoning with incomplete information, in particular logic programming systems employing negation as failure. Certain properties particular to sentential logic, *e.g.*, that there is a decision procedure for validity, will reflect on aspects of minimal consequence. On the other hand, sentential logic already contains enough complexity so that many issues concerning important model theoretic features of more interesting languages are raised. Significant aspects of minimal consequence can thus be articulated and studied in a clear and natural way. The formulation and detailed study of minimal consequence presented in this chapter, not only subserves the applicability of logic programming formalisms, but also illuminates the development and study of minimal consequence for more sophisticated languages.

Section 2 gives the definitions of s-minimal model and s-minimal consequence, followed by examples that illustrate some of the key features of these notions.

[1] This chapter is an expanded version of [Papalaskari & Weinstein 90].

Section 3 contains an exposition of the central model theoretic properties of minimal consequence, namely minimal satisfiability of satisfiable theories and non-compactness. In section 4 we undertake a study of some interesting fragments; here we define subconditional theories and compare them to conditional (or Horn) theories; since this is of great computational import with regard to the consistent application of negation as failure, we offer a syntactic characterization of the class of subconditional theories and show that they are the largest class of theories that remain consistent under the application of negation as failure. Section 5 deals with complexity theoretic aspects of minimal consequence and subconditional theories; we show that the minimal consequence relation is Π_2^0 and not Σ_2^0 and that even in the case of theories with unique minimal models it is Δ_2^0 and neither r.e. or co-r.e., while the question of determining whether a theory has a unique minimal model is also Π_2^0 and not Σ_2^0.

2 Minimal Consequence in Sentential Logic

We begin by defining the notions of s-minimal model and s-minimal consequence.

Definition 2.1 $\mathcal{M} \models_s \Gamma$ (*$\mathcal{M}$ is an s-minimal model of* Γ) iff $\mathcal{M} \models \Gamma$ and $\forall \mathcal{N}(\mathcal{N} \models \Gamma \Rightarrow \mathcal{N} \not\subset \mathcal{M})$.

Definition 2.2 $\Gamma \models_s \phi$ (*ϕ is an s-minimal consequence of* Γ) iff $\forall \mathcal{M}(\mathcal{M} \models_s \Gamma \Rightarrow \mathcal{M} \models \phi)$.

Thus, a model $\mathcal{M}$ of a theory is s-minimal if the theory has no models that are proper submodels of $\mathcal{M}$ and a sentence ϕ is an s-minimal consequence of a theory Γ (or Γ *s-minimally entails* ϕ) if ϕ is true in all s-minimal models Γ. Note that the symbol "$\models_s$" is used both as a relation between models and theories and as a relation between theories and sentences, but no confusion should arise, since the meaning will always be clear from the context. Within this chapter we will drop the "s-" prefix from s-minimal model and s-minimal consequence. It was included in the definitions so that these notions could be differentiated from their first order counterparts, and will only be used when the distinction needs to be emphasized. Similarly we will often write "minimally satisfiable" instead of "s-satisfiable."

The following examples illustrate some interesting features of minimal consequence. In each of them it is assumed that Γ is a theory of a countable sentential language $\mathcal{L}$, determined by the set of sentence letters $S = \{p_i |\, i \in \omega\}$.

Example 2.1 Let $\Gamma = \emptyset$. The set of models of Γ is $\mathcal{P}(S)$ and so $\emptyset$ is the unique minimal model for Γ. Thus $\Gamma \models_s \neg p_1 \wedge \ldots \wedge \neg p_n$, for $p_i \in S$, $1 \leq i \leq n$. $\Gamma \cup \{p_1 \wedge \ldots \wedge p_n\}$ has a unique minimal model also, namely $\{p_i |\, 1 \leq i \leq n\}$, so $\Gamma \cup \{p_1 \wedge \ldots \wedge p_n\} \not\models_s \neg p_1 \wedge \ldots \wedge \neg p_n$.

We immediately see from this simple example that the relation $\models_s$ is non-monotonic.

Example 2.2 Let $\Gamma = \{p_{2i} \vee p_{2i+1} |\, i \in \omega\}$. Every model of Γ must contain either p_{2i} or p_{2i+1}, or both, for each $i \in \omega$. The minimal models of Γ will be the ones that contain exactly one of p_{2i} or p_{2i+1}. So the set of minimal models of Γ is of cardinality 2^ω, ie, the cardinal number of the continuum.

These examples depict two extreme cases: $Cn_s(\Gamma)$ in example 2.1 is complete, whereas $Cn_s(\Gamma)$ in example 2.2 has continuum many models.

Intuitively, the number of models of a theory is an indication of its degree of completeness. As more sentences are added to a theory, its number of models decreases (if these sentences were not already consequences of the theory). Complete theories have unique models and hence, unique minimal models. As we will see, the number of minimal models of a theory is, roughly, an indication of the ease with which it can be completed, while retaining its basic structure.

One systematic way in which a theory can be so completed is by adding the negation of some set of sentences which are not consequences of the theory. Using minimal consequence, the theory of example 2.1 can be completed by the addition of the negation of every sentence letter in the language. The theory of example 2.2, however can be completed in 2^ω different ways, but there is no apparent reason for choosing one over any of the others. The difference between those two examples lies in that the former is completed in a very natural way, *i.e.*, positive sentences which are not consequences are now refuted by the theory, while in the latter, no matter how we decide to complete it, some of the positive sentences which were not consequences will now be refutable and others will become consequences.

Note that, in this, positive and negative sentences are not treated evenly. In particular, positive information appears to be favoured, in the sense that it is taken to be more significant — and significant in the sense that it is assumed that it will be given if in fact it holds. Of course, we could take the opposite approach. That is, Γ of example 2.2 can also be completed by the addition of the sentences $\{p_i | i \in \omega\}$, thus making every positive literal valid (and thus favouring negative information, in the sense that "what is not known is assumed to be true"), but then all structure of Γ is lost. (The same, of course, holds for Γ of example 2.1.)

These observations indicate that adding the negation of a set of sentences that are not consequences of a theory is a plausible way to complete that theory; it is worthy of consideration, since, typically, it is not destructive to the structure of the original theory and appears to conform with some human intuition, namely that what is not asserted does not hold (although, at this point we would not like to press this last issue). It remains to make this process more precise and examine the cases where it is most likely to succeed.

Recall that another (equivalent) way to view the process of completing a theory is as singling out one of its models. If a theory has a unique minimal model, this would be an obvious choice. For theories that do not possess unique minimal models, we can proceed in the same manner, by first narrowing down to its minimal models and then picking one among those. The second step will necessarily embody some arbitrariness. The idea of completing a theory by selecting one among its minimal models corresponds to adding the negations of atoms which are not consequences. Whether it is possible to consistently add all such sentences to the theory, on the other hand, depends on whether the theory has a unique minimal model. For the theory of example 2.1, which has a unique minimal model, we can indeed add the negation of every atom (since Γ has no consequences that are not tautologies). For Γ of example 2.2, which is very far from having a unique minimal model, we can add the set of sentences $\Sigma = \{\neg(p_{2i} \wedge p_{2i+1}) | \, i \in \omega\}$, but, although for any $i \in \omega$, neither p_{2i} nor p_{2i+1} are consequences of Γ, $\Gamma \cup \{\neg p_{2i}, \neg p_{2i+1}\}$ is inconsistent. Adding Σ to Γ corresponds to the first step above, of narrowing down to Γ's minimal models, for it is easy to see that the set of models of $\Gamma \cup \Sigma$ is precisely the set of minimal models of Γ. The next step involves arbitrarily adding either $\neg p_{2i}$ or $\neg p_{2i+1}$ to Γ (but not both).

We can now rephrase an earlier comment about the effect of the number of minimal models of a theory. The number of minimal models of a theory is proportional to the degree of arbitrariness required in completing the theory and thus theories that do not have unique minimal models will pose limits to the minimal models method. This exposition, which was chosen because it separates the well defined from the arbitrary aspects of this strategy of completing a theory, was in terms of two steps, first narrowing down to the minimal models and then picking one among these. Note, however that no connection is drawn between the first step of narrowing down to the minimal models and the set of sentences added to the theory. It may be felt that, since any arbitrariness is contained in the second step for theories with more than one model, it would be adequate to only partially complete a theory, by performing only the first step. However, as we see from the next example, this is not always possible.

Example 2.3 Let $\Gamma = \{p_i \vee p_j | \, i \in \omega, \, j \in \omega, \, i \neq j\}$. Γ has sentences of the form $\neg p_i \rightarrow p_j$, for $j \neq i$, so in any model $\mathcal{M}$ of Γ, if $p_i \notin \mathcal{M}$, then for all $j \neq i$, $p_j \in \mathcal{M}$. Thus the models of Γ are those where at most one element is missing and the minimal models of Γ are those where exactly one element is missing.

The importance of this example lies in the fact that the set of structures $\Re = \{\mathcal{M}_i | \, p_i \notin \mathcal{M}_i, \, p_j \in \mathcal{M}_i, \, i \neq j\}$ cannot be characterized as the set of models of any theory, as can be shown by a straightforward application of the compactness theorem. Hence, $Cn_s(\Gamma) = Cn(\Gamma)$ and there is no set of sentences Σ such that the models of $\Gamma \cup \Sigma$ are exactly the minimal models of Γ.

We conclude that, in completing a theory, it is not always possible to first partially complete it (down to the minimal models) as was suggested in the discussion of examples 2.1 and 2.2; as we see here, there is no set of sentences that could be added to Γ so as to achieve this, since no extension of Γ has as its models exactly the set of minimal models of Γ. On the other hand, if a theory has a finite number of minimal models, it is always possible to find an extension of that theory that has these as its models, since any finite set of structures can be characterized as the models of some theory. Thus, it is possible to separate the well defined from the arbitrary steps in completing a theory with a finite number of minimal models, and the arbitrariness will be proportional to the number of minimal models. For theories with unique minimal models the first step will be sufficient to complete them. Thus, theories with unique minimal models form a special class and section 4 will be devoted to their study.

Although there are sets of structures that cannot be characterized as the models of any theory, but that can be characterized as the set of minimal models of some theory, there are also sets of structures which cannot be characterized as the minimal models of any theory. Trivially, any set that contains structures $\mathcal{M}$ and $\mathcal{N}$ with $\mathcal{M} \subset \mathcal{N}$ is such a set. The question now arises whether there are sets that do not contain any two structures such that one includes the other (in this sense, *incomparable structures*), and which cannot be characterized as the set of minimal models of any theory. The answer to this question is necessarily positive, because the number of theories of a countably infinite language is of cardinality 2^{ω}, while there are $2^{2^{\omega}}$ sets of incomparable structures. To show that there are $2^{2^{\omega}}$ sets of incomparable structures, first note that there are $2^{2^{\omega}}$ sets of structures for any countably infinite language. Next, we define a 1-1 mapping h from the set of all structures into itself, such that, if $\mathcal{M}$ and $\mathcal{N}$ are two distinct structures, $h(\mathcal{M})$ and $h(\mathcal{N})$ are incomparable, thus embedding the set of all structures into the set of all incomparable structures (from which it follows that they are of the same cardinality). Let $f : \mathcal{P}(S) \times \omega \rightarrow \{0,1\}$, $g : \mathcal{P}(S) \times \omega \rightarrow \{0,1\}$ and

$h : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$ be defined as follows:

$$
\begin{aligned}
f(\mathcal{M}, i) &= \begin{cases} 1 & \text{if } p_i \in \mathcal{M} \\ 0 & \text{otherwise} \end{cases} \\
g(\mathcal{M}, i) &= (f(\mathcal{M}, \lfloor \tfrac{i}{2} \rfloor) + i) \bmod 2 \\
h(\mathcal{M}) &= \{p_i | \, g(\mathcal{M}, i) = 1\}
\end{aligned}
$$

($\lfloor \frac{n}{m} \rfloor$ denotes integer division.) Intuitively, the effect of f is to represent a structure as a string of 0's and 1's; g creates a new string by mapping each 0 in the string created by f to a pair "0 1" and each 1 to a pair "1 0"; h then maps the strings created by g to structures. Clearly, if $\mathcal{M} \subset \mathcal{N}$, then some element of $\mathcal{N}$ not in $\mathcal{M}$ will be mapped by f to a 0 in $\mathcal{M}$ and to a 1 in $\mathcal{N}$; thus, some element will be mapped by g to a pair "0 1" in $\mathcal{M}$ and to a pair "1 0" in $\mathcal{N}$; hence, $h(\mathcal{M})$ will contain an element (corresponding to the 1 in the "0 1") which is not in $h(\mathcal{N})$. Therefore, for any two distinct structures $\mathcal{M}$, $\mathcal{N}$, $h(\mathcal{M})$ and $h(\mathcal{N})$ are incomparable, so any set of structures can be mapped to a set of incomparable structures. Since the set of all sets of structures is of cardinality 2^{2^ω}, it follows that there are 2^{2^ω} sets of incomparable structures, while there are only 2^ω theories for any countably infinite language. Thus, there are (many!) sets of incomparable structures that cannot be characterized as the minimal models of any theory.

A specific example of a collection of structures (in fact, a countable collection) that cannot be characterized as the set of minimal models of any theory is the set $\Re' = \{\{p_i\} | \, i \in \omega\}$. As was the case with the minimal models of the theory in example 2.3, any theory that is satisfied by all $\mathcal{M} \in \Re'$, will also be satisfied by the empty model (again, by compactness); but in this case, the empty model will be the unique minimal model of that theory.

3 Existence of Minimal Models

Due to the simplicity of sentential logic, the intersection of every chain (under the submodel relation) of models of a theory will be a model of the theory and thus every satisfiable theory is *minimally satisfiable*, *i.e.*, has a minimal model. This is the content of proposition 3.1[1].

Proposition 3.1 *Every satisfiable theory is minimally satisfiable.*

Proof: It suffices[4] to show that any maximal chain of models of a theory (ordered by the submodel relation) will contain a minimum element.[2]

Suppose that the set of models of a theory Γ contains a chain of models $\mathcal{M}_i$ such that $\mathcal{M}_i \subseteq \mathcal{M}_{i-1}$ and $\bigcap_i \mathcal{M}_i = \mathcal{M}$, but $\mathcal{M} \not\models \Gamma$.[3] Let $\Sigma = \mathcal{M} \cup \{\neg \boldsymbol{p} \mid \mathcal{M} \models \neg \boldsymbol{p}\}$. $\mathcal{M}$ is the only model of Σ, thus $\Gamma \cup \Sigma$ is inconsistent. By the compactness theorem, there is a finite $\Delta \subseteq \Sigma$ such that $\Gamma \cup \Delta$ is inconsistent.

Let $\Delta = \{p_1, p_2, \ldots, p_m, \neg q_1, \neg q_2, \ldots, \neg q_n\}$. It follows that

$$\Gamma \models \neg(p_1 \wedge p_2 \wedge \ldots \wedge p_m \wedge \neg q_1 \wedge \neg q_2 \wedge \ldots \wedge \neg q_n).$$

---

[1]This proposition is related the well known result of [Van Emden & Kowalski 76], which states that the intersection of the Herbrand models of a set of Horn clauses is itself a model, in fact the unique minimal Herbrand model. The connection is via the correspondence of the Herbrand models of a skolemized set of sentences to the (sentential) models of the Herbrand expansion of that set of sentences.

[2]By "chain of models" here we mean a set of models of a theory that is linearly ordered by the substructure relation (*i.e.*, the subset relation, since we are now working in sentential logic). A maximal chain of models is a chain of models such that there is no chain of models that properly includes it.

[3]Note that we place no restriction on the cardinality of the language of Γ or the chain of models $\mathcal{M}_i$.

[4]By Zorn's Lemma.

Since $\Delta \subseteq \Sigma$, we have that $\mathcal{M} \models p_1 \wedge \ldots \wedge p_m$ and, therefore, for all $\mathcal{M}_i$, $\mathcal{M}_i \models p_1 \wedge \ldots \wedge p_m$. We also have that, for each j, $q_j \notin \mathcal{M}$, and hence there is an $\mathcal{M}_k$ such that, for each j, $q_j \notin \mathcal{M}_k$. (Recall that $\mathcal{M}$ is the intersection of a *chain* of models $\mathcal{M}_i$.)

It follows that $\mathcal{M}_k \models p_1 \wedge p_2 \wedge \ldots \wedge p_m \wedge \neg q_1 \wedge \neg q_2 \wedge \ldots \wedge \neg q_n$, and thus $\mathcal{M}_k$ is not a model of Γ. This is a contradiction, so we conclude that $\mathcal{M} \models \Gamma$. □

Corollary 3.1 *Minimal satisfiability is compact.*

Proof: Immediate from the proposition and the compactness of satisfiability in sentential logic. □

Proposition 3.1 is of relevance with respect to some of the intended applications of minimal consequence, where, in general, it is crucial that any sentences added to the consequences of a theory preserve consistency. As discussed earlier these have to do with completing a theory by adding sentences true in all minimal models (but not necessarily all models) to the theory. Clearly, if it were the case that some theory has no minimal models, it would mean adding every sentence of the language, which would result in an inconsistent theory. No such problem arises in sentential logic, its language and model theory being so simple that every element of a model corresponds to a proposition letter in the language, and there is no way of constructing a chain of submodels such that their intersection is not a model of the theory. However, minimal satisfiability of satisfiable theories is not retained in the usual notions of minimal consequence for first order languages.[Davis 80]

Corollary 3.1 suggests that minimal consequence bears a strong similarity to logical consequence and, in particular, that the minimal consequence relation may be compact. As discussed in Chapter 2, the two (equivalent) formulations of compactness for sentential logic are the following:

(i) If $\Gamma \models \phi$ then there exists a finite $\Delta \subseteq \Gamma$ such that $\Delta \models \phi$. (*Compactness of logical consequence*)

(ii) If every finite subset of Γ is satisfiable, then Γ is satisfiable. (*Compactness of satisfiability*)

The equivalence follows from the fact that $\Gamma \models \phi$ iff $\Gamma \cup \{\neg\phi\}$ is unsatisfiable. Due to non-monotonicity phenomena, however, the corresponding formulations of compactness for minimal consequence and minimal satisfiability are *not* equivalent. That is, it is possible that $\Gamma \models_s \phi$, although $\Gamma \cup \{\neg\phi\}$ is (minimally) satisfiable. (For example, $\emptyset \models_s \neg p$, but $\{p\}$ clearly has a minimal model.) Thus the question as to whether it is at all possible to construct a complete logical calculus for minimal consequence is not settled by corollary 3.1 and to this end it will be necessary to inquire whether, if a sentence is a minimal consequence of a theory, then it is a consequence of a finite subset of that theory. The answer to this question is no.

Proposition 3.2 *The minimal consequence relation is not compact, ie, there is a set of sentences Γ and a sentence ϕ such that (i) $\Gamma \models_s \phi$, and (ii) $\forall\Delta \subseteq \Gamma(\Delta$ is finite $\Rightarrow \Delta \not\models_s \phi)$.*

Proof: Let $\Gamma = \Gamma_1 \cup \Gamma_2$, where

$$\begin{aligned} \Gamma_1 &= \{p_1 \vee \bigwedge_{3 \leq i \leq 2n+1} p_i | \, n \geq 1\}, \text{ and} \\ \Gamma_2 &= \{p_2 \vee \bigwedge_{3 \leq i \leq 2n} p_i | \, n \geq 1\}. \end{aligned}$$

Note that $\Gamma \models_s (p_1 \leftrightarrow p_2) \wedge (p_1 \vee p_3)$, but for all finite $\Delta \subseteq \Gamma$, $\Delta \not\models_s (p_1 \leftrightarrow p_2) \wedge (p_1 \vee p_3)$. This is because Γ has exactly two minimal models: one in which p_1 and p_2 are true and all other letters are false and another where p_1 and p_2 are false and all other letters are true; both of these make $(p_1 \leftrightarrow p_2) \wedge (p_1 \vee p_3)$ true. On the other hand, any finite subset Δ of Γ will fail to minimally entail $(p_1 \leftrightarrow p_2) \wedge (p_1 \vee p_3)$ for one of the following reasons:

(a) $\Delta \cap \Gamma_1 = \emptyset$ and thus $\Delta \not\models_s p_1 \vee p_3$;

(b) $\Delta \cap \Gamma_2 = \emptyset$ and thus $\Delta \models_s \neg p_2$, but $\Delta \not\models_s \neg p_1$, so $\Delta \not\models_s p_1 \leftrightarrow p_2$;

(c) there will be a largest m for which either the sentence $p_1 \vee \bigwedge_{3 \leq i \leq m} p_i$ or the sentence $p_2 \vee \bigwedge_{3 \leq i \leq m} p_i$ will be in Δ; supposing the former is the case, Δ will then have exactly three minimal models:

(i) $\mathcal{M}_1 = \{p_1, p_2\}$

(ii) $\mathcal{M}_2 = \{p_3, \ldots, p_m\}$

(iii) $\mathcal{M}_3 = \{p_1, p_3, p_4, \ldots, p_{m'}\}$, for $m' < m$

and $p_1 \leftrightarrow p_2$ will fail to hold in $\mathcal{M}_3$. Similarly for the case where the largest m for which p_m occurs in Γ appears in a sentence of the form $p_2 \vee \bigwedge_{3 \leq i \leq m} p_i$.

We have exhibited a theory Γ and a sentence ϕ such that $\Gamma \models_s \phi$, but for no finite subset Δ of Γ, $\Delta \models_s \phi$. Thus, the minimal consequence relation is not compact. □

The above result settles negatively the question of the existence of a complete finitary logical calculus for minimal consequence. Although the practical implications at this point may appear very grim, there are, as we see in the next section, fragments for which it is of demonstrated practical value. Moreover, the absence of compactness attests to a greater richness of expression, a fact that is very relevant to the obvious need in AI for more powerful formalisms.

4 Unique Minimal Models

The idea of adding the negation of every sentence letter which is not provable, is not a novel one; it has been studied extensively in Computer Science in connection with theorem proving for Horn clauses and logic programming, and is generally referred to as *negation as failure*. In general, negation as failure as applied to a consistent, incomplete theory results in an inconsistent theory. Consider, for example the theory $\Gamma = \{p \vee q\}$. Γ is consistent but $\Gamma' = \Gamma \cup \{\neg p, \neg q\}$ is inconsistent, although neither p nor q is a consequence of Γ. This motivates the search for a class of consistent theories that remain consistent under the application of negation as failure. A necessary and sufficient condition to this end is that a theory has a unique minimal model and is given in the following lemma:

Lemma 4.1 *Let Γ be a theory in a language $\mathcal{L}$. $\Gamma \cup \{\neg \boldsymbol{p} \mid \Gamma \not\models \boldsymbol{p}$ and $\boldsymbol{p} \in \mathcal{L}\}$ is consistent $\iff$ Γ has a unique minimal model.*

Proof: ($\Rightarrow$) Let $\Gamma^{neg} = \{\neg \boldsymbol{p} \,|\, \Gamma \not\models \boldsymbol{p} \text{ and } \boldsymbol{p} \in \mathcal{L}\}$ and suppose that $\Gamma \cup \Gamma^{neg}$ is consistent. Let $\mathcal{M} \models \Gamma \cup \Gamma^{neg}$. Then $\mathcal{M} = \{\boldsymbol{p} \,|\, \Gamma \models \boldsymbol{p}\}$. Hence, if $\mathcal{N} \models \Gamma$, then $\mathcal{M} \subseteq \mathcal{N}$. Therefore, $\mathcal{M}$ is the unique minimal model of Γ.

($\Leftarrow$) Suppose Γ has a unique minimal model $\mathcal{M}$. Note that

$$\Gamma^{neg} = \{\neg \boldsymbol{p} |\, \Gamma \not\models \boldsymbol{p} \text{ and } \boldsymbol{p} \in \mathcal{L}\} = \{\neg \boldsymbol{p} |\, \exists \mathcal{N}(\mathcal{N} \models \Gamma \text{ and } \mathcal{N} \models \neg \boldsymbol{p}) \text{ and } \boldsymbol{p} \in \mathcal{L}\}$$

and, since $\mathcal{M}$ is the minimal model of Γ, the latter equals $\{\neg \boldsymbol{p} |\, \mathcal{M} \models \neg \boldsymbol{p}\}$. Thus $\mathcal{M} \models \Gamma^{neg}$ and, therefore, $\Gamma \cup \Gamma^{neg}$ is consistent. □

Thus, the class of theories that can be completed using negation as failure is exactly the class of theories with unique minimal models. Observe also that a theory Γ with a unique minimal model shows a form of completeness: for each sentence, ϕ in the language of Γ, either $\Gamma \models_s \phi$ or $\Gamma \models_s \neg\phi$ (this fact follows directly from the definition of minimal consequence). For this reason, theories that have a unique minimal model will be referred to as minimally complete.

Definition 4.1 A theory is *minimally complete* iff it has a unique minimal model.

As was noted above, minimally complete theories are significant to computer scientists via their connection to negation as failure and therefore the question to be addressed next is their syntactic characterization. The first candidate class of minimally complete theories are conditional theories:

Definition 4.2 *Conditional theories* (or *Horn theories*) are sets of sentences each of which is in of one of the following forms:

(i) $\boldsymbol{p}$

(ii) $\neg \boldsymbol{q}_1 \vee \ldots \vee \neg \boldsymbol{q}_m$

(iii) $\neg \boldsymbol{q}_1 \vee \ldots \vee \neg \boldsymbol{q}_m \vee \boldsymbol{p}$

Horn theories have been extensively studied in computer science and it is a well known result that they remain consistent under negation as failure. Apart from the interest they offer in connection to negation as failure, conditional theories have been an object of study in model theory because they exhibit a preservation property for model intersections. A theory is said to preserve intersections iff for any set of its models, their intersection is also a model. A converse of the above is also true, namely, every theory that preserves intersections is equivalent to a conditional theory. Clearly, a theory that preserves intersections must have a unique minimal model. Thus conditional theories provide a partial characterization of the class of minimally complete theories and, therefore, by lemma 4.1, of the class of theories where negation as failure can be consistently applied. The converse, however is not true: it is possible for a theory to be minimally complete without preserving intersections. This is clear from the following trivial example. Let $\Gamma = \{\neg p \vee q \vee r\}$; Γ has a unique minimal model, namely the empty set, but it also has the models $\{p, q\}$ and $\{p, r\}$ whose intersection is not a model of Γ. Of course, Γ is not a conditional theory either, but Γ can be consistently completed via negation as failure in this case, since it is minimally complete. Examples such as this suggest a weaker preservation property for minimally complete theories, namely that the intersection of a set of models of the theory contains a model of the theory, and motivate the following definition:

Definition 4.3 A theory Γ is *subconditional* iff every sentence of Γ is in one of the following the forms:

(i) $p_n,\ n \in I,\ I \subseteq \omega$

(ii) $\bigvee_{n\in J} \neg p_n \vee \bigvee_{n\in K} p_n,\ J \subseteq \omega,\ K \subseteq \omega$, J and K finite, and $J \not\subseteq I$

In intuitive terms, a subconditional theory consists of a set of positive literals (indexed by a set I) and a set of disjunctions; each of the disjunctions contains one or more negative literals (indexed by a finite set J), at least one of which does not appear in isolation as a positive literal, and zero or more positive literals (indexed by a finite set K). The theory mentioned above is a subconditional theory. Another simple example is the theory $\{p_4, p_7, \neg p_1 \vee p_2 \vee p_3, \neg p_2 \vee \neg p_4 \vee p_3 \vee p_7\}$.

The aim of this section is to provide a complete characterization of minimally complete theories. We proceed by showing that the class of subconditional theories contains exactly those theories that are equivalent to some minimally complete theory. As was suggested earlier conditional theories fail in this respect due to a slightly stronger preservation property. The preservation property required for minimally complete theories is given in proposition 4.1 below.

Proposition 4.1 *A consistent theory* Γ *is minimally complete* $\iff$ *the intersection of every set of models of* Γ *contains a model of* Γ.

Proof: ($\Rightarrow$) Let Γ be a minimally complete theory and let $\mathcal{M}^*$ be the unique minimal model of Γ. Since $\mathcal{M}^*$ is minimal, it is contained in every model of Γ, and thus, in the intersection of any set of models of Γ.

($\Leftarrow$) Let Γ be a consistent theory and suppose the intersection of any set of models of Γ contains a model of Γ. If $\mathcal{N}$ and $\mathcal{K}$ are minimal models of Γ, then their intersection will contain a model $\mathcal{M}^*$ of Γ, but since $\mathcal{N}$ and $\mathcal{K}$ are minimal, $\mathcal{N} = \mathcal{K} = \mathcal{M}^*$. □

Although proposition 4.1 is useful in emphasising the distinction between minimally complete theories (which preserve a submodel of intersections) and conditional theories (which preserve all intersections), rather than showing that subconditional theories have the desired preservation property, we will directly show that subconditional theories are the desired characterization and thus obtain the preservation property as a corollary.

Proposition 4.2 *A theory* Γ *is minimally complete* $\iff$ Γ *is equivalent to a subconditional theory.*

Proof: ($\Rightarrow$) Suppose Γ has a unique minimal model $\mathcal{M}$. Let

$$\begin{aligned} I &= \{n \mid p_n \in \mathcal{M}\}, \\ \Gamma^p &= \{p_n \mid n \in I\} \text{ and} \\ \Gamma^c &= \{\phi \mid \phi \text{ is the conjunctive normal form of } \psi \text{ and } \psi \in \Gamma\} \end{aligned}$$

Thus,

$$\Gamma^c = \{ \bigvee_{n\in J_i} \neg p_n \vee \bigvee_{n\in K_i} p_n | i \in \omega\}$$

for some sequence of (finite) sets J_i and K_i. Note that $\Gamma^p \cup \Gamma^c$ is equivalent to Γ.

Let $\Gamma^{c^-} = \{\bigvee_{n\in J_i - I} \neg p_n \vee \bigvee_{n\in K_i} p_n | i \in \omega\}$. $\Gamma^p \cup \Gamma^{c^-}$ is equivalent to $\Gamma^p \cup \Gamma^c$ and therefore to Γ (for $q \in J_i \cap I$, $\{p_q, \bigvee_{n\in J_i} \neg p_n \vee \bigvee_{n\in K_i} p_n\}$ is equivalent to $\{p_q, \bigvee_{n\in J_i - \{q\}} \neg p_n \vee \bigvee_{n\in K_i} p_n\}$).

Let $\Gamma^{c^*} = \{\bigvee_{n\in J_i - I} \neg p_n \vee \bigvee_{n\in K_i} p_n | i \in \omega \text{ and } J_i - I \neq \emptyset\}$. $\Gamma^p \cup \Gamma^{c^*}$ is a subconditional theory, so now it suffices to show that $\Gamma^p \cup \Gamma^{c^*}$ is equivalent to $\Gamma^p \cup \Gamma^c$. $\Gamma^{c^*} \subseteq \Gamma^{c^-}$, so $\forall \mathcal{N}(\mathcal{N} \models \Gamma^p \cup \Gamma^{c^-} \Rightarrow \mathcal{N} \models \Gamma^p \cup \Gamma^{c^*})$. Let $\phi \in \Gamma^{c^-} - \Gamma^{c^*}$. Then $\phi = \bigvee_{n\in K_i} p_n$ for some i. Now, since $\Gamma^p \cup \Gamma^{c^-}$ is equivalent to Γ, $\mathcal{M}$ is a model of ϕ, so $K_i \cap I \neq \emptyset$, and therefore ϕ contains a (positive) literal already in Γ^p. Hence $\Gamma^p \cup \Gamma^{c^*} \models \phi$. Therefore, $\forall \phi \in \Gamma^{c^-}(\Gamma^p \cup \Gamma^{c^*} \models \phi)$, so $\Gamma^p \cup \Gamma^{c^*}$ is equivalent to $\Gamma^p \cup \Gamma^{c^-}$.

($\Leftarrow$) Suppose Γ is equivalent to some subconditional theory Γ^*. Let $\mathcal{M} = \{p_i | i \in I\}$. Clearly, $\mathcal{M}$ satisfies every positive sentence in Γ^*. Let $\phi = \bigvee_{n\in J_i} \neg p_n \vee \bigvee_{n\in K_i} p_n \in \Gamma^*$. By the definition of subconditional theories, ϕ contains a negative literal $\neg p_i$ such that $p_i \notin \mathcal{M}$, so $\mathcal{M} \models \phi$. $\mathcal{M}$ is minimal since for any $\mathcal{N} \subseteq \mathcal{M}$ there is a $p_i \in \Gamma^*$ such that $\mathcal{N} \not\models p_i$. □

Corollary 4.1 (Preservation theorem for subconditional theories) *A theory Γ is equivalent to a subconditional theory $\iff$ for every set of models of Γ, their intersection contains a model of Γ.*

Proof: Immediate, by propositions 4.1 and 4.2. □

The class of theories that are expressible as subconditional theories is the largest class of theories that can be completed via negation as failure. This class properly contains any theory equivalent to a Horn theory. In section 5.2 we will compute the complexity of determining whether a theory is minimally complete (and, hence, the complexity of determining whether it is equivalent to a subconditional theory).

5 Complexity

So far in this chapter we have encountered various properties of minimal consequence for sentential logic that indicate that minimal consequence cannot be formalised in the usual sense. In Section 3 we showed that, although minimal satisfiability is compact, minimal consequence is not, thus settling negatively the question of the existence of a finatary logical calculus for minimal consequence. Minimally complete theories appear to have a relatively simple set of minimal consequences and in the previous section we offered a complete characterisation for minimally complete theories in terms of subconditional theories. It is thus interesting to ask: How difficult is it to determine whether a theory is equivalent to a subconditional theory? The method suggested by the proof of proposition 4.2 would be to put the theory in conjunctive normal form and then check to see if it meets the conditions — obviously not an easy task. One may then wonder whether there is an alternative characterisation of minimally complete theories that would simplify this task, or whether the problem of determining whether or not a theory is minimally complete is intrinsically difficult.

It is the aim of this section to make the above questions and observations precise by means of complexity considerations, *i.e.*, by determining whether the relations or properties involved are recursive, recursively enumerable or of higher complexity. Note that in this, sentences of a countable theory can be viewed as natural numbers and theories can be viewed as sets of natural numbers or functions from natural numbers into $\{0,1\}$ (notation: ${}^{\omega}2$). Thus, $\models_s$ is studied here as a relation on ${}^{\omega}2 \times \omega$.

The arithmetical hierarchy classifies sets (relations) according to the quantifier complexity of their syntactic definition:

Definition 5.1 (The Arithmetical Hierarchy).

(i) A set B is *in* Σ^0_0 (Π^0_0) iff B is recursive.

(ii) For $n \geq 1$, B is *in* Σ^0_n (written $B \in \Sigma^0_n$) if there is a recursive relation $R(x, y_1, \ldots, y_n)$ such that

$$x \in B \iff \exists y_1 \forall y_2 \exists y_3 \cdots Q y_n R(x, y_1, \ldots, y_n)$$

where Q is $\exists$ if n is odd, and $\forall$ if n is even. B is *in* Π^0_n (written $B \in \Pi^0_n$) if

$$x \in B \iff \forall y_1 \exists y_2 \forall y_3 \cdots Q y_n R(x, y_1, \ldots, y_n)$$

where Q is $\exists$ if n is even, and $\forall$ if n is odd.

(iii) B is *in* Δ^0_n if $B \in \Sigma^0_n \cap \Pi^0_n$.

(iv) B is *arithmetical* if $B \in \bigcup_{n \in \omega}(\Sigma^0_n \cup \Pi^0_n)$.

The class of arithmetical relations is thus the smallest class of relations containing the recursive relations and closed under number quantification. Σ^0_n relations are relations definable by a formula with n alternations of quantifiers, beginning with an existential quantifier, and similarly, Π^0_n relations are definable by a formula with n alternations of quantifiers, beginning with a universal quantifier. Note that $\Delta^0_0 = \Delta^0_1 = \Sigma^0_0 = \Pi^0_0 =$ the recursive relations; $\Sigma^0_1 =$ the recursively enumerable (r.e.) relations; and $\Pi^0_1 =$ the complements of recursively enumerable (co-r.e.) relations. For relations A and B we say that A is *many-one reducible (m-reducible)* to B iff there exists a recursive function f such that $x \in A$ iff $f(x) \in B$. A relation is called Σ^0_n-*hard* iff all Σ^0_n relations on ω are m-reducible to it; a relation is called Σ^0_n-*complete* iff it is in Σ^0_n and all Σ^0_n relations on ω are reducible to it (*i.e.*, B is Σ^0_n-complete iff $B \in \Sigma^0_n$ and B is Σ^0_n-hard). Similarly for Π^0_n-*hard* and Π^0_n-*complete*. (Note that the notions of "complete" and "hard" are only defined for relations on natural numbers and that if a relation is "hard" or "complete" at some level of the arithmetical hierarchy, then it is not included in any level lower than that.)

The *analytical hierarchy* is defined in a parallel manner, with quantifiers ranging over subsets of the natural numbers (as opposed to natural numbers). The levels of the analytical hierarchy are similarly denoted Δ^1_n, Σ^1_n, and Π^1_n; these again correspond to relations definable by formulas with n alternations of *set* quantifiers.

For a countable theory Γ and a sentence ϕ in the language of Γ, we will consider the complexity of the following questions:

1. $\Gamma \models_s \phi$?

2. Is Γ minimally complete?

3. $\Gamma \models_s \phi$? for Γ minimally complete

The equivalent questions for logical consequence (in classical sentential logic) are: (1) r.e.; (2) Π^0_2; and (3) recursive. For minimal consequence, as we will see, the complexity of these questions will be: (1) Π^0_2 (and not Σ^0_2); (2) Π^0_2 (and not Σ^0_2); and (3) Δ^0_2 (and neither Σ^0_1 nor Π^0_1).

5.1 Complexity of the Minimal Consequence Relation

The failure of compactness for minimal consequence indicates that the minimal consequences of a theory Γ may not be recursively enumerable in Γ. The aim of this section is to show that, in fact, the minimal consequence relation is Π^0_2 and not Σ^0_2 . Indeed, we will exhibit a recursive theory Γ such that $\{\phi | \Gamma \models_s \phi\}$ is Π^0_2-complete.

From the definition of minimal consequence we can obtain the first, very loose, upper bound; recall that

$$\Gamma \models_s \phi \iff \forall \mathcal{M}(\mathcal{M} \models_s \Gamma \Rightarrow \mathcal{M} \models \phi)$$

and

$$\mathcal{M} \models_s \Gamma \iff \mathcal{M} \models \Gamma \text{ and } \forall \mathcal{N}(\mathcal{N} \models \Gamma \Rightarrow \mathcal{N} \not\subset \mathcal{M}).$$

Thus, $\Gamma \models_s \phi$ can be expressed by the following formula:

$$\forall \mathcal{M} \exists \mathcal{N}(\mathcal{M} \not\models \Gamma \text{ or} (\mathcal{N} \subset \mathcal{M} \text{ and } \mathcal{N} \models \Gamma) \text{ or } \mathcal{M} \models \phi)$$

The structures in sentential logic for a countable language are countable, so the quantifiers of this formula can be taken to range over subsets of ω; $\mathcal{M} \models \Gamma$ and $\mathcal{N} \subset \mathcal{M}$ are arithmetical relations; therefore it follows that $\models_s$ is a Π^1_2 relation, *i.e.*, is definable by an $\forall\exists$ formula with the quantifiers ranging over sets. The strict Π^0_2 upper bound will be obtained by showing that $\models_s$ is definable by an $\forall\exists$ formula with the quantifiers ranging over numbers, in proposition 5.2. Lower bounds will be established via many-one reduction of sets of known complexity; proposition 5.1 gives the strict lower bound: Π^0_2 .

As a prelude to this, first note that it is easy to modify the non-compactness example from proposition 3.2 and obtain a many-one reduction of $\overline{K}$, the set of indices of functions that diverge on the diagonal. $\overline{K}$ is a Π^0_1-complete set. From this it follows that the minimal consequence relation is not r.e.. Indeed, lemma 5.2 employs an even simpler construction to this end, namely that of a theory of the form: $\{p \vee q_1,\, p \vee q_2,\, \ldots\} \cup \{q_n |\, n \in I\}$, for some set $I \subseteq \omega$. It is easy to see that, if $I = \omega$, then all the disjuncts of the above theory are subsumed by the second part containing isolated literals and, therefore, p will be false in all its minimal models. A generalization of this observation is the content of the following lemma:

Lemma 5.1 *Let* $\Gamma = \Gamma^1 \cup \Gamma^2 \cup \Gamma^3$, *where* $\Gamma^1 = \{p \vee q_n |\, n \in I\}$, $\Gamma^2 = \{q_n |\, n \in J\}$, *and* Γ^3 *is a consistent theory which does not involve* p *or* q_i, *for any* $i \in I \cup J$.

Then $\Gamma \models_s \neg p \iff I \subseteq J$.

Proof: Since Γ^3 does not involve any of the same letters as Γ^1 and Γ^2 we have that

$$\{\mathcal{M} |\, \mathcal{M} \models_s \Gamma\} \;=\; \{\mathcal{M} \cup \mathcal{N} |\, \mathcal{M} \models_s \Gamma^1 \cup \Gamma^2 \text{ and } \mathcal{N} \models_s \Gamma^3\} \qquad (3.1)$$

We also have that

$$\mathcal{M} \models \Gamma^1 \cup \Gamma^2 \iff (p \in \mathcal{M} \text{ or } \forall i \in I\, q_i \in \mathcal{M}) \text{ and } \forall i \in J\, q_i \in \mathcal{M}$$

and thus

$$\begin{aligned} \mathcal{M} \models_s \Gamma^1 \cup \Gamma^2 \quad \iff \quad \mathcal{M} \;&=\; \{q_i |\, i \in I \cup J\} \text{ or} \\ &(\mathcal{M} = \{p\} \cup \{q_i |\, i \in J\} \text{ and } I \not\subseteq J). \end{aligned}$$

From (3.1) it now follows that

$$\Gamma \models_s \neg p \iff I \subseteq J.$$

□

The notation $[p]^n$ will be used to denote a conjunction of p, n times: $[p]^1 = p$ and $[p]^{n+1} = ([p]^n \wedge p)$; $\langle n, m \rangle$ will be used to denote an encoding of the ordered pair (n, m) ($\langle\, , \rangle$ is a recursive bijection between $\omega \times \omega$ and ω). We assume a fixed enumeration of the partial recursive functions. Thus φ_n will denote the n^{th} partial recursive function. For a given partial recursive function φ_n we say that $\varphi_n(x)$ *converges* iff φ_n is defined for input x and write $\varphi_n(x) \downarrow$; otherwise we say that $\varphi_n(x)$ *diverges* and write $\varphi_n(x) \uparrow$. Recall that $Cn_s(\Gamma)$ denotes $\{\phi | \, \Gamma \models_s \phi\}$.

Lemma 5.2 *Minimal consequence is not r.e., i.e., there is a recursive set of sentences, Γ, such that $Cn_s(\Gamma)$ is Π^0_1-hard.*

Proof: We show that minimal consequence is not r.e. by exhibiting a recursive set of sentences Γ such that

$$\forall n(\Gamma \models_s \neg p_n \iff n \in \overline{K}). \tag{3.2}$$

where

$$K = \{e | \, \varphi_e(e) \downarrow\}$$

i.e., the set of indices of partial recursive functions that are defined on the diagonal. K is a Σ^0_1-complete set (r.e.) and thus $\overline{K}$ (its complement) is Π^0_1-complete, so from (3.2) we obtain that $\models_s$ is not r.e..

Let $\{p_n | n \in \omega\} \cup \{q_{\langle n,x \rangle} | \, n, x \in \omega\}$ be a set of pairwise distinct sentence letters. To show (3.2), let $\Gamma = \bigcup_{n \in \omega} \Gamma_n$, where

$$\Gamma_n \quad = \quad \{p_n \vee q_{\langle n,x \rangle} | \, n, x \in \omega\} \cup \{q_{\langle n,x \rangle} | \, \neg T(n, n, x), \, i \in \omega\}$$

$$T(e, n, m) \iff \text{the } e^{th} \text{ Turing machine with input } n \text{ has converged in } m \text{ steps}$$

(T is the recursive Kleene T-predicate)

Γ is a recursive set of sentences. By lemma 5.1 and the definitions of T and K we have that:

$$\Gamma \models_s \neg p_n \iff \{i | \, \neg T(n, n, i), \, i \in \omega\} = \omega \iff \varphi_n(n) \uparrow \iff n \in \overline{K}$$

□

Proposition 5.1 *There is a recursive set of sentences* Γ *such that* $Cn_s(\Gamma)$ *is* Π^0_2-*hard*

Proof: Let $X = \{n \mid \forall x \exists y R(n, x, y)\}$, R a recursive relation, be a complete Π^0_2 set of numbers. We show how to construct a recursive set of sentences Γ such that X is m-reducible to $Cn_s(\Gamma)$.

Let $\{p_n | n \in \omega\} \cup \{q_{\langle n,x\rangle} | n, x \in \omega\}$ be a set of pairwise distinct sentence letters. Let $\Gamma = \{p_n \vee q_{\langle n,x\rangle} | n, x \in \omega\} \cup \{[q_{\langle n,x\rangle}]^y | R(n, x, y)\}$. Note that, by lemma 5.1, (i) for each $n, x \in \omega$, $\Gamma \models_s q_{\langle n,x\rangle}$ iff $\exists y R(n, x, y)$, and (ii) $\Gamma \models_s \neg p_n$ iff for each x, $\Gamma \models q_{\langle n,x\rangle}$. Hence, $n \in X \iff \Gamma \models_s \neg p_n$. □

Corollary 5.1 $\models_s$ *is not a* Σ^0_2 *relation on* ${}^\omega 2 \times \omega$.

Proof: The corollary follows directly from proposition 5.1. □

Proposition 5.1 and corollary 5.1 establish a lower bound on the complexity of minimal consequence. Intuitively, proposition 5.1 expresses the fact that deciding whether an arbitrary sentence is a minimal consequence of a given theory is at least as hard as deciding, given an index of a Turing machine, whether that Turing machine eventually halts at every input. The next proposition establishes that the lower bound on the complexity of the minimal consequence relation given in corollary 5.1 is the best possible.

Proposition 5.2 $\models_s$ *is a* Π^0_2 *relation on* ${}^\omega 2 \times \omega$.

Proof: The proof proceeds by showing that $\Gamma \not\models_s \phi$ is a Σ^0_2 relation, from which it follows that the relation "$\models_s$" is Π^0_2 . Note that $\Gamma \not\models_s \phi$ is equivalent to:

$$\exists \mathcal{M}(\mathcal{M} \models_s \Gamma \text{ and } \mathcal{M} \not\models \phi) \tag{3.3}$$

Let $Con(\Gamma) \iff \Gamma \not\models p \wedge \neg p$ (*i.e.*, Γ is a consistent theory). Con is a Π^0_1 predicate. It suffices then to show that (3.3) is equivalent to:

$$\exists \Delta(\Delta \text{ is finite } \& \ Con(\Gamma \cup \Delta) \ \& \ \neg Con(\Delta \cup \{\phi\}) \ \&$$
$$\forall i(p_i \in \Delta \Rightarrow \neg Con(\Gamma \cup \Delta^- \cup \{\neg p_i\})),) \tag{3.4}$$

where $\Delta^- = \{\neg\boldsymbol{p} \mid \neg\boldsymbol{p} \in \Delta\}$. (Note that $\forall i$ is a bounded quantifier and hence that (3.4) defines a Σ_2^0 relation on ${}^{\omega}2 \times \omega$.)

((3.3) $\Rightarrow$ (3.4)) Let $M = \{\boldsymbol{p} \mid \boldsymbol{p} \in \mathcal{M}\} \cup \{\neg\boldsymbol{p} \mid \boldsymbol{p} \notin \mathcal{M}\}$. Observe that

$$\mathcal{M} \not\models \phi \Rightarrow \exists \Delta \subseteq M(\Delta \text{ finite } \& \neg Con(\Delta \cup \{\phi\})) \tag{3.5}$$

and

$$\begin{aligned}
\mathcal{M} \models_s \Gamma \;&\Rightarrow\; \forall \Delta \subseteq M(Con(\Gamma \cup \Delta)\,\& \\
&\qquad \forall i(p_i \in \Delta \Rightarrow \neg Con(\Gamma \cup M^- \cup \{\neg p_i\}))) \\
&\Rightarrow\; \forall \Delta \subseteq M(Con(\Gamma \cup \Delta)\,\&\, \forall i(p_i \in \Delta \Rightarrow \\
&\qquad \exists \Sigma_{M,i} \subseteq M^-(\Sigma_{M,i} \text{ finite } \& \neg Con(\Gamma \cup \Sigma_{M,i} \cup \{\neg p_i\})))) \\
&\qquad\qquad\qquad \text{(by compactness)} \\
&\Rightarrow\; \forall \Delta \subseteq M(Con(\Gamma \cup \Delta)\,\&\, \forall i(p_i \in \Delta \Rightarrow \\
&\qquad \exists \Sigma_{M,i} \subseteq M^-(\Sigma_{M,i} \text{ finite } \& \\
&\qquad\qquad \neg Con(\Gamma \cup \bigcup_{\{i \mid p_i \in \Delta\}} \Sigma_{M,i} \cup \{\neg p_i\})))) \\
&\Rightarrow\; \forall \Delta \subseteq M\, \exists \Sigma \subseteq M^-(\Sigma \text{ finite } \& \, Con(\Gamma \cup \Delta)\,\& \\
&\qquad \forall i(p_i \in \Delta \Rightarrow \neg Con(\Gamma \cup \Sigma \cup \{\neg p_i\})))
\end{aligned} \tag{3.6}$$

From (3.5), (3.6) with (3.3), we obtain (3.4).

((3.4) $\Rightarrow$ (3.3)) Suppose (3.4) and let Δ be such that

$$\begin{aligned}
&\Delta \text{ is finite } \& \; Con(\Gamma \cup \Delta)\,\&\, \neg Con(\Delta \cup \{\phi\})\,\& \\
&\forall i(p_i \in \Delta \Rightarrow \neg Con(\Gamma \cup \Delta^- \cup \{\neg p_i\}))
\end{aligned} \tag{3.7}$$

Note that for all Δ'

$$\neg Con(\Delta' \cup \{\phi\}) \Rightarrow \forall \mathcal{M}(\mathcal{M} \models \Delta' \Rightarrow \mathcal{M} \not\models \phi) \tag{3.8}$$

and

$$Con(\Gamma \cup \Delta') \Rightarrow \exists \mathcal{M}(\mathcal{M} \models_s \Gamma \cup \Delta'). \tag{3.9}$$

From (3.7), (3.8), and (3.9) it follows that

$$\exists \mathcal{M}(\mathcal{M} \models_s \Gamma \cup \Delta \,\&\, \mathcal{M} \not\models \phi). \tag{3.10}$$

Now, it suffices to show that

$$\mathcal{M} \models_s \Gamma \cup \Delta \Rightarrow \mathcal{M} \models_s \Gamma, \tag{3.11}$$

since (3.3) follows immediately from (3.10) and (3.11).

Suppose $\mathcal{M} \models_s \Gamma \cup \Delta$ and $\exists \mathcal{N} \subset \mathcal{M}(\mathcal{N} \models_s \Gamma)$. From this we have that

$$\exists \mathcal{N} \subset \mathcal{M}(\mathcal{N} \models \Gamma \cup \Delta^- \,\&\, \mathcal{N} \not\models \Gamma \cup \Delta).$$ [1]

So

$$\exists i(p_i \in \Delta \,\&\, Con(\Gamma \cup \Delta^- \cup \{\neg p_i\}))$$

which contradicts (3.7). □

5.2 Complexity of the Minimal Completeness Property

In a computational setting it is necessary to have a way of determining whether, given an arbitrary theory, it is minimally complete, in a straightforward and mechanical fashion. Of course, one such method is suggested by the construction in the proof of proposition 4.2, which involves checking whether the theory is equivalent to a subconditional theory, but this requires that the entire theory be taken into consideration, even in the case where only one new sentence is being added to a minimally complete theory. Thus the characterisation of minimally complete theories in terms of subconditional theories seems lacking. In light of complexity considerations, however, we will see that the perceived defects of a characterisation based on subconditional theories stem from the intrinsic complexity of the problem of deciding minimal completeness.

Proposition 5.3 *The set MC of minimally complete theories is Π^0_2 and not Σ^0_2.*

Proof: For any theory Γ we have that

$$\Gamma \in MC \iff \forall \phi(\Gamma \models_s \phi \text{ or } \Gamma \models_s \neg\phi). \tag{3.12}$$

It follows from 3.12 and proposition 5.2 that MC is Π^0_2 and it follows from (the proof of) proposition 5.1 that MC is not Σ^0_2. □

[1] The first conjunct follows from $\mathcal{N} \models_s \Gamma$, $\mathcal{N} \subset \mathcal{M}$, and $\mathcal{M} \models_s \Gamma \cup \Delta$; the second follows from $\mathcal{M} \models_s \Gamma \cup \Delta$ and $\mathcal{N} \subset \mathcal{M}$.

In order to assess the usefulness of minimal completeness it is necessary to consider a further question: what is the complexity of minimal consequence for minimally complete theories? Consider the case of a pure conditional theory or a recursive set of sentences which is known to be minimally complete; does this knowledge affect the complexity of deciding minimal consequence, which as we have seen is Π^0_2 and not Σ^0_2? The answer to this question is negative, as one might suspect by inspecting the proof of proposition 4.2; notice that the construction in that proof assumes that the theory Γ has a unique minimal model, corresponding exactly to the atomic consequences of Γ, but the difficulty in generating the model from an arbitrary minimally complete set of sentences is apparent.

Proposition 5.4 *For minimally complete theories, minimal consequence is* Δ^0_2 *and neither* Σ^0_1 *or* Π^0_1.

Proof: In a minimally complete theory Γ

$$\forall\phi(\Gamma \models_s \phi \iff \Gamma \not\models_s \neg\phi) \tag{3.13}$$

So minimal consequence is Δ^0_2 for minimally complete theories. Suppose it is either Π^0_1 or Σ^0_1. From (3.13) we obtain that it is in fact both, so we have that it is Δ^0_1 . Let $\Gamma = \{[p_e]^k | \, T(e,e,k)\}$. Note that Γ is a recursive set of sentences and it is minimally complete. We now have that

$$\Gamma \models_s p_e \iff e \in K$$

and thus a decision procedure for K, which is a contradiction. □

6 Discussion

Having motivated the study of minimally complete theories by computational considerations, it is natural to ask what bearing these results have on computational problems, since they were generally obtained in connection to infinite theories. As was discussed in Chapter 2, in semantics of logic programming, one is often led to consider the set of ground instances of a (finite) database; in general, this is an infinite propositional theory. Hence, these results suggest interesting extensions of logic programming. Second, although no clear connection has been made to date, there is much evidence that the complexity of a problem restricted to finite objects is generally connected to that of the unrestricted case (see, *e.g.*, [Stockmeyer 87]); for example, a very high complexity for the problem of deciding minimal consequence between a theory and a sentence suggests a high degree of intractability for the problem of deciding minimal consequence between two sentences.

Chapter 4

Minimal Consequence in First Order Logic

1 Introduction

This chapter presents a study of minimal consequence in first order logic. A detailed development of the theoretical aspects of minimal consequence in first order logic provides a firm basis for the various circumscription formalisms that are of interest to researchers in AI. The study of minimal consequence in sentential logic has prepared the ground for this elaboration, by exposing many pertinent properties of minimal consequence for a very simple language. Many such properties extend naturally to first order languages, but here we must expect that with the added richness of expression there will be an ensuing increase in complexity, all around. To begin with, many different notions of minimality appear very natural and equally worthy of investigation. We will restrict attention to three of these, which underlie McCarthy's domain, predicate, and formula circumscription. We will begin by considering domain minimal consequence both for historical and motivational reasons.

Historically, domain circumscription was the first concept of circumscription introduced [McCarthy 77], and intuitively, it is the clearest, as it relies on familiar model theoretic notions (*e.g.*, that of a submodel). In order to define the relation of *domain minimal consequence*, to be written $\models_d$, (and, subsequently, of *predicate minimal consequence* and *formula minimal consequence*, to be written $\models_p^{(\mathbf{R})}$ and $\models_f^{(\mathbf{R},\mathbf{U})}$ respectively), we will need some model theoretic notions. All such notions assumed here are defined in a standard way, as in [Chang & Keisler 73] — Chapter 3 provides the necessary background to the key definitions and notation used here.

Sections 2, 3, and 4 provide definitions for domain, predicate, and formula minimal consequence, respectively, by analogy with the usual definition of logical consequence in first order logic. The definitions are followed by a number of examples that, in each case, illustrate the more immediate properties of the relation defined. In contrast with the familiar notion of logical consequence, which is resilient to expansions of the non-logical vocabulary, minimal consequence is very sensitive to such changes; for example, a theory that has no minimal models in one language may have minimal models in another. The lack of minimal models for satisfiable theories has been a matter of concern in the circumscription literature (for example, see [Davis 80,Etherington 86]) because it lights a danger signal for circumscription-based computational systems: any system that uses circumscription to make "common sense" conjectures, runs the risk of introducing inconsistencies into its database, in so doing. Sections 2, 3, and 4 conclude with results on this topic, showing that any satisfiable theory can be extended to a theory (in an extended language) that has minimal models and makes true the same sentences of the original language as the original theory. Moreover, if the language is finite, this can be accomplished with the addition of finitely many new symbols to the language and finitely many sentences to the theory; in fact, in the case of domain minimal consequence, the extension to the language suffices — the original theory in the extended language will have minimal models. Next, in section 5, we examine the relationship between domain, predicate, and formula minimal consequence and present results on the exact connections among these. Section 6 will review and extend the investigation of model theoretic properties of minimal consequence, with particular emphasis on properties underlying practical applications.

As was the case with most results in the previous chapter, the results in this chapter are obtained restricted to finite or countable first order languages — in this case, languages consisting of a finite number of relation, function, and constant symbols — since uncountable languages do not appear to be of much practical use. In most cases, however, this restriction is not necessary and the results are easily extended to arbitrary first order languages.

2 Domain Minimal Consequence

2.1 Preliminaries

Domain minimal consequence (or *d-minimal consequence*) aims to capture an intuitive notion that "the objects required by a certain theory are all there are" (in this sense it "circumscribes" the domains of the models of a theory). Domain minimal consequence was first introduced by McCarthy in [McCarthy 80], where it is called *minimal entailment*, as the semantic counterpart of domain circumscription (discussed in Chapter 2).

Definition 2.1 Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathcal{L}$-structures. $\mathcal{M} \subseteq_d \mathcal{N}$ ($\mathcal{M}$ is a *d-submodel* of $\mathcal{N}$) iff

(i) $M \subseteq N$;

(ii) $R^{\mathcal{M}}$ is the restriction of $R^{\mathcal{N}}$ to M, for each relation symbol $R \in \mathcal{L}$;

(iii) $f^{\mathcal{M}}$ is the restriction of $f^{\mathcal{N}}$ to M, for each function symbol $f \in \mathcal{L}$ and $c^{\mathcal{M}} = c^{\mathcal{N}}$ for each constant symbol $c \in \mathcal{L}$.

Note that the definition of a d-submodel is the same as the standard definition of a submodel (or substructure) in model theory. It is included here for completeness and so that it can be compared and contrasted with the definitions of p-submodel and f-submodel which are to follow.

Definition 2.2 Let $\mathcal{M}$ be an $\mathcal{L}$-structure and Γ an $\mathcal{L}$-theory. $\mathcal{M} \models_d \Gamma$ ($\mathcal{M}$ is a *d-minimal model* of Γ) iff $\mathcal{M} \models \Gamma$ and for every $\mathcal{N}$ such that $\mathcal{N} \models \Gamma$, if $\mathcal{N} \subseteq_d \mathcal{M}$, then $\mathcal{N} = \mathcal{M}$.

A d-minimal model is thus a model with no proper d-submodels.

Definition 2.3 Let Γ be an $\mathcal{L}$-theory and ϕ an $\mathcal{L}$-sentence. $\Gamma \models_d \phi$ (ϕ is a *d-minimal consequence* of Γ) iff ϕ is true in every d-minimal model of Γ.

Example 2.1 Let $\Gamma = \emptyset$ (the empty theory), in a language that contains no constant or function symbols, but contains at least one one-place predicate symbol, P. The d-minimal models of Γ are those with singleton domains. Since all such models make true sentences of the form $\exists x P(x) \rightarrow \forall x P(x)$, for any predicate symbol P in the language, we have that $\Gamma \models_d \exists x P(x) \rightarrow \forall x P(x)$.

This example illustrates the non-monotonicity of d-minimal consequence. Let $\Delta = \{\exists x P(x), \exists x \neg P(x)\}$. Clearly, $\Gamma \subset \Delta$, but $\Delta \not\models_d \exists x P(x) \rightarrow \forall x P(x)$, since the minimal models of Δ have exactly two elements.

Notice that the phenomenon illustrated by example 2.1, namely the non-monotonicity of d-minimal consequence, results from the following feature of d-minimal consequence. A sentence is a d-minimal consequence of a theory if it is true in all the *d-minimal models* of the theory. Whereas an extension Γ' of a theory Γ always has at most the models that Γ has, it may have d-minimal models which Γ lacks (of course, Γ may also have d-minimal models which Γ' lacks). This is what accounts for the non-monotonicity of notions of minimal consequence in general.

Example 2.2 Let $\Gamma = \{P(a), P(b)\}$. The d-minimal models of Γ are one or two element structures in which the extension of P contains the whole domain. So we have, for example, $\Gamma \models_d \forall x P(x)$ and $\Gamma \models_d \forall x(x = a \vee x = b)$.

This example serves to clarify what is meant by circumscribing the domain of a theory, as defined here: d-minimal consequence neither forces different interpretations for the constant symbols nor does it force the same interpretation, as we admit as d-minimal models both those where elements have unique names and those where they do not. Thus there can be no connection between d-minimal consequence and the unique names assumption.

Example 2.3 [Davis 80] Let

$$\Gamma = \{\exists x \forall y \neg s(y) = x,\ \forall x \forall y (s(x) = s(y) \rightarrow x = y)\}.$$

Here Γ is intended as a theory of the natural numbers with the usual successor function. Any model of Γ contains an infinite chain of elements generated by the function s, isomorphic to the natural numbers (usually referred to as a standard chain), and possibly other chains unbounded in one or both directions (since the theory only requires one element without predecessor) or finite cycles. Thus, every model of Γ contains a proper d-submodel isomorphic to the natural numbers, obtained by dropping all but one standard chain and one or more elements from the origin of the standard chain. Since any such d-submodel satisfies Γ, it follows immediately from the definition that Γ has no d-minimal model.

Observe that from this it follows that Γ d-minimally entails every sentence, since by the definition it is vacuously true that every sentence holds in all d-minimal models of Γ, as Γ has no d-minimal models. The following definition is natural in this connection:

Definition 2.4 Γ is *d-satisfiable* iff Γ has a d-minimal model.

Γ as defined in example 2.3 is thus a satisfiable first order theory which is not d-satisfiable.

2.2 Existence of Domain Minimal Models

The existence of satisfiable theories with no minimal models has been of some concern in the circumscription literature (for example, see [Davis 80,Etherington 86]). For the purposes of modelling common sense reasoning, it is desirable to have a computational system that jumps to conclusions which are always consistent with its database (given that the database was originally consistent). Note that although the concern is valid, it does not suggest there is anything wrong with the concept of minimal consequence nor with circumscription formalisms. Rather, it indicates that caution must be exercised in employing circumscription in a computational system to generate plausible conjectures. Nevertheless, we feel that, given the computational barriers posed by consistency checking, this is indeed a real problem, to be addressed at the theoretical level, as opposed to the applications level.

In brief, there are two ways in which to dispense with the difficulty of satisfiable theories that are not minimally satisfiable. One is to redefine the submodel relation and the notion of minimality so that some (or all) models of a theory have a minimal submodel — if this is possible. The other is to identify large classes of theories where this difficulty does not arise and to argue that these cover all cases (theories) that would be of interest in our application, except perhaps for a small number of contrived examples. Efforts in the first direction could, for example, propose to redefine a minimal submodel to be a model of the theory with no proper non-isomorphic submodels; under this definition the theory of example 2.3 would be d-satisfiable, but it is still possible to construct theories that are not (see Chapter 6). Efforts in the second direction will typically formulate a syntactic restriction that is judged to be general enough to include all the theories that are of interest in applications or, at least, a logically equivalent theory for each of these (this type of approach is taken in Chapter 3, Section 4, where we define subconditional theories).

This thesis explores the latter approach. This section addresses the problem by showing that, although there are satisfiable theories that are not d-satisfiable, every satisfiable theory is d-satisfiable in a finitely extended language. Clearly, this solves the problem and more: in classical first order logic extending the language of a theory cannot possibly alter the set of its consequences (in the original language), thus it can do no harm to suppose there are some extra symbols in the language that are not used in the theory and, therefore, every satisfiable theory (not only those of interest in practical applications) is d-satisfiable. More importantly, this can be accomplished with the addition of at most finitely many new symbols.

Example 2.3 suggests the possibility that theories which are not d-satisfiable may have d-satisfiable extensions. This is in fact the case, as shown by the theory $\Gamma' = \Gamma \cup \{\forall x \neg s(x) = 0\}$. Note that, up to isomorphism, Γ' has as its unique d-minimal model the natural numbers with the usual successor function. This is due to the fact that some of the expansions of the models of the theory in the old language to the new language will become d-minimal. In the example above, intuitively what happens is that the constant 0 somehow "grounds" the infinite chain of submodels; 0 can be interpreted as any of the elements of the domain of a model of Γ' (since there are no axioms governing its interpretation) but if it is interpreted as the first element of an infinite chain of successors (*i.e.*, its usual interpretation), then the model will have no proper d-submodel, since a d-submodel must give the same interpretation to 0. Thus, if a model interprets 0 and s in such a way that 0 is the first element of an infinite chain of terms defined by the interpretation of s, that model will be d-minimal. Thus Γ' is a d-satisfiable extension of a theory which is not d-satisfiable.

It is interesting to note that all d-satisfiable extensions of Γ use new vocabulary. Observe that every model for a satisfiable extension of Γ contains a standard chain, thus any such extension within the language $\{s\}$ will have no minimal model (by the same argument that Γ is not d-satisfiable). Moreover, the addition of the sentence $\{\forall x \neg s(x) = 0\}$ to Γ is not necessary in order to make Γ d-satisfiable and it serves only as a means of introducing the constant 0 to the language; it would suffice simply to consider Γ to be a theory of the language $\{s, 0\}$, instead of $\{s\}$. Indeed, we see that the addition of a single constant symbol to the language of Γ, not only renders it d-satisfiable, but also forces all its d-minimal models to be isomorphic to the d-minimal models of Γ', so that Γ and Γ' in the language $\{s, 0\}$ have the same d-minimal consequences.

The fact that the same theory can have different d-minimal consequences, depending on the language of which it is considered to be a theory, represents a radical departure from familiar first order model theory. In the case of logical consequence we can leave the language of a theory unspecified and it can safely be assumed to be the set of symbols occurring in sentences of the theory, since they must be in the language; it makes no difference to the consequences of the theory in this restricted language, if there are additional (unmentioned) symbols in the language. For d-minimal consequence, on the other hand, we see that the language of a theory plays a very important role. We will continue to define it implicitly, as the set of symbols occurring in an axiomatization of the theory, but note now that accordance to this rule is significant.

It is also useful to pause briefly to consider some of the particular effects of the choice of language on the d-minimal consequences of a theory in that language. As noted above, the addition of a new constant symbol to the language can turn a theory that is not d-satisfiable into a d-satisfiable one. This is due to the fact that some of the expansions of the models of the theory in the old language to the new language will become d-minimal and it suggests the significance of the availability of names in the language for elements of the domain of its models. The following definition will be useful:

Definition 2.5 Let $\mathcal{M}$ be an $\mathcal{L}$-structure. $\mathcal{M}$ is called *nameable* iff for all $\mathbf{a} \in M$ there is a closed term $t \in \mathcal{L}$ such that $t^{\mathcal{M}} = \mathbf{a}$.

Nameable structures can be useful in shedding some light into the general question of existence of d-satisfiable extensions for satisfiable theories. For example, once all the elements of the domain of a model $\mathcal{M}$ are linked to a term (*i.e.*, they are "named"), if any of them is eliminated from the domain, the resulting structure will no longer be a submodel. Any $\mathcal{L}$-structure can be expanded to a nameable $\mathcal{L}'$-structure for an appropriate choice of $\mathcal{L}'$ (see lemma 2.1 below) and hence, any satisfiable $\mathcal{L}$-theory is a d-satisfiable $\mathcal{L}'$-theory for appropriate choice of $\mathcal{L}'$. Moreover it is possible to achieve this with a finite extension of the language. This is the content of proposition 2.1 below.

Lemma 2.1 *Any countable $\mathcal{L}$-structure $\mathcal{M}$ can be expanded to a nameable $\mathcal{L}'$-structure $\mathcal{M}'$ such that* *(i)* $\mathcal{L} \subseteq \mathcal{L}'$;

(ii) $\mathcal{L}' - \mathcal{L}$ *is finite; and*

(iii) $\mathcal{M}'$ *and* $\mathcal{M}$ *make true the same* $\mathcal{L}$*-sentences.*

Proof: For the case where the $\mathcal{L}$-structure $\mathcal{M}$ is finite, let $\mathcal{L}' = \mathcal{L} \cup \{c_1, \ldots, c_n\}$, where $n = |M|$ and $c_i \notin \mathcal{L}$, $i = 1, \ldots, n$. Now, let $\mathcal{M}' = \langle \mathcal{M}, \mathbf{a_1}, \ldots, \mathbf{a_n} \rangle$, where $\{\mathbf{a_1}, \ldots, \mathbf{a_n}\} = M$, *i.e.*, the constants $c_1, \ldots, c_n$ are interpreted by the elements of the domain of $\mathcal{M}$ (thus $\mathcal{M}'$ is nameable) and all other symbols receive the same interpretation in $\mathcal{M}'$ as in $\mathcal{M}$, and thus make true the same $\mathcal{L}$-sentences.

If $\mathcal{M}$ is countable, *i.e.*, if $M = \{\mathbf{a_i} | \mathbf{i} \in \omega\}$, where $\{\mathbf{a_i} | \mathbf{i} \in \omega\}$ is an enumeration of M without repetitions, let $\mathcal{L}' = \mathcal{L} \cup \{c, f\}$, $f \notin \mathcal{L}$, $c \notin \mathcal{L}$. Let $\mathcal{M}' = \langle \mathcal{M}, \mathbf{a_0}, \{< \mathbf{a_i}, \mathbf{a_{i+1}} > | \mathbf{i} \in \omega\} \rangle$, *i.e.*, c is interpreted by $\mathbf{a_0}$ and f by the successor function on the elements of the domain of $\mathcal{M}$ (relative to the ordering imposed by their enumeration $\{\mathbf{a_i} | \mathbf{i} \in \omega\}$), and all other symbols receive the same interpretation in $\mathcal{M}'$ as in $\mathcal{M}$, and thus make true the same $\mathcal{L}$-sentences. For every $\mathbf{a_i}$, $\mathbf{a_i} = f(f(\cdots f(c) \cdots))^{\mathcal{M}'}$ (i applications of f) and thus $\mathcal{M}'$ is nameable. □

Lemma 2.2 *If* $\mathcal{M}$ *is a nameable* $\mathcal{L}$*-structure then* $\mathcal{M}$ *is a d-minimal model of every* $\mathcal{L}$*-theory of which it is a model.*

Proof: Suppose $\mathcal{M}$ is a nameable $\mathcal{L}$-structure. Then $\mathcal{M}$ has no proper submodel, *i.e.*, if $\mathcal{M}' \subseteq_d \mathcal{M}$, then $\mathcal{M}' = \mathcal{M}$, since for any $\mathbf{a} \in M$ we have:

$$\begin{aligned} \mathbf{a} \in M &\Rightarrow \exists t \in \mathcal{L}\ \mathbf{a} = t^{\mathcal{M}} \text{ (since } \mathcal{M} \text{ is nameable)} \\ &\Rightarrow \exists t \in \mathcal{L}\ \mathbf{a} = t^{\mathcal{M}'} \text{ (since } \mathcal{M}' \subseteq_d \mathcal{M}) \\ &\Rightarrow \mathbf{a} \in M' \end{aligned}$$

But since $\mathcal{M}$ has no proper submodel, if $\mathcal{M} \models \Gamma$, then $\mathcal{M} \models_d \Gamma$. □

Proposition 2.1 *Any satisfiable* $\mathcal{L}$*-theory is a d-satisfiable* $\mathcal{L}'$*-theory, for some* $\mathcal{L}' \supseteq \mathcal{L}$*, where* $\mathcal{L}' - \mathcal{L}$ *is finite.*

Proof: Let Γ be a satisfiable $\mathcal{L}$-theory. By the downward Lowenheim-Skolem theorem[1] Γ has a countable model, $\mathcal{M}$. By lemma 2.1, $\mathcal{M}$, can be expanded to a nameable $\mathcal{L}'$-structure, $\mathcal{M}'$, by the addition of a finite number of new constant and function symbols to $\mathcal{L}$ and, moreover, $\mathcal{M}$ and $\mathcal{M}'$ make true the same sentences in $\mathcal{L}$. Thus, if $\mathcal{M} \models \Gamma$, then $\mathcal{M}' \models \Gamma$ and by lemma 2.2 $\mathcal{M}'$ will be d-minimal, so Γ is a d-satisfiable $\mathcal{L}'$-theory. □

As we will see in the following sections, similar results apply to other notions of minimality (predicate and formula); sections 3 and 4 also conclude with some results on this topic, showing that it is possible trivially to extend satisfiable theories with no minimal models to theories in an expanded language that have minimal models.

3 Predicate Minimal Consequence

3.1 Preliminaries

Predicate minimal consequence (or *p-minimal consequence*) aims to capture the notion that "the objects that can be shown to satisfy a certain property P by reasoning from certain facts A are all the objects that satisfy P." In this sense it "circumscribes" the extensions of one or more predicates *i.e.*, it minimizes the sets interpreting some of the relation symbols of a theory. Predicate minimal consequence was first introduced by McCarthy in [McCarthy 80], where it is called *minimal entailment*, as the semantic counterpart of predicate circumscription (discussed in Chapter 2).

Definition 3.1 Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathcal{L}$-structures and let $\mathbf{R}$ be a set of relation symbols in $\mathcal{L}$. $\mathcal{M} \subseteq_p^{(\mathbf{R})} \mathcal{N}$ ($\mathcal{M}$ is a *p-submodel* of $\mathcal{N}$ with respect to $\mathbf{R}$) iff

(i) $M = N$;

[1]See, *e.g.*, [Enderton 72], page 141.

(ii) $R^{\mathcal{M}} \subseteq R^{\mathcal{N}}$, for each relation symbol R;

(iii) $R^{\mathcal{M}} = R^{\mathcal{N}}$, for every relation symbol $R \notin \mathbf{R}$;

(iv) $f^{\mathcal{M}} = f^{\mathcal{N}}$ and $c^{\mathcal{M}} = c^{\mathcal{N}}$, for each function symbol f and constant symbol c.

We refer to the relation symbols in $\mathbf{R}$ as the *circumscribed predicates.* Note that a p-submodel of a model $\mathcal{M}$ is generally *not* a submodel of $\mathcal{M}$ (in the usual model theoretic sense). In fact, it follows directly from the definition of a p-submodel that if $\mathcal{M}$ is a submodel of $\mathcal{N}$ and $\mathcal{M} \subseteq_p^{(\mathbf{R})} \mathcal{N}$, then $\mathcal{M} = \mathcal{N}$.

The definitions of p-minimal model and p-minimal consequence are analogous to the respective ones for d-minimal model and d-minimal consequence:

Definition 3.2 $\mathcal{M} \models_p^{(\mathbf{R})} \Gamma$ ($\mathcal{M}$ is a *p-minimal model* of a theory Γ with respect to a set of predicates $\mathbf{R}$) iff $\mathcal{M} \models \Gamma$ and for every $\mathcal{N}$ such that $\mathcal{N} \models \Gamma$, if $\mathcal{N} \subseteq_p^{(\mathbf{R})} \mathcal{M}$, then $\mathcal{N} = \mathcal{M}$.

Definition 3.3 $\Gamma \models_p^{(\mathbf{R})} \phi$ (ϕ is a *p-minimal consequence* of Γ with respect to $\mathbf{R}$) iff ϕ is true in all p-minimal models of Γ.

Example 3.1 Let $\Gamma = \emptyset$ and let $\mathbf{R} = \{P\}$. The p-minimal models of Γ with respect to $\mathbf{R}$ are those that interpret the predicate symbol P by the empty set, *i.e.*, those where the extension of P is empty. All such models of Γ make true all sentences of the form $\neg P(t)$, where t is a term in the language. They also make true the sentence $\forall x \neg P(x)$. For other predicate symbols $Q \neq P$, there will be p-minimal models, $\mathcal{M}$, such that $\mathcal{M} \models \forall x \neg Q(x)$ and $\mathcal{N}$ such that $\mathcal{N} \models \forall x Q(x)$. Thus it follows that $\Gamma \models_p^{(\mathbf{R})} \forall x \neg P(x)$, but $\Gamma \not\models_p^{(\mathbf{R})} \forall x \neg Q(x)$.

It is easy to see from this example that p-minimal consequence is non-monotonic since, although $\Gamma \models_p^{(\mathbf{R})} \forall x \neg P(x)$, $\Gamma \cup \{P(a)\} \not\models_p^{(\mathbf{R})} \forall x \neg P(x)$. As with d-minimal consequence, this is due to the fact that an extension of a theory Γ may have p-minimal models which Γ lacks.

Example 3.2 Let $\Gamma = \{P(a), P(b), Q(a)\}$. The p-minimal models of Γ with respect to $\mathbf{R} = \{P\}$ are those models in which the interpretation of P contains only the interpretations of the constants a and b. Depending whether a and b are given distinct interpretations, the interpretation of P in those models will contain either one or two elements. There will be such p-minimal models $\mathcal{M}$ in which the extension of Q contains only $a^{\mathcal{M}}$ and ones where, in addition, it contains any number of other elements of M. Thus, while $\Gamma \models_p^{(\mathbf{R})} \forall x(P(x) \rightarrow x = a \vee x = b)$, $\Gamma \not\models_p^{(\mathbf{R})} \forall x(Q(x) \rightarrow x = a)$. In fact, it is easy to see that for any term t,

$$\Gamma \models_p^{(\mathbf{R})} Q(t) \iff \Gamma \models Q(t), \text{ and}$$
$$\Gamma \models_p^{(\mathbf{R})} \neg Q(t) \iff \Gamma \models \neg Q(t)$$

We also have that, for any ground term t, $\Gamma \models_p^{(\mathbf{R})} P(t) \iff \Gamma \models P(t)$, *i.e.*, p-minimal consequence does not produce any new positive ground instances of the predicate circumscribed.

3.2 Limitations of Predicate Minimal Consequence

In standard AI applications of common sense reasoning, it is generally felt that it should be possible to obtain new positive ground sentences, as in the typical bird example, where it is known that Tweety is a bird and typical birds can fly, and where it is argued that a common sense conclusion is that Tweety can fly. Many elaborations exist of this example, "axiomatized" in progressively more peculiar ways, so as to render it explicable by circumscription. One difficulty lies in obtaining a positive conclusion, namely that Tweety can fly. The fact that predicate circumscription (and, indeed, p-minimal consequence) does not produce any new positive ground instances of the predicate circumscribed, and any new positive or negative ground instances of other predicates, is not due to an idiosyncracy of the particular theory Γ of this example, but is characteristic of a certain class of theories which includes universal theories (and Γ happens to be such) [Etherington et al 85]. Nevertheless, this fact may appear surprising; at first glance it would seem that, in order to obtain new positive ground instances of a predicate P via p-minimal consequence, it would suffice to define P in the theory as the negation of another predicate Q and circumscribe Q — thus obtaining new negative ground instances of Q, from which the respective positive ground instances of P would then follow. Upon some reflection on the definition of p-minimal consequence, however, we see that this is not the case, as illustrated by the next example.

Example 3.3 Let $\Gamma = \{Q(a), \forall x(\neg Q(x) \rightarrow P(x)), \neg a = b, \forall x(x = a \vee x = b)\}$ and $\mathbf{R} = \{Q\}$. This theory has only models with two element domains. In addition, each of its models is isomorphic to one of the following models:

$\mathcal{M}_1 : Q^{\mathcal{M}_1} = \{\mathbf{a}\},\ P^{\mathcal{M}_1} = \{\mathbf{b}\};$

$\mathcal{M}_2 : Q^{\mathcal{M}_2} = \{\mathbf{a}\},\ P^{\mathcal{M}_2} = \{\mathbf{a}, \mathbf{b}\};$

$\mathcal{M}_3 : Q^{\mathcal{M}_3} = \{\mathbf{a}, \mathbf{b}\},\ P^{\mathcal{M}_3} = \emptyset;$

$\mathcal{M}_4 : Q^{\mathcal{M}_4} = \{\mathbf{a}, \mathbf{b}\},\ P^{\mathcal{M}_4} = \{\mathbf{a}\};$

$\mathcal{M}_5 : Q^{\mathcal{M}_5} = \{\mathbf{a}, \mathbf{b}\},\ P^{\mathcal{M}_5} = \{\mathbf{b}\};$

$\mathcal{M}_6 : Q^{\mathcal{M}_6} = \{\mathbf{a}, \mathbf{b}\},\ P^{\mathcal{M}_6} = \{\mathbf{a}, \mathbf{b}\};$

where $M_1 = M_2 = M_3 = M_4 = M_5 = M_6 = \{\mathbf{a}, \mathbf{b}\}$. Notice that $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ are all p-minimal, so it follows then that $\Gamma \not\models_p^{(\mathbf{R})} \neg Q(b)$ and $\Gamma \not\models_p^{(\mathbf{R})} P(b)$.

Thus, again we have that p-minimal consequence yields no new positive ground instances of the predicate circumscribed and no new positive or negative ground instances of any of the other predicates for this theory. Moreover it becomes clear that ingenuity is not going to help in getting around this problem; its source lies in the way p-minimal consequence is defined, namely under (iii) of the definition of p-submodel, which forces the extensions of predicates not being circumscribed to be the same in p-submodels. If a theory is *well founded*, *i.e.*, if all its models have a p-minimal submodel (for every set of circumscribed predicates), then it is clear that positive ground instances of the predicates circumscribed will be true in all the minimal models only if they are true in all the models of the theory; similarly for any ground instance — positive or negative — of predicates not being circumscribed [Etherington et al 85]. Most theories that are of interest to AI are well founded, including the theory of example 3.3, theories of abnormal birds, *etc.* Yet, from the point of view of intuitive motivation for p-minimal consequence, these are the cases where it would be most desirable to be able to minimize the extension of a predicate, even when the theory logically connects that predicate to another predicate. In essence this would involve allowing extensions of some predicates to vary (in either direction) during the minimization. This observation led to the development of a new minimal consequence relation, which is the subject of the next section.

Returning to our example, apart from the fact that in the situation it exemplifies no new positive ground instances of the predicate circumscribed could be obtained via p-minimal consequence, it is interesting to ask whether there are other types of theories where this is not the case. So the question is the following: is there a theory Γ such that $\Gamma \models_p^{(\mathbf{R})} P(a)$ and $\Gamma \not\models P(a)$, where $\mathbf{R} = \{P\}$? The next example gives a positive answer to this question.

Example 3.4 Let Γ be the theory of a (strict) linear ordering R and a non-empty upward closed property P, with a name for the top element, if this exists. For clarity, the definition of Γ is broken down into the definitions of a linear ordering ($\mathrm{lo}(R)$), non-empty upward closed property ($\mathrm{neupc}(R,P)$) and an axiom which states that if a top element exists then a is the top element ($\mathrm{top}(R) \Rightarrow \mathrm{top}(R,a)$).

$$\begin{aligned}
\mathrm{lo}(R) &= \{\forall x \forall y (R(x,y) \vee R(y,x) \vee x = y), \\
&\quad\ \forall x \forall y (R(x,y) \rightarrow \neg R(y,x)), \\
&\quad\ \forall x \forall y \forall z ((R(x,y) \wedge R(y,z)) \rightarrow R(x,z))\} \\
\mathrm{neupc}(R,P) &= \{\exists x P(x), \forall x \forall y ((R(x,y) \wedge P(x)) \rightarrow P(y))\} \\
\mathrm{top}(R) \Rightarrow \mathrm{top}(R,a) &= \{\forall y (\forall x \neg R(y,x) \rightarrow y = a)\} \\
\Gamma &= \mathrm{lo}(R) \cup \mathrm{neupc}(R,P) \cup \mathrm{top}(R) \Rightarrow \mathrm{top}(R,a)
\end{aligned}$$

A typical model of this theory is $\mathcal{M}_0 = \langle \mathrm{N}, <, \{\mathrm{n}|\mathrm{n} \geq \mathbf{42}\}, 0\rangle$, where N is the set of natural numbers, "$<$" interprets R and is the usual "less than" relation on the natural numbers, "$\{\mathrm{n}|\mathrm{n} \geq \mathbf{42}\}$" interprets P as the natural numbers greater than or equal to 42, and the constant a is interpreted by 0. In this model we clearly have $\neg P(a)$, so it follows that $\Gamma \not\models P(a)$.

Another model of Γ is $\mathcal{M}_1 = \langle [0,1], <, \{\mathrm{n}|\mathrm{n} \geq \frac{1}{2}\}, 1\rangle$, where $[0,1]$ is the closed interval from 0 to 1 on the real line. We see that $a^{\mathcal{M}_1}$ is the top element with respect to the relation $R^{\mathcal{M}_1}$ and $a^{\mathcal{M}_1} \in P^{\mathcal{M}_1}$, so it follows that $\Gamma \not\models \neg P(a)$.

Observe that all the models of Γ fall into two categories: those that have a top element (*e.g.*, $\mathcal{M}_1$) and those that are unbounded (*e.g.*, $\mathcal{M}_0$). Now consider what happens when we circumscribe the predicate P in this theory; those models which have a top element will have a p-minimal submodel $\mathcal{M}$, which also has a top element, where $P^{\mathcal{M}} = \{a^{\mathcal{M}}\}$; those models that are unbounded will have an infinite sequence of p-submodels, each of which interprets P by a proper subset of the previous model (for example, $\mathcal{M}_0$ has a p-submodel where P is interpreted by "$\{\mathrm{n}|\mathrm{n} \geq \mathbf{43}\}$," one where it is "$\{\mathrm{n}|\mathrm{n} \geq \mathbf{44}\}$," one where it is "$\{\mathrm{n}|\mathrm{n} \geq \mathbf{45}\}$," ... and so on) and so will have no p-minimal p-submodels. So the p-minimal models of Γ are the p-minimal p-submodels of the models with top elements and hence all p-minimal models of Γ have a top element. In other words, Γ p-minimally entails that the relation R has a top element. Moreover, since P is non-empty upward closed, it will be the case that in every model $\mathcal{M}$ with a top element $\{a^{\mathcal{M}}\} \subseteq P^{\mathcal{M}}$ and that in every p-minimal model (which will necessarily also have a top element) $\{a^{\mathcal{M}}\} = P^{\mathcal{M}}$. Thus we have that $\Gamma \models_p^{(\mathbf{R})} P(a)$, although $\Gamma \not\models P(a)$. (In fact, we also have that $\Gamma \models_p^{(\mathbf{R})} \exists! x P(x)$.)

This example serves to illustrate the fact that p-minimal consequence can yield new positive ground instances of the predicate being circumscribed. By slightly modifying this theory we have an example of a theory Γ' which p-minimally entails new positive and negative ground instances of a predicate which is not among those circumscribed. Let $\Gamma' = \Gamma \cup \{a \neq b\}$. Clearly, $\Gamma' \models_p^{(\mathbf{R})} R(b,a)$ and $\Gamma' \models_p^{(\mathbf{R})} \neg R(a,b)$, although $\Gamma' \not\models R(b,a)$ and $\Gamma' \not\models \neg R(a,b)$. Of course, neither Γ nor Γ' is well founded, and this is precisely the reason that these theories yield new positive ground instances of the circumscribed predicate, or new ground instances of the other predicates.

3.3 Existence of Predicate Minimal Models

A variation on the theme of example 2.3 yields an example of a satisfiable theory with no p-minimal models:

Example 3.5 [Etherington et al 85] Let

$$\begin{aligned}\Gamma = \{ \quad & \exists x(N(x) \wedge \forall y(N(y) \rightarrow \neg s(y) = x)), \\ & \forall x(N(x) \rightarrow N(s(x))), \\ & \forall x \forall y(s(x) = s(y) \rightarrow x = y)\}\end{aligned}$$

and let $\mathbf{R} = \{N\}$. Every model of Γ has an infinite chain of p-submodels, in each of which the extension of N corresponds to the natural numbers greater than n. There is no p-minimal model for this theory, so the set of its p-minimal consequences consists of all the sentences of the language, and is thus unsatisfiable.

This suggests the following definition:

Definition 3.4 Γ is *p-satisfiable* iff Γ has a p-minimal model with respect to every set of predicates $\mathbf{R}$.

As with d-minimal consequence, there are theories which are not p-satisfiable, but possess p-satisfiable extensions. This is exemplified by the theory $\Gamma' = \Gamma \cup \{N(0)\}$. Note that, up to isomorphism, Γ has as its unique p-minimal model that in which N is interpreted by the natural numbers and s by the usual successor function. Thus Γ' is a p-satisfiable extension of a theory which is not p-satisfiable.

An interesting question that comes to mind is whether any satisfiable theory can be extended to a p-satisfiable theory. In the case of d-minimal consequence it was observed that any satisfiable theory can be made d-satisfiable by the mere addition of new symbols to the language. This is not generally the case for p-minimal consequence, but the extension of Γ to Γ' above suggests that the addition of constant symbols to the language and a set of sentences to the theory, stating that the constant symbols introduced lie in the extension of the circumscribed predicate, will result in a p-satisfiable theory in the new language.

Proposition 3.1 *Any satisfiable $\mathcal{L}$-theory Γ can be extended to a p-satisfiable $\mathcal{L}'$-theory Γ' such that:*

(i) Γ and Γ' make true the same $\mathcal{L}$-sentences; and

(ii) if $\mathcal{L}$ is finite, then both $\mathcal{L}' - \mathcal{L}$ and $\Gamma' - \Gamma$ are also finite.

Proof: First note that if Γ has finite models, then Γ is already p-satisfiable (since any finite model has a p-minimal submodel) and the proposition holds trivially. We therefore assume that Γ has only infinite models.

The proof proceeds in a similar manner to that of Proposition 2.1, by exhibiting a uniform construction that can be applied to expand any (infinite) countable $\mathcal{L}$-structure $\mathcal{M}$ to an $\mathcal{L}'$-structure $\mathcal{M}'$ and to extend any $\mathcal{L}$-theory Γ to an $\mathcal{L}'$-theory Γ', in such a way that, if $\mathcal{M}$ is a model of Γ, then $\mathcal{M}'$ is a p-minimal model of Γ'. The language of Γ is extended by the addition of new constant and function symbols; the intention here is to "name" individuals in the extension that each relation symbol in the language receives in $\mathcal{M}$; the sentences added to Γ require that, for each relation symbol of the language, individuals denoted by these new terms "remain" in the extension which that relation symbol receives in $\mathcal{M}$— thus these cannot be dropped to obtain a p-submodel of $\mathcal{M}'$, so $\mathcal{M}'$ is then a p-minimal model of Γ'. In this manner, from a model of Γ, we construct a p-minimal model of an extension of Γ.

We exhibit this construction for a finite language $\mathcal{L}$ containing a single binary relation symbol R, so as to avoid cumbersome notation that would obscure the argument; the proof generalizes easily to arbitrary first order languages, by replication of this construction for each relation symbol in the language.

Let $\mathcal{L}' = \mathcal{L} \cup \{c_1, c_2, f_1, f_2\}$. We define the structure $\mathcal{M}' = \langle M', R^{\mathcal{M}'}, c_1^{\mathcal{M}'}, c_2^{\mathcal{M}'}, f_1^{\mathcal{M}'}, f_2^{\mathcal{M}'} \rangle$ from $\mathcal{M}$ as follows. Let $M' = M$ and $R^{\mathcal{M}'} = R^{\mathcal{M}}$. Since M is countable, $M \times M$ is also countable; hence $R^{\mathcal{M}'}$ is also countable ($R^{\mathcal{M}'} \subseteq M \times M$). Let $\langle \alpha_i, \beta_i \rangle_{i \in \omega}$ be an enumeration without repetitions of $R^{\mathcal{M}'}$. (We can again assume, without loss of generality, that $R^{\mathcal{M}'}$ is infinite.) The interpretations of the constant and function symbols in $\mathcal{M}$ are defined as follows:

$$\begin{aligned} c_1^{\mathcal{M}'} &= \alpha_1 \\ c_2^{\mathcal{M}'} &= \beta_1 \\ f_1^{\mathcal{M}'} &= \{\langle\langle\alpha_i, \beta_i\rangle, \alpha_{i+1}\rangle \mid i \in \omega\} \cup \{\langle\langle \mathbf{x}, \mathbf{y}\rangle, \mathbf{x}\rangle \mid \langle \mathbf{x}, \mathbf{y}\rangle \notin R^{\mathcal{M}'}\} \\ f_2^{\mathcal{M}'} &= \{\langle\langle\alpha_i, \beta_i\rangle, \beta_{i+1}\rangle \mid i \in \omega\} \cup \{\langle\langle \mathbf{x}, \mathbf{y}\rangle, \mathbf{x}\rangle \mid \langle \mathbf{x}, \mathbf{y}\rangle \notin R^{\mathcal{M}'}\} \end{aligned}$$

Let $\Gamma' = \Gamma \cup \{R(c_1, c_2), \forall x \forall y (R(x, y) \rightarrow R(f_1(x, y), f_2(x, y)))\}$.

Note that $\mathcal{M}$ and $\mathcal{M}'$ have the same domain and the same interpretation for R, so they make true the same $\mathcal{L}$-sentences. The interpretations of the new terms (built from the new constant and function symbols) "follow" the enumeration $R^{\mathcal{M}'}$. (The set $\{\langle\langle \mathbf{x}, \mathbf{y}\rangle, \mathbf{x}\rangle \mid \langle \mathbf{x}, \mathbf{y}\rangle \notin R^{\mathcal{M}'}\}$ is included in the interpretations of f_1 and f_2 because these must be defined on *all* pairs in $M \times M$ — the particular interpretation for pairs lying outside $R^{\mathcal{M}'}$ is irrelevant.) Γ' extends Γ with an axiom that requires that the interpretation of these terms be in the extension of R. Thus, $\mathcal{M}' \models_p^{\{R\}} \Gamma'$.

Any satisfiable theory Γ in a countable language has a countable model (by the downward Lowenheim-Skolem Theorem). The above construction can be applied to any (infinite) countable model $\mathcal{M}$ of Γ to obtain an expansion $\mathcal{M}'$ of $\mathcal{M}$ which is a p-minimal model of $\Gamma' \supseteq \Gamma$; thus Γ' is p-satisfiable. Clearly, if $\mathcal{L}$ is finite, the extensions to $\mathcal{L}$ and to Γ are both finite. It remains to show that Γ and Γ' make true the same $\mathcal{L}$-sentences.

Let ϕ be an $\mathcal{L}$-sentence. Since $\Gamma \subseteq \Gamma'$, $\Gamma \models \phi \Rightarrow \Gamma' \models \phi$. Now suppose $\Gamma \not\models \phi$, *i.e.*, for some $\mathcal{N}$, $\mathcal{N} \models \Gamma$, but $\mathcal{N} \not\models \phi$; again, by the downward Lowenheim-Skolem Theorem there has to be a countable such model, so we assume that $\mathcal{N}$

is countable. We can thus apply the above construction to $\mathcal{N}$ to obtain an $\mathcal{L}'$-structure $\mathcal{N}'$ expanding $\mathcal{N}$ which is a model of Γ'. As noted before, $\mathcal{N}$ and $\mathcal{N}'$ make true the same $\mathcal{L}$-sentences. Hence, $\mathcal{N}' \models \neg\phi$ and, since $\mathcal{N}'$ is a model of Γ', we have that $\Gamma' \not\models \phi$. □

Throughout this discussion on p-satisfiability, and in example 3.5 above, the careful reader may have observed some similarities with issues encountered in d-minimal consequence. In particular, the theory of example 3.5, and the argument for the lack of existence of p-minimal models, seem almost a repetition of those of example 2.3. There is indeed a close relationship between domain and predicate minimal consequence, which will be studied in a later section. As we will see, parallel results of this sort can be obtained in a more straightforward manner, once the correspondence between the various notions of minimal consequence is made precise.

4 Formula Minimal Consequence

4.1 Preliminaries

Formula minimal consequence is a generalization of predicate minimal consequence. Here, while the extensions of some predicates are being minimized, those of some of the other predicates are allowed to vary. As we will see, this has the effect of minimizing the set of objects satisfying an open formula.

Definition 4.1 Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathcal{L}$-structures and let $\mathbf{R}$ and $\mathbf{U}$ be sets of relation symbols in $\mathcal{L}$. $\mathcal{M} \subseteq_f^{(\mathbf{R},\mathbf{U})} \mathcal{N}$ ($\mathcal{M}$ is an *f-submodel* of $\mathcal{N}$ with respect to ($\mathbf{R}$, $\mathbf{U}$)) iff

(i) $M = N$;

(ii) $R^{\mathcal{M}} \subseteq R^{\mathcal{N}}$, for each relation symbol $R \in \mathbf{R}$;

(iii) $R^{\mathcal{M}} = R^{\mathcal{N}}$, for every relation symbol R, $R \notin \mathbf{R}$ and $R \notin \mathbf{U}$;

(iv) $f^{\mathcal{M}} = f^{\mathcal{N}}$ and $c^{\mathcal{M}} = c^{\mathcal{N}}$, for each function symbol f and constant symbol c.

Note that an f-submodel of a model $\mathcal{M}$ is generally *not* a submodel of $\mathcal{M}$ (in the usual model theoretic sense). This, as noted in the previous section, was also the case for p-submodels and again it follows directly from the definition of an f-submodel that if $\mathcal{M}$ is a submodel of $\mathcal{N}$ and $\mathcal{M} \subseteq_f^{(\mathbf{R},\mathbf{U})} \mathcal{N}$, then $\mathcal{M} = \mathcal{N}$. However, whereas if $\mathcal{M} \subseteq_p^{(\mathbf{R})} \mathcal{N}$ and $\mathcal{N} \subseteq_p^{(\mathbf{R})} \mathcal{M}$, then $\mathcal{M} = \mathcal{N}$, this is not the case for $\subseteq_f^{(\mathbf{R},\mathbf{U})}$, *i.e.*, the former but not the latter relation is antisymmetric. This accounts for a slight deviation in the definition of f-minimal model from the analogous definitions of d-minimal and p-minimal model:

Definition 4.2 $\mathcal{M} \models_f^{(\mathbf{R},\mathbf{U})} \Gamma$ ($\mathcal{M}$ is a *f-minimal model* of a theory Γ with respect to the sets of predicates ($\mathbf{R}$,$\mathbf{U}$)) iff $\mathcal{M} \models \Gamma$ and for every $\mathcal{N}$ such that $\mathcal{N} \models \Gamma$, if $\mathcal{N} \subseteq_f^{(\mathbf{R},\mathbf{U})} \mathcal{M}$, then $\mathcal{M} \subseteq_f^{(\mathbf{R},\mathbf{U})} \mathcal{N}$.

Definition 4.3 $\Gamma \models_f^{(\mathbf{R},\mathbf{U})} \phi$ (ϕ is an *f-minimal consequence* of Γ with respect to the sets of predicates ($\mathbf{R}$,$\mathbf{U}$) iff for all $\mathcal{M}$, $\mathcal{M} \models_f^{(\mathbf{R},\mathbf{U})} \Gamma \Rightarrow \mathcal{M} \models \phi$

Example 4.1 Let $\Gamma = \emptyset$, $\mathbf{R} = \{P\}$ and $\mathbf{U} = \emptyset$. The f-minimal models of Γ are those models in which P is given an empty extension. These are exactly the same as its p-minimal models. Moreover, this follows trivially from the above definition, so, in general, we have that if $\Gamma \models_f^{(\mathbf{R},\mathbf{U})} \phi$ and $\mathbf{U} = \emptyset$, then $\Gamma \models_p^{(\mathbf{R})} \phi$.

Thus we see that f-minimal consequence is a generalization of p-minimal consequence. From this fact it immediately follows that f-minimal consequence is also non-monotonic.

Example 4.2 Let $\Gamma = \{\forall x(E(x) \leftrightarrow \phi(x))\}$, where ϕ is a formula of one free variable which only involves the predicate P, in the language of Γ. Let $\mathbf{R} = \{E\}$ and $\mathbf{U} = \{P\}$. The f-minimal models of Γ are those in which the extension of E is smallest, and hence where the set of objects satisfying the formula ϕ is minimized.

This example serves as an explanation for the name "formula minimal consequence." So we see that formula minimal consequence can (equivalently) be defined in terms of minimizing the extension of a formula, as is done in [Etherington 86] (this is in fact more immediate from McCarthy's exposition of formula circumscription, in [McCarthy 84]). The definition given above (which is due to [Schlipf 87]) was nevertheless chosen because it is more direct and thus simplifies subsequent proofs.

Example 4.3 Let $\Gamma = \{Q(a), \forall x(\neg Q(x) \rightarrow P(x)), \neg a = b, \forall x(x = a \vee x = b)\}$, $\mathbf{R} = \{Q\}$, and $\mathbf{U} = \{P\}$. This is the theory of example 3.3, which has only models with two element domains. Recall that each of its models is isomorphic to one of six models, repeated here for ease of reference:

$\mathcal{M}_1 : Q^{\mathcal{M}_1} = \{\mathbf{a}\}$, $P^{\mathcal{M}_1} = \{\mathbf{b}\}$;

$\mathcal{M}_2 : Q^{\mathcal{M}_2} = \{\mathbf{a}\}$, $P^{\mathcal{M}_2} = \{\mathbf{a}, \mathbf{b}\}$;

$\mathcal{M}_3 : Q^{\mathcal{M}_3} = \{\mathbf{a}, \mathbf{b}\}$, $P^{\mathcal{M}_3} = \emptyset$;

$\mathcal{M}_4 : Q^{\mathcal{M}_4} = \{\mathbf{a}, \mathbf{b}\}$, $P^{\mathcal{M}_4} = \{\mathbf{a}\}$;

$\mathcal{M}_5 : Q^{\mathcal{M}_5} = \{\mathbf{a}, \mathbf{b}\}$, $P^{\mathcal{M}_5} = \{\mathbf{b}\}$;

$\mathcal{M}_6 : Q^{\mathcal{M}_6} = \{\mathbf{a}, \mathbf{b}\}$, $P^{\mathcal{M}_6} = \{\mathbf{a}, \mathbf{b}\}$;

where $M_1 = M_2 = M_3 = M_4 = M_5 = M_6 = \{\mathbf{a}, \mathbf{b}\}$. Notice that, whereas $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ are all p-minimal, only $\mathcal{M}_1$ and $\mathcal{M}_2$ are f-minimal, so, although $\Gamma \not\models_p^{(\mathbf{R})} \neg Q(b)$ and $\Gamma \not\models_p^{(\mathbf{R})} P(b)$, we have that $\Gamma \models_f^{(\mathbf{R},\mathbf{U})} \neg Q(b)$ and $\Gamma \models_f^{(\mathbf{R},\mathbf{U})} P(b)$.

Thus we see that f-minimal consequence can yield new positive ground instances of a predicate in this kind of simple theory. In this manner, it is possible to obtain new positive ground instances of any of the predicates P of a theory, by adding an axiom which defines P as the negation of a new predicate symbol Q and then using f-minimal consequence with $\mathbf{R} = \{Q\}$ and $\mathbf{U} = \{P\}$. The reader is referred to [McCarthy 84] for several more examples of the use of f-minimal consequence involving abnormal birds.

4.2 Existence of Formula Minimal Models

By analogy to p-satisfiablility and d-satisfiability we can define f-satisfiability:

Definition 4.4 Γ is *f-satisfiable* iff Γ has an f-minimal model with respect to every pair of sets of predicates $\mathbf{R}, \mathbf{U}$.

It is clear that f-minimal consequence is a generalization of p-minimal consequence and, as such, shares some of its properties. In particular, f-minimal consequence reduces to p-minimal consequence (or, p-minimal consequence is a special case of f-minimal consequence) in the case where $\mathbf{U} = \emptyset$. It immediately follows that not every satisfiable theory is f-satisfiable. As was the case for p-minimal consequence, however, we again have that every satisfiable theory can be extended to an f-satisfiable theory.

Proposition 4.1 *Any satisfiable $\mathcal{L}$-theory Γ can be extended to an f-satisfiable $\mathcal{L}'$-theory Γ' such that:*

(i) Γ and Γ' make true the same $\mathcal{L}$-sentences; and

(ii) if $\mathcal{L}$ is finite, then both $\mathcal{L}' - \mathcal{L}$ and $\Gamma' - \Gamma$ are also finite.

Proof: The proof of this proposition is identical to that of proposition 3.1, by the following observation: if $\mathcal{N} \subseteq_f^{(\mathbf{R},\mathbf{U})} \mathcal{M}'$ and $\mathcal{M}' \not\subseteq_f^{(\mathbf{R},\mathbf{U})} \mathcal{N}$, then it follows that for some $P \in \mathbf{R}$, $P^{\mathcal{N}} \subset P^{\mathcal{M}'}$. □

The next section examines the relationship between domain, predicate, and formula minimal consequence more closely.

5 Relationship Between Domain, Predicate and Formula Minimal Consequence

The intuitive motivation for the notions of domain, predicate, and formula minimal consequence has been to articulate a method of conjecturing theories in which certain properties of objects which do not follow from the original theory are refutable. For domain minimal consequence we could, in rather vague and intuitive terms, view this property as that of the existence itself of those objects. For predicate and formula minimal consequence this property is the predicate or formula being circumscribed.

As is to be expected, there is much in common among the various notions of minimal consequence. This section draws precise connections between the domain, predicate, and formula minimal consequences of a theory, namely:

- The p-minimal consequences of a theory with respect to a set of predicates **R** are exactly those sentences which are, for every set of predicates **U**, the f-minimal consequences of that theory with respect to (**R**,**U**).

- The relativizations[2] of the d-minimal consequences of a theory (to some new predicate letter in the language, P) are exactly the p-minimal consequences, with respect to $(\{P\})$, of that theory relativized to P.

5.1 Predicate vs. Formula Minimal Consequence

Predicate minimal consequence is a special case of formula minimal consequence, in which the set of predicate parameters (those relation symbols which are allowed to vary under the f-submodel operation) is empty (see section 4, example 4.1). This follows directly from the definitions of $\subseteq_p^{(\mathbf{R})}$ and $\subseteq_f^{(\mathbf{R},\mathbf{U})}$. A somewhat tighter

[2]Defined in Section 5.2 below.

connection is reflected in the minimal consequences of a theory under these two submodel relations. We begin with a lemma relating the p-minimal and f-minimal models of a theory Γ:

Lemma 5.1 $\forall\mathcal{M}\forall\mathbf{R}\forall\mathbf{U}(\mathcal{M} \models_f^{(\mathbf{R},\mathbf{U})}\Gamma \Rightarrow \mathcal{M} \models_p^{(\mathbf{R})}\Gamma)$.

Proof: Suppose $\mathcal{M} \models_f^{(\mathbf{R},\mathbf{U})}\Gamma$, thus

$$\forall\mathcal{N}((\mathcal{N} \models \Gamma \,\&\, \mathcal{N} \subseteq_f^{(\mathbf{R},\mathbf{U})}\mathcal{M}) \Rightarrow \mathcal{M} \subseteq_f^{(\mathbf{R},\mathbf{U})}\mathcal{N}), \tag{4.1}$$

and let $\mathcal{N}$ be such that

$$\mathcal{N} \models \Gamma \,\&\, \mathcal{N} \subseteq_p^{(\mathbf{R})}\mathcal{M}. \tag{4.2}$$

Now, $\mathcal{N} \subseteq_p^{(\mathbf{R})}\mathcal{M} \Rightarrow \mathcal{N} \subseteq_f^{(\mathbf{R},\mathbf{U})}\mathcal{M}$, by the definitions of $\subseteq_p^{(\mathbf{R})}$ and $\subseteq_f^{(\mathbf{R},\mathbf{U})}$, so from (4.1) and (4.2) it follows that $\mathcal{M} \subseteq_f^{(\mathbf{R},\mathbf{U})}\mathcal{N}$; but, since $\mathcal{N} \subseteq_p^{(\mathbf{R})}\mathcal{M}$, we have that $\mathcal{M} = \mathcal{N}$ (again, by the definitions of $\models_p^{(\mathbf{R})}$ and $\models_f^{(\mathbf{R},\mathbf{U})}$). Thus, $\mathcal{M} \models_p^{(\mathbf{R})}\Gamma$. □

Proposition 5.1 $\Gamma \models_p^{(\mathbf{R})}\phi \iff \forall\mathbf{U}(\Gamma \models_f^{(\mathbf{R},\mathbf{U})}\phi)$.

Proof: ($\Rightarrow$) Suppose $\forall\mathcal{M}(\mathcal{M} \models_p^{(\mathbf{R})}\Gamma \Rightarrow \mathcal{M} \models \phi)$ and that for some $\mathcal{M}$ and $\mathbf{U}$ $\mathcal{M} \models_f^{(\mathbf{R},\mathbf{U})}\Gamma$. From the latter, by Lemma 5.1, we then have $\mathcal{M} \models_p^{(\mathbf{R})}\Gamma$, and hence, from the former, we have that $\mathcal{M} \models \phi$. Thus, $\forall\mathbf{U}\forall\mathcal{M}(\mathcal{M} \models_f^{(\mathbf{R},\mathbf{U})}\Gamma \Rightarrow \mathcal{M} \models \phi)$, *i.e.*, $\forall\mathbf{U}(\Gamma \models_f^{(\mathbf{R},\mathbf{U})}\phi)$.

($\Leftarrow$) Suppose $\forall\mathbf{U}(\Gamma \models_f^{(\mathbf{R},\mathbf{U})}\phi)$. For $\mathbf{U} = \emptyset$ we have $\Gamma \models_p^{(\mathbf{R})}\phi$. □

Proposition 5.1 can be interpreted as saying that f-minimal consequence is stronger, in the sense that, given a first order theory, a sentence conjectured from that theory by p-minimal consequence with respect to a set **R** of minimized predicates, can also be conjectured by f-minimal consequence, with respect to **R**, *regardless of which predicates are allowed to vary.* The crux of this argument is the observation that all f-minimal submodels of a theory Γ with respect to (**R**,**U**) are also p-minimal with respect to **R**. On the other hand, there are f-minimal consequences that are not p-minimal consequences (see example 4.3), so, in this sense, f-minimal consequence is strictly stronger.

For practical applications, where the need for such conjectures has been felt, it seems clear that f-minimal consequence can simply replace p-minimal consequence: no generality is lost and some apparently needed strength is gained. The study of p-minimal consequence, however, is not rendered fruitless by this outcome since most properties of p-minimal consequence can thus be generalized to f-minimal consequence. In particular, since there exists a satisfiable theory that is not p-satisfiable, it follows that there is a theory that is not f-satisfiable, no matter which predicates are allowed to vary.

5.2 Domain vs. Predicate Minimal Consequence

The relationship between domain and predicate minimal consequence has received some attention in the circumscription literature, although no clear connection has been established to date[3]. In order to establish a relation between p-minimal (or f-minimal) consequence and d-minimal consequence the aim is to, in a sense, "transfer" the minimization from a predicate's extension to the domain of a model. The following definitions will be useful:

Definition 5.1 Let $\mathcal{M} = \langle M, P^{\mathcal{M}}, R_0^{\mathcal{M}}, R_1^{\mathcal{M}}, \ldots, f_0^{\mathcal{M}}, f_1^{\mathcal{M}}, \ldots, c_0^{\mathcal{M}}, c_1^{\mathcal{M}}, \ldots \rangle$ be a model for the language $\mathcal{L} = \{P, R_0, R_1, \ldots, f_0, f_1, \ldots, c_0, c_1, \ldots\}$ such that

(i) P is a one place predicate symbol,

(ii) $P^{\mathcal{M}}$ is non-empty and closed under each of the $f_0^{\mathcal{M}}, f_1^{\mathcal{M}}, \ldots$, and

[3]McCarthy claimed that domain circumscription is reducible to predicate circumscription [McCarthy 80]. His argument, which was made on syntactic grounds, was an attempt to make precise the intuition that we can view the existence of objects as a predicate which is being circumscribed, but it contained some (fatal) faults. We will not discuss McCarthy's argument, nor its faults; a detailed exposition of this is given in [Etherington & Mercer 86]. On model theoretic grounds alone it is clear that his conclusion is incorrect — as we have observed, a p-submodel is generally not a d-submodel (see section 3).

(iii) $c_i^{\mathcal{M}} \in P^{\mathcal{M}}$, for each $c_i \in \mathcal{L}$.

The *inner model* of $\mathcal{M}$ defined by P (written $\mathcal{M}^P$), is a model for the language $\mathcal{L}' = \mathcal{L} - \{P\}$ such that:

(i) $M^P = P^{\mathcal{M}}$;

(ii) $c_i^{\mathcal{M}^P} = c_i^{\mathcal{M}}$, for each $c_i^{\mathcal{M}} \in \mathcal{L}'$; and

(iii) the interpretation for all relation and function symbols of $\mathcal{L}'$ in $\mathcal{M}^P$ is the restriction of their $\mathcal{M}$ interpretations to M^P.

Observe that, although $\mathcal{M}^P$ is not defined for every $\mathcal{L} \cup \{P\}$-structure $\mathcal{M}$, every $\mathcal{L}$-structure $\mathcal{M}$ is the inner model defined by P of *some* $\mathcal{L} \cup \{P\}$-structure $\mathcal{N}$. Thus

$$\forall \mathcal{M} \exists \mathcal{N} (\mathcal{N}^P \text{ is defined } \& \mathcal{N}^P = \mathcal{M}) \tag{4.3}$$

Definition 5.2 Let Γ be a theory in a language $\mathcal{L}$ and $P \in \mathcal{L}$, a one place predicate symbol. The *relativization* of Γ to P (written Γ^P) is the theory obtained by replacing each sentence of the form $\forall x \phi(x)$ in Γ by the sentence $\forall x(P(x) \rightarrow \phi(x))$; replacing each sentence of the form $\exists x \phi(x)$ in Γ by the sentence $\exists x(P(x) \wedge \phi(x))$; adding the sentence $\exists x P(x)$; and adding the sentences $P(c)$ for each constant symbol $c \in \mathcal{L}$ and $\forall x(P(x) \rightarrow P(f(x)))$ for each function symbol $f \in \mathcal{L}$.

Note that, by the definition of a relativized theory Γ^P, its models always satisfy the conditions of Definition 5.1 (since it is the case that in all the models of Γ^P, P is given a non-empty extension[4], closed under the function interpreting each function symbol, and containing the interpretation of each of the constant symbols) so the

[4]This is assured by the inclusion of the sentence $\exists x P(x)$ to Γ^P in the case where there are no constant symbols in the language.

inner models of Γ^P are always defined. The Relativization Lemma[5] connects the inner models of a theory defined by a predicate P and the models of that theory relativized to P:

$$\mathcal{M} \models \Gamma^P \iff \mathcal{M}^P \models \Gamma. \tag{4.4}$$

The Lemma below gives the correspondence between the d-minimal submodels of the inner models and the p-minimal models of the relativized theory:

Lemma 5.2 $\mathcal{M} \models_p^{\{P\}} \Gamma^P \iff \mathcal{M}^P \models_d \Gamma$.

Proof: This follows immediately from the Relativization Lemma and the fact that $\mathcal{N} \subseteq_p^{\{P\}} \mathcal{M}$ iff $\mathcal{N}^P \subseteq_d \mathcal{M}^P$. □

So we see that the models of a theory Γ are the inner models, defined by a one-place predicate P, of Γ relativized to P and also that the d-minimal models of Γ are the inner models defined by P of the p-minimal models of Γ relativized to P. A further conclusion can be drawn at this point, relating d-minimal consequence and p-minimal consequence:

Proposition 5.2 *Let* Γ *be an* $\mathcal{L}$*-theory and* $P \notin \mathcal{L}$*. For any sentence* $\phi \in \mathcal{L}$

$$\Gamma \models_d \phi \iff \Gamma^P \models_p^{\{P\}} \phi^P.$$

Proof:

$$\begin{aligned}
\Gamma \models_d \phi &\iff \forall \mathcal{M}(\mathcal{M} \models_d \Gamma \Rightarrow \mathcal{M} \models \phi) \\
&\iff \forall \mathcal{N}((\mathcal{N}^P \text{ is defined } \& \mathcal{N}^P \models_d \Gamma) \Rightarrow \mathcal{N}^P \models \phi) \\
&\qquad \text{(from 4.3)} \\
&\iff \forall \mathcal{N}((\mathcal{N}^P \text{ is defined } \& \mathcal{N}^P \models_d \Gamma) \Rightarrow \mathcal{N} \models \phi^P) \\
&\qquad \text{(from 4.4)}
\end{aligned}$$

[5] See *e.g.*, [Ebbinghaus et al 84], pp 122-124.

$$\iff \forall \mathcal{N}((\mathcal{N}^P \text{ is defined } \& \mathcal{N} \models_p^{\{P\}} \Gamma^P) \Rightarrow \mathcal{N} \models \phi^P)$$

(from Lemma 5.2)

$$\iff \Gamma^P \models_p^{\{P\}} \phi^P$$

□

Thus we see that, although d-minimal consequence is not reducible to p-minimal consequence, there is a correspondence between the d-minimal consequences of a theory and the p-minimal consequences of the relativized theory. For example, note that if a theory is d-satisfiable, then its relativization to a one place predicate P will be p-satisfiable, and conversely. This correspondence can be exploited in the investigation of theoretical properties of the minimal consequence relations. If we have shown that there is a theory for which the d-minimal consequence relation has a certain property, then it follows immediately that the relativized theory will also have an analogous property. For example, in section 2 it was shown that there is a theory that has no d-minimal models (example 2.3) and subsequently, in section 3, it was shown that there is a theory which has no p-minimal models (example 3.5). The apparent similarity in these two theories is now made precise; we see that the latter is the relativization of the former, with respect to the circumscribed predicate. Since any theory can be relativized to a new one-place predicate symbol, the fact that there is a satisfiable theory which is not p-satisfiable follows immediately from the fact that there is a satisfiable theory which is not d-satisfiable.

Similarly, if we have shown that every theory has a certain property with regard to p-minimal consequence, then we also obtain that every theory will have the corresponding property with regard to d-minimal consequence. For example, we can obtain Proposition 2.1 from Proposition 3.1 as follows. Proposition 3.1 states that any satisfiable $\mathcal{L}$-theory can be extended to a p-satisfiable $\mathcal{L}'$-theory; therefore the relativization of any satisfiable $\mathcal{L}$-theory can be extended to a p-satisfiable $\mathcal{L}'$-

theory.[6] We also have that if the relativization of a theory is p-satisfiable, then that theory is d-satisfiable, so it follows that any satisfiable $\mathcal{L}$-theory can be extended to a d-satisfiable $\mathcal{L}'$-theory (Proposition 2.1).

In general, examples of theories with a certain domain minimal consequence property can be easily transformed into examples of theories with an analogous predicate (and formula) minimal consequence property; proofs of predicate (or formula) minimal consequence properties that hold for any theory can be transformed into proofs that the corresponding property holds for domain minimal consequence. In this sense, domain minimal consequence is a weaker notion than predicate minimal consequence, which is a weaker notion (since it is a special case) than formula minimal consequence.

6 Significant Properties of Minimal Consequence and Minimal Satisfiability for First Order Languages

So far in this chapter we have a model theoretic exposition of the notions of domain, predicate, and formula minimal consequence and some of their properties. This section presents a study of further, more general properties of minimal consequence in first order logic.

Common to all minimal consequence relations is their non-monotonicity, which in essence provides the initial motivation for their study. Also common to the minimal consequence relations presented in this chapter is that there are satisfiable

[6]At this point we will also need that extending the relativization of any satisfiable theory to a p-satisfiable theory will still be a relativization of a satisfiable theory, but this is clearly so, by the construction of the extended p-satisfiable theory (see Section 3).

theories with no minimal models and, therefore, theories for which minimal consequence will produce inconsistent conjectures. Although this fact in itself is rather natural and by no means invalidates the study of minimal consequence or circumscription, it has been felt that circumscription may be of no practical use in AI, since it was introduced primarily as a method of conjecturing common sense conclusions drawn from incomplete information. It is generally agreed, among researchers who favour logical approaches to modelling common sense reasoning, that these conjectures should at least be consistent with the available information. Thus, the fact that not all satisfiable theories are minimally satisfiable necessitates either checking theories for minimal satisfiability before using minimal consequence to generate conjectures (a highly undecidable problem), or restricting the use of minimal consequence to theories of a special syntactic form that ensures the existence of minimal models for satisfiable theories. The first of these alternatives is obviously not practicable. Etherington [Etherington 86] explores the second alternative by introducing and studying the properties of *well founded theories.* A first order theory is well founded iff each of its models has a minimal submodel. A syntactic characterization of well founded theories is not available, but there are well known classes of theories that are well founded, such as universal theories. Nevertheless, minimal consequence is found to be of limited use for AI purposes when restricted to well founded theories. This is most evident in the case of p-minimal consequence, which yields no new positive ground instances of any of the predicates circumscribed and no new (positive or negative) ground instances of any of the other predicates, as shown by [Etherington 86].

Our approach to this problem has not been to attempt to define the set of minimally satisfiable theories, but rather, to show how satisfiable theories can be extended to minimally satisfiable theories. Earlier in this chapter we discussed the significance of the language of a theory and the availability of names. This exposes another aspect in which minimal consequence differs from logical consequence, which can be exploited to provide a satisfactory solution to the problem of satisfiable theories with no minimal models. Namely, the logical consequences of a theory in a restricted language, are consequences of the same theory in an extended language (and vice versa), *i.e.*, an $\mathcal{L}$-sentence ϕ is a consequence of an $\mathcal{L}$-theory Γ iff ϕ is a consequence of Γ taken as an $\mathcal{L}'$-theory. This is not true for minimal consequence, as we saw in section 2, where the addition of a new constant symbol to the language of a theory that was not d-satisfiable rendered it d-satisfiable — thus, minimal consequences of the theory in the restricted language (*i.e.*, all the sentences of this language) are not necessarily minimal consequences of the theory in the extended language (since this is d-satisfiable).

For domain, predicate, and formula notions of minimal satisfiability, every satisfiable theory can be expanded to a minimally satisfiable theory in an extended language. In the case of d-minimal consequence, in fact, we found that this expansion was trivial — it is sufficient to extend the language (see proposition 2.1). In this section we will consider some further model theoretic properties of minimal consequence. Apart from the theoretical interest such explorations offer, there is a significant computational motivation for their pursuit. We will examine issues of compactness and [discuss the work of [Schlipf 87] on the] complexity of the minimal consequence relations, as these directly relate to the existence and feasibility of formal proof systems.

6.1 Elementary Properties

Some, but not all, properties of logical consequence find natural analogues in properties of minimal consequence. The lack of monotonicity for minimal consequence invalidates some of the properties one would take for granted. The following propositions assure us that some very fundamental properties still hold (here the symbol "$\models_m$" is used to denote minimal consequence in general):

Lemma 6.1 *If $\mathcal{M} \models_m \Gamma$, $\mathcal{M} \models \Gamma'$, and $\Gamma \subseteq \Gamma'$, then $\mathcal{M} \models_m \Gamma'$*

Proof: Let $\mathcal{M} \models_m \Gamma$, $\mathcal{M} \models \Gamma'$, and $\Gamma \subseteq \Gamma'$. Suppose $\mathcal{N} \subseteq_m \mathcal{M}$ and $\mathcal{N} \models \Gamma'$; then $\mathcal{N} \models \Gamma$ (by the monotonicity of logical consequence), so $\mathcal{M} \subseteq_m \mathcal{N}$ (since $\mathcal{M} \models_m \Gamma$). Thus, $\mathcal{M} \models_m \Gamma'$. □

Proposition 6.1 (Deduction Theorem)*([Shoham 87])* $\Gamma \cup \{\phi\} \models_m \psi \Rightarrow \Gamma \models_m \phi \rightarrow \psi$.

Proof: Suppose $\Gamma \cup \{\phi\} \models_m \psi$. From Lemma 6.1 we have

$$\begin{aligned} \mathcal{M} \models_m \Gamma \,\&\, \mathcal{M} \models \phi &\Rightarrow \mathcal{M} \models_m \Gamma \cup \{\phi\} \\ &\Rightarrow \mathcal{M} \models \psi \end{aligned}$$

(since $\Gamma \cup \{\phi\} \models_m \psi$). Hence $\mathcal{M} \models_m \Gamma \Rightarrow (\mathcal{M} \models \phi \Rightarrow \mathcal{M} \models \psi)$, *i.e.*, $\Gamma \models_m \phi \rightarrow \psi$.
□

Proposition 6.2 (Modus Ponens) *If* $\Gamma \models_m \phi$ *and* $\Gamma \models_m \phi \rightarrow \psi$, *then* $\Gamma \models_m \psi$.

Proof: Suppose $\Gamma \models_m \phi$ and $\Gamma \models_m \phi \rightarrow \psi$. So

$$\forall \mathcal{M}(\mathcal{M} \models_m \Gamma \Rightarrow (\mathcal{M} \models \phi \,\&\, \mathcal{M} \models \phi \rightarrow \psi)).$$

Therefore $\forall \mathcal{M}(\mathcal{M} \models_m \Gamma \Rightarrow \mathcal{M} \models \psi)$. □

The deduction theorem and modus ponens attest to some similarity of minimal consequence to the familiar notion of logical consequence. Of course, the extent of this similarity is very limited; the converse of the deduction theorem fails for minimal consequence. For example, even in sentential logic, $\emptyset \models_s p \rightarrow q$ (since $\emptyset \models_s \neg p \wedge \neg q$), but $p \not\models_s q$.

In chapter 3 we have encountered another significant effect of the non-monotonicity of minimal consequence: the breakdown of the equivalence between the two widely accepted formulations of compactness, *i.e.*, the compactness of the satisfiability property and the compactness of the consequence relation. This of course did not depend on the particular definition of minimal consequence in sentential logic and is also the case for first order notions of minimal consequence. Similarly, minimal consequence in sentential logic was found to be a significantly more complex notion than its usual logical counterpart.

6.2 Compactness

There are a number of important model theoretic properties of first order languages which underlie their applicability in a variety of ways. Perhaps the most basic property of a first order language from a model theoretic point of view is the compactness of the first order consequence relation. As discussed in chapter 2 (section 3), compactness underlies the possibility of elaborating a deductive calculus which captures the consequence relation.

In a computational setting, a syntactically defined logical calculus is, generally speaking, a requirement. One route, of course, is to attempt its definition straight off and, failing that, restrict it to a "reasonable" fragment, where it can be made sound and complete. We will, however, proceed in a different way, which is to consider first whether it is at all possible to construct such a deductive calculus by determining first whether the consequence relation which it aims to formalize is compact; in this manner we generally uncover interesting model theoretic properties of the consequence relation, which are useful, either in proving completeness (in the case where a complete formalization is possible), or in the search for appropriate fragments that can be formalized (in the case where the consequence relation is not compact).

As we have seen in chapter 3, section 3, the minimal consequence relation is not compact even for a propositional language. From this it clearly follows that minimal consequence is not compact for a first order language (by a simple transformation of the example given in the proof of proposition 3.2). In section 3, it was also pointed out that the compactness of minimal satisfiability and the compactness of minimal consequence must be distinguished, because, although these are equivalent notions in classical logic, they are no longer equivalent in the case of a non-monotonic consequence relation. In fact, as was shown there, minimal satisfiability for sentential logic is compact, whereas minimal consequence is not. Proposition 6.3 below shows that minimal satisfiability is also not compact, for any of the notions of minimal satisfiability that we study in first order logic.

Proposition 6.3 *(a) d-satisfiability is not compact. (b) p-satisfiability is not compact. (c) f-satisfiability is not compact.*

Proof:

(a) Let $\Gamma = \{E^i | i \in \omega\}$, where

$$E^n \equiv \exists x_1 \ldots \exists x_n \bigwedge_{1 \leq i < j \leq n} x_i \neq x_j$$

Γ has only infinite models, but no d-minimal model, since from each of its models we can construct a submodel by dropping any finite number of elements of its domain. Any finite subset of Γ, however, will have some finite models, so it will have d-minimal models. Thus, we have here an infinite theory which is not d-satisfiable, but every finite subset of which is d-satisfiable. Since Γ is not d-satisfiable it will d-minimally entail a sentence (for example, $\exists x(x \neq x)$), which will not be a consequence of any of its finite subsets, since they are all d-satisfiable.

(b) Immediate from the non-compactness of d-satisfiability, by proposition 5.2.

(c) Immediate from the non-compactness of p-satisfiability, by proposition 5.1. □

A natural question that arises is whether minimal consequence might be compact in a slightly narrower sense, that is, with respect to minimally satisfiable theories. This is not the case though. We will examine this question in the context of d-minimal consequence — it is clear that the argument generalizes to p- and f-minimal consequence. The argument rests partly on an observation made after example 2.3, namely that augmenting the language of a theory may change its d-minimal consequences. We observed that the theory $\Gamma = \{\exists x \forall y \neg s(y) = x, \forall x \forall y (s(x) = s(y) \rightarrow x = y)\}$ has no minimal models when thought of as a theory in the language $\{s\}$ (the language with only one function symbol, s), but becomes d-satisfiable when the language is expanded to $\{s, 0\}$. Similarly, Γ in the proof of proposition 6.3, which we have implicitly assumed is in the language of pure predicate calculus (the language contains no non-logical symbols) and which is not d-satisfiable, becomes d-satisfiable when we introduce a constant symbol, c, and a function symbol, f, to the language. This fact alone is quite surprising, but what is rather striking is that, with this addition, all the minimal models that now come into existence are isomorphic to the natural numbers with the usual successor operation. The reason for this is that, by the addition of the new symbols to the language of Γ, some of the models of Γ, precisely the countable models isomorphic to the natural numbers with the successor operation, will be nameable structures and thus d-minimal. Any other model (for example, structures in which the interpretation of the constant symbol has a predecessor, or structures with cycles) will not be nameable and will have proper submodels. Thus, all the d-minimal models of Γ are the models isomorphic to the natural numbers, so we have that $\Gamma \models_d \forall x \neg f(x) = c$. Now note that there are finite models of each of the E^i which are not nameable, but are d-minimal. In particular, there are (finite) minimal models of each of the E^i in which f is the identity function and where, clearly, everything has a predecessor — itself — and, therefore, for each E^i, $E^i \not\models_d \forall x \neg f(x) = c$. So we see that d-minimal consequence is not compact, even restricted to d-satisfiable theories.

6.3 Complexity

So far in this chapter we have seen some properties of minimal consequence in first order logic that suggest that it is considerably more "complicated" than ordinary logical consequence in first order logic, or minimal consequence in sentential logic. Minimal consequence in sentential logic, viewed as a relation between sets of sentences and a sentence is Π^0_2 and not Σ^0_2 (chapter 3, proposition 5.2 and corollary 5.1). Recall that in developing these results, we began with an upper bound of Π^1_2 for minimal consequence, determined by the fact that minimal consequence is definable by a $\forall\exists$ formula with quantifiers ranging over countable sets. This upper bound was later found to be too loose, in proposition 5.2, where the strict Π^0_2 bound was given. The strict bound was reached by specific considerations on the relation for sentential logic.

By considering a more sophisticated language than that of sentential logic, it may appear that the complexity of minimal consequence will be considerably higher. Note, however, that this is not necessarily the case; it is well known that logical consequence is semi-decidable for both sentential logic and first order logic. In other cases our intuition is confirmed. For example, for finite theories in sentential logic, logical consequence is decidable — a decision procedure is given by the method of truth tables — and is known to be *co-NP-complete*. In the case of finite first order theories the complexity of the logical consequence relation is higher than that for sentential logic — although the consequences of a first order theory can be effectively enumerated, there can be no decision procedure for determining whether a sentence is a consequence of another — and is known to be Σ_1^0-*complete*.

The above considerations suggest that, for first order languages, it is useful to also determine the complexity of minimal consequence for finite theories. Such an investigation is pursued in [Schlipf 87]. Recall that the first, loose upper bound for minimal consequence in sentential logic was given by its definition. As a first attempt in this line, we can begin by considering the definition of a (general) minimal consequence relation in first order logic, "$\models_m$":

$$\Gamma \models_m \phi \iff \forall \mathcal{M}(\mathcal{M} \models_m \Gamma \Rightarrow \mathcal{M} \models \phi)$$

and

$$\mathcal{M} \models_m \Gamma \iff \mathcal{M} \models \Gamma \text{ and } \forall \mathcal{N}(\mathcal{N} \models \Gamma \Rightarrow \mathcal{N} \not\subset_m \mathcal{M}).$$

Thus, $\Gamma \models_m \phi$ can be expressed by the following formula:

$$\forall \mathcal{M} \exists \mathcal{N}(\mathcal{M} \not\models \Gamma \text{ or} (\mathcal{N} \subset_m \mathcal{M} \text{ and } \mathcal{N} \models \Gamma) \text{ or } \mathcal{M} \models \phi)$$

In the case of sentential logic, this immediately yielded an upper bound on the complexity of minimal consequence, because $\mathcal{M} \models \phi$ in sentential logic is an arithmetical relation and because the cardinality of models in sentential logic is determined by the cardinality of the language; the languages that we considered were countable, which meant that the quantifiers in this formula were over countable sets and the matrix contained only number quantification (since all the relations are arithmetical). The difference in first order logic is that the cardinality of a first order structure is arbitrarily high (since it has a domain that is an arbitrary set, independent of the language). If we restrict attention to countable structures (*i.e.*, structures whose domains are countable), it is possible to proceed with the computation of the upper bound, as before. Note that for countable first order structures and theories, $\mathcal{M} \models \Gamma$ is a Δ_1^1 relation (*i.e.*, both Π_1^1 and Σ_1^1); $\mathcal{N} \subset_m \mathcal{M}$

will, in general, be arithmetic — and definitely for the orderings considered in this thesis. From these observations it follows that, restricted to countable structures, the above formula is again (as in sentential logic) Π^1_2 — since the matrix is Δ^1_1 and the leading quantifiers range over countable sets. This upper bound on the complexity of minimal consequence restricted to countable models is not lowered by a futher restriction to finite theories, since, even for sentences, $\mathcal{M} \models \phi$ is Δ^1_1.

As with sentential logic, the next question to be addressed is whether minimal consequence is indeed as uncomputable as its syntactic form suggests. By specific model theoretic considerations on sentential logic we found that minimal consequence is a Π^0_2 relation, *i.e.*, definable by an $\forall\exists$ formula with quantifiers ranging over numbers (rather than sets). In order to answer this question for first order logic it is necessary to take into account specific model theoretic properties of the particular notion of minimality under consideration. As a relation on numbers (encoding sentences of the language), minimal consequence is recursive for sentential logic, but not so for first order logic; thus, before determining its complexity for arbitrary theories, it is useful to first consider its complexity restricted to finite theories. The outcome of this, however, makes it unnecessary to consider the case of arbitrary theories, since, restricted to countably infinite models, $\models^{(\mathbf{R})}_p$ and $\models^{(\mathbf{R},\mathbf{U})}_f$ are Π^1_2-*complete* relations on $\omega \times \omega$, as shown in [Schlipf 87]; p-minimal and f-minimal consequence are thus as complex as their definition allows them to be, even in the case of finite theories.

7 Discussion

The very high complexity of minimal consequence in first order logic can inspire one of two kinds of reaction:

1. If minimal consequence is really that complicted, then there is no point in considering it further.

2. All these results are obtained by very involved pathological constructions which, we can rest assured, will never occur in practice.

The merits of minimal consequence, or any other notion that claims to capture some aspect of human reasoning, lie in how well it performs this function and should be kept separate from efficiency and effectiveness considerations. If minimal consequence is an interesting notion in itself, or if it bears some relation to human reasoning, then the study of various pathologies will lead to alternative definitions excluding these, or further our insight into problems in modelling human reasoning. For example, the theories of examples 2.3 (Section 2) and 3.5 (Section 3.3) have no minimal models. We can exclude such theories from consideration since they are not well founded — in fact, not universal — and, as has been observed with non-well founded theories in general, they do not seem to conform with human intuitions and so there can be no harm in excluding them. We can also argue that, although theories which are not well founded do not occur "naturally" — and the theories in question are prime examples — theories that express much the same content can and do occur very naturally. After all, both of these theories describe a very familiar structure, that of a linear ordering with a least element; this could, for instance, be the natural numbers or an ancestral relation.

Clearly, we need a means to represent and reason about such structures, but perhaps this is not done best in the form offered in these examples; a simple addition to the vocabulary and a trivial extension to the theory (for the latter theory), is enough to remedy the situation, as far as the existence of minimal models is concerned. In fact, as we show in propositions 2.1, 3.1, and 4.1 all three notions of minimality that we consider enjoy this property — satisfiable theories can always be extended to minimally satisfiable theories and, if the original language is finite, its extension and the extension to the theory will also be finite.

Chapter 5

Related Work

1 Introduction

This thesis has explored the logical foundations of one approach to non-monotonic reasoning through the notion of minimal consequence. In particular, we have analysed a semantic approach to minimal consequence, paying little attention to explicit syntactical questions about notions of provability for deriving the minimal consequences of a given theory. As noted, however, our results on the complexity of minimal consequence do bear crucially on such syntactic questions and imply that further research needs to seek to identify highly expressive fragments of the logical languages under consideration, for which inferring minimal consequences (of certain classes of theories) is computationally feasible.

The present chapter discusses other research which bears on the main themes of this thesis.

2 Semantic Approaches to the Study of Circumscription

There is a fast growing body of literature in AI and logic programming on non-monotonic reasoning, much of which deals with syntactic inference mechanisms that (directly or indirectly) attempt to capture some aspect of minimal consequence. These are either tailored to specific applications or are not developed and understood to the point where they may be useful. The most significant of these elaborations were discussed in Chapter 2, since they undeniably attest to the interest in minimal consequence and play a large part in motivating our work. Thus, the goal of this thesis has been to provide a theoretical framework which can serve as a basis for the development of robust systems for non-monotonic reasoning across a range of domains. The main theme has been the development of the semantics of a number of notions of minimal consequence based on notions of minimality of structures derived from classical model theory, on the one hand, and McCarthy's (and others') investigation of circumscription, on the other, to the point where the results become relevant to the design of computational systems for non-monotonic reasoning.

In the literature, there have been some other results reported which are relevant to this theme. As discussed in Chapter 4, the work of Schlipf provides extensive information about the model theory of an interesting semantic formulation of McCarthy's predicate and formula circumscription for first order logic, and develops surprising results about the complexity of these consequence relations [Schlipf 87]. In Chapter 3 we undertake a similar study of the complexity of minimal consequence relations in sentential logic. Although the results of Schlipf and our complexity results are similar in nature, their impact is very different — the former concerning the applicability of circumscription formalisms and, in general, first order minimal consequence relations, the latter concerning the applicability of closed world reasoning of the kind employed in logic programming, *i.e.*, minimal consequence relations restricted to Herbrand models.

A few earlier semantic studies of circumscription-related notions in first order logic contributed to a better understanding of minimal consequence relations [Davis 80], [Lifschitz 86a], and [Etherington 86]. This work, which is also discussed in earlier chapters, was primarily concerned with the existence of minimal models and with finding ways to ensure it. Part of our work in sentential logic had a similar motive: to characterize the class of theories that have unique minimal models. The work of Schlipf, on the other hand, is instead aimed at exploiting the "sparseness" of minimal models for certain theories, to prove strong properties about minimal consequence. Another part of our work in sentential logic was in the spirit of Schlipf's approach, exploiting the fact that there exist theories whose minimal consequences comprise highly undecidable sets, to prove results about the complexity of the minimal consequence relation.

The work of Etherington, which covers a broad range of themes in reasoning with incomplete information, also includes some results about minimal consequence in first order logic [Etherington 86]. The main problems addressed by Etherington, in this line, are: (1) the existence of minimal models; (2) the use of p-minimal consequence in obtainining new ground sentences; (3) the relationship between predicate and domain circumscription; and (4) the (semantic) formulation of f-minimal consequence based on formula circumscription.

With respect to (1) and (2), Etherington introduces and studies the properties of well founded theories[1]. Etherington does not give a characterization of well founded theories, but shows that: (a) universal theories are well founded for domain, predicate, and formula minimal consequence, and (b) every positive ground literal of a circumscribed predicate is a p-minimal consequence of a well founded theory iff it is a logical consequence of that theory; and every ground literal (positive or negative) of any of the other predicates is a p-minimal consequence of a well founded theory iff it is a logical consequence of that theory — thus p-minimal consequence (and circumscription) does not produce any new positive ground instances of the circumscribed predicates or any new (positive or negative) ground

[1] Recall that a theory is well founded iff each of its models have a minimal submodel.

instances of the other predicates. This work was discussed in Chapter 4, where it was compared with the results of our work on this subject. We discussed the significance of the language of the theory to the availability of names for the objects in the domain or the extension of a relation in a structure. Our approach to (1) (*i.e.*, to ensuring the existence of minimal models for satisfiable theories) exploited this fact to show that satisfiable theories can be extended to minimally satisfiable theories in an extended language. The extensions make true the same sentences in the original language as the original theory. We showed this for all three notions of minimality, but in the particular case of d-minimal consequence we indeed have a stronger result: The extension to the language of a theory suffices — the same (satisfiable) theory, now taken as a theory of the extended language, will be d-satisfiable. With respect to (2), we gave examples of theories that establish the converses of Etherington's results, namely: (a) a non-well founded theory which yields new positive ground instances of the circumscribed predicate by p-minimal consequence and (b) a non-well founded theory which yields new positive and negative ground instances of other predicates (not among those circumscribed) by p-minimal consequence.

With respect to (3), Etherington shows that d-minimal consequence is not subsumed by p-minimal consequence (as was claimed in [McCarthy 80]), by showing that d-minimal consequence can be used to conjecture domain closure axioms[2], whereas p-minimal consequence cannot. In Chapter 4 we showed that, although d-minimal consequence is not a special case of p-minimal consequence, there is a definite connection between the two notions. We employed the model theoretic notions of inner models and relativizations to make this connection precise, namely, to show that the relativization of a sentence ϕ is a p-minimal consequence of the relativization of a theory Γ iff ϕ is a d-minimal consequence of Γ. This type of correspondence between d-minimal consequence and p-minimal consequence can be exploited in further theoretical investigation of these notions. Indeed, later in

[2]See Chapter 2, Section 4.

Chapter 4, it was used to show the non-compactness of p-minimal consequence from the non-compactness of d-minimal consequence.

With respect to (4), Etherington formulates f-minimal consequence, based on McCarthy's formula circumscription [McCarthy 84]. The definition that we used in Chapter 4 is different, although as we observed (and as Etherington also observes) they are equivalent. Formula circumscription and formula minimal consequence are, by definition, extensions of predicate circumscription and predicate minimal consequence, respectively. In Chapter 4 we showed a stronger connection, namely that the p-minimal consequences of a theory with respect to a set of predicates $\mathbf{R}$ are exactly those sentences which, for every set of predicates $\mathbf{U}$ are the f-minimal consequences of that theory with respect to $\mathbf{R}$ and $\mathbf{U}$. Using this, later in Chapter 4, we established the non-compactness of f-minimal consequence from the non-compactness of p-minimal consequence.

In this thesis we have pointed out the distinction among two notions of compactness in logic: compactness of satisfiability and compactness of consequence. While in classical logic these notions are interchangeable, the distinction is very significant in the study of non-monotonic consequence relations (see Chapter 2). Thus, although from the fact that every satisfiable sentential theory is minimally satisfiable it follows immediately that minimal satisfiability is compact in sentential logic, it is still possible that minimal consequence is not compact, as indeed we showed in Chapter 3.

It should be pointed out that our results on non-compactness can also be obtained (more directly) as corollaries to complexity results — those given in Chapter 3, for minimal consequence in sentential logic, and Schlipf's results, on the complexity of first order minimal consequence relations. The interest of our proofs of non-compactness, however, lies in exhibiting a particular (infinite) theory and a particular sentence, in each case, such that the sentence is a minimal consequence of the theory, but not a minimal consequence of any of its finite subsets.

3 Alternative Notions of Minimality and Applications

As noted several times in the course of this thesis, the notion of minimal consequence is sensitive to the notion of minimality of structures in terms of which it is defined. This fact has also been observed by other researchers (see *e.g.*, [McCarthy 84], [Lifschitz 84], [Etherington 86], and [Shoham 87]) and it suggests that the investigation of other notions of minimality — that is, notions based on orderings of structures different from those we consider — might provide valuable approaches to non-monotonic reasoning in certain application domains.

Shoham casts existing notions in non-monotonic reasoning in this framework and draws comparisons between them. Since he considers orderings among structures that are not readily seen as relations of "size," he terms these *preferential orderings*, and their suprema *preferred models*. For example, we can express d-minimal consequence in Shoham's terminology by defining a preference ordering among structures such that $\mathcal{M}$ is preferred over $\mathcal{N}$ iff $\mathcal{M}$ is a submodel of $\mathcal{N}$; then, we see that a structure is a d-minimal model of a theory Γ iff it is a most preferred model of Γ, *i.e.*, it is a supremum with respect to the preference ordering[3]. Shoham's framework is indeed very general and is, in fact, what in this thesis has been laid out as background in Chapter 2 (in other works also, *e.g.*, [Etherington 86], [Lifschitz 84], [Kautz 86], *etc*). As we have seen, at this level of generality, very little can be said about non-monotonic consequence relations, except for some simple properties that we discuss in Chapter 4, Section 6.1. Shoham's work is rather directed at showing how this framework can be used to compare some seemingly different approaches to non-monotonic reasoning, such as

[3]Note that there is a reversal in the direction of the ordering here, but since we are dealing with dual notions this is insignificant.

autoepistemic reasoning (in particular, the non-monotonic modal logic developed in [Halpern & Moses 85]) and circumscription.

Motivated by McCarthy's minimization ideas, several researchers have striven to define new orderings among structures that give rise to consequence relations that are more appropriate for particular aspects of common sense reasoning. Examples of this type of work are [Lifschitz 86b], [Kautz 86], and [Hintikka 88].

In the first of these, Lifschitz defines a notion of pointwise circumscription, which is a generalization of formula circumscription, based on the following idea: predicate extensions are minimized and a model is said to be minimal iff no single element can be dropped from the extension of the minimized predicate while still satisfying the theory. This idea is adapted as a generalization of formula circumscription, where several predicates are jointly minimized while others are allowed to vary. The difference between a minimal model in this sense and a p-minimal model, say, can be seen from a simple example. Let $\Gamma = \{P(a) \leftrightarrow P(b)\}$. The model where the extension of P contains two distinct interpretations of a and b is a pointwise minimal model of Γ (since no single element can be dropped from the extension of P), but is not p-minimal (because both elements may be dropped from the extension of P). Lifschitz discusses the application of pointwise circumscription to problems that arise in common sense reasoning and argues for the flexibility of his approach.

The work of Kautz concerns one particular application of minimal consequence, to temporal reasoning, which provides a solution to the "persistence problem" — given that no relevant action, or perhaps no action at all, occurred over a stretch of time, one may need to infer that certain facts do not change their truth values over that time.

Unlike Lifschitz, who adds some sophistication to the familiar methods of defining minimal models, and Kautz, who shows how these methods can be applied to give a satisfactory solution to a particular problem in common sense reasoning, Hintikka proposes a new approach to minimal consequence (restricted to the case of theories with finite models) that employs his notion of an m-automorphism. Hintikka argues that his approach captures McCarthy's idea of "small models" better than circumscription and yields conclusions that are more natural. It would appear that his proposal attempts to combine d-minimal and p-minimal consequence to produce a consequence relation that is superior to both (in the sense of being more natural), but no serious comparison is drawn in the paper. Moreover, the exposition of the main technical notion — that of an m-automorphism — is too fragmentary to allow such a comparison.

Another piece of research concerned with an application of minimal consequence is [Papalaskari & Bundy 84]. This work shows how the use of predicate circumscription in (natural language) question answering can be guided by contextual information, in order to produce conjectures that accord with the maxims of cooperative conversation [Grice 75]. In particular, the topic of a question may be used to choose the predicate which is to be circumscribed, and the conjectures thus produced conform with Grice's quantity maxims.

4 Logic Programming and Databases

This thesis has not directly dealt with any first order minimal consequence relations restricted to Herbrand structures (such as those of concern to logic programming or databases), relying on the fact that interesting issues about these can be accomodated via the study of minimality for sentential theories. We maintain the superiority of this approach, because it isolates the important features of minimal consequence of this kind, but we also acknowledge the limitations of our particular undertaking. First, it is necessary to make explicit the connection of our results to real issues encountered in logic programming. In some cases this is not trivial and requires further research.[4] Second, our study of minimal consequence in sentential logic covers only one type of minimality of structures. Recent research

[4]An example of a trivial connection is between Proposition 3.1 (Chapter 3, Section 3), which shows that every satisfiable theory is minimally satisfiable, and the well known result that states that every positive disjunctive database has a minimal Herbrand model.

in logic programming has introduced new orderings of the Herbrand models of databases [Przymusinski 87]. These aim to capture the "intended semantics" of logic programs and arose out of work in stratification (see *e.g.*, [Apt et al 86] and [Van Gelder 86]). Przymusinski refers to these as *preference orderings*[5] and to models such that no other models are preferred to them as *perfect models.* The objective in defining perfect models is to "choose" among the minimal Herbrand models of the database those that are most likely to reflect the intended interpretation of the database — perfect models are always minimal, although not all minimal models are perfect. preference orderings and perfect models are dependent on the syntax of the database. This represents a radical departure from all the orderings considered so far in this thesis and elsewhere. The intuition behind such a proposal has its source in a hypothesis that what the database designer chooses to represent as the head of a clause depends on the intended interpretation of the database. Indeed, Przymusinski argues that perfect models semantics leads to the "correct" (intuitive) interpretation of a database. A study of the recursion theoretic complexity of the perfect (Herbrand) models of stratified logic programs is given in [Apt & Blair 88]. In this very interesting paper, Apt and Blair show that these models lie arbitrarily high in the arithmetic hierarchy, but, under certain strict syntactic restrictions on the form of theories, the set of consequences can be recursively enumerable.

Viewed in a different light, perfect models represent an approach to completing a database ("more" than is afforded by the method of minimal Herbrand models). Our concerns with respect to completion of databases centered in specifying necessary and sufficient conditions for their completion via minimal models (*i.e.*, conditions under which databases have unique minimal models). A careful study of alternative notions of minimality in sentential logic, such as those suggested by Przymusinski's perfect models, would thus address concerns complementary to those explored in this thesis.

---

[5] Note that this is totally unrelated to Shoham's use of the same term.

Chapter 6

Further Work

1 Introduction

The main theme of this thesis has been the development of the semantics of minimal consequence relations as an approach to reasoning with incomplete information. Minimal consequence relations have been the subject of considerable attention in recent years as an approach to reasoning with incomplete information. Although most often they are introduced syntactically, in retrospect it has been observed that there are serious difficulties in formalizing the underlying intuitive notions which are clearly and precisely articulated in semantic terms. The notions of minimality of structures that we studied were chosen for their model theoretic simplicity, their applicability to problems in Logic Programming, and because of a generally held belief that they, in some sense, reflect important aspects of human reasoning. Apart from the immediate contribution to the understanding of these consequence relations, this work identifies questions that are pertinent to the evaluation of alternative proposals for minimal consequence relations and in fact suggests alternative proposals that avoid certain shortcomings of the consequence relations that we have studied. This chapter discusses directions for future work on minimal consequence and reasoning with incomplete information, roughly falling under the following topics:

1. Alternative formulations of minimality.

2. Connections with Logic Programming.

3. Polynomial time complexity of decidable problems.

Interesting technical problems arise under all of the above topics. These are discussed in Section 2 below. Next, in Section 3, we propose a broader perspective for further research which touches on some of these topics and engages notions of learnability.

2 Mathematical Development of Minimal Consequence Relations

2.1 Alternative Formulations of Minimality

One of the views that we take of minimal consequence here, is that of a method of completing a theory, by selecting a subclass of models that seem more "natural" than the class of all models of that theory. In Chapter 4, we discussed problems arising from the fact that the definitions of minimal models in first order logic that we consider do not ensure the existence of minimal models for every satisfiable theory. Although our work indicates that this is not as serious a problem as it appears, it would be useful to study similar types of orderings of structures that *are* well founded. Alternative definitions of minimal consequence could then be advanced that are similar to the ones studied in this thesis, but which more accurately conform to the above description of selecting a non-empty class of "natural" models. Consider the theory Γ of example 2.3 (Chapter 4, Section 2), which is intended as a theory of natural numbers with the usual successor function. Recall that every model of Γ contains a standard chain and possibly any number of finite cycles and z-chains (chains unbounded in both directions); every model of Γ contains a proper d-submodel isomorphic to the natural numbers and thus Γ has no d-minimal model. Now, all these isomorphic standard chains are, in a sense, the "same" model, and we might like to say, the minimal model of Γ. We could thus modify the definition of minimal consequence to view structures in an isomorphic chain as minimal. For example, for d-minimality we would have:

$$\mathcal{M} \models_d \Gamma \iff \mathcal{M} \models \Gamma \,\&\, \forall \mathcal{N} (\mathcal{N} \subseteq_d \mathcal{M} \Rightarrow \mathcal{N} \cong \mathcal{M})$$

where $\cong$ denotes isomorphism — similarly for p-minimality and f-minimality. Under this definition, the models of Γ that consist of a standard chain would be d-minimal, but none of those containing non-standard elements would be d-minimal. Although it works for this example, this idea does not accomplish our objective, which is to ensure that every satisiable theory is minimally satisfiable[1], but illustrates the sort of approach that may be of interest to a further investigation of this question.

Research of this character could similarly aim to "correct" other shortcomings of the minimal consequence relations that we studied, although we feel that the complexity aspects would be better served by an approach that instead seeks to identify highly expressive fragments of the languages under consideration for which minimal consequence is tractable.

2.2 Connections with Logic Programming

In Logic Programming we encounter a fragment, the class of universal Horn theories (or Horn clauses), which has certain nice properties. In this thesis we have studied Horn theories in sentential logic and have shown how to extend this class while retaining at least one of its nice properties, *i.e.*, the uniqueness of minimal models. Interesting and useful work in this area would aim to lift this type of result to first order languages, exploiting connections with sentential logic offered by Herbrand's theorem. We feel that an approach similar to that taken in Chapter 3, via preservation theorems, would be most fruitful.

---

[1]The theory $\Gamma = \{\exists x \bigwedge_{i=0}^{n} L_i(x) |\, L_i = P_i \text{ or } L_i = \neg P_i\}$ is satisfiable, but has no d-minimal model, even with this definition. Moreover, each finite subset of Γ is satisfiable, so this notion of minimal satisfiability (and minimal consequence) is not compact.

In Chapter 2, it was briefly noted that a form of what we might call "minimal Herbrand consequence" is employed in Logic Programming to handle negation and the closed world assumption. It would be interesting to study this relation in isolation and connect it to d-minimal and p-minimal consequence. For example, we can define a combination of the above by looking at models that are both d-minimal and p-minimal (for all the relation symbols in the language) and then compare this relation to minimal Herbrand consequence.

2.3 Complexity of Decidable Problems

We have emphasised the importance of our complexity results and noted that the fact that they were obtained in connection to infinite theories in no way makes them irrelevant to Computer Science. Once clear connections are made to fragments of first order logic that are relevant to logic programming, it would also be fruitful to settle questions of tractability for the decidable problems. The questions considered in Chapter 3 for unrestricted sentential theories or unrestricted subconditional theories become decidable when restricted to finite theories. Their placement in the polynomial time hierarchy would involve very interesting mathematical work. It would, perhaps, be difficult to argue that this type of research would be of great practical use, since the high complexity of the unrestricted problems suggests a high degree of intractability for the finite cases. The interest is more theoretical: the polynomial time hierarchy is barely explored beyond the first two levels (P and NP–co-NP) and, as is well known, it is not even "established" (a positive answer to P=NP would collapse it)[2]. Of course, no clear connection has been made to date between the polynomial time complexity of a problem restricted to finite objects and the complexity of the unrestricted problem (in the arithmetical hierarchy). Nevertheless, there is much evidence indicating such a connection, so it is very likely that the complexity of some of the problems that we study in this thesis can be placed in one of the higher levels of the polynomial

[2] See [Stockmeyer 87] for an excellent survey of work on the polynomial time hierarchy.

time hierarchy (*i.e.*, beyond NP and co–NP). Such questions, which are naturally occuring (in the sense that they are not conceived for the purpose of being highly intractable) and are located in the higher levels of the polynomial time hierarchy, are few and can yield valuable insights that would further the understanding of the polynomial time hierarchy.

As mentioned in the discussion of alternative notions of minimality, tractability considerations should play an important role in formulating alternative notions of minimality or fragments.

3 Learnability Considerations

In formulating a good approach to reasoning with incomplete information, as with most problems in AI, there are two competing objectives. One of these tends towards Cognitive Science and concerns the modelling of human reasoning in the presence of incomplete information. The other tends more towards Computer Science and concerns the development of computational systems that successfully reason with incomplete information. Often these two objectives can be pursued simultaneously, and in AI it has often been the case that they are identified; for clearly, modelling the way people solve a problem may in fact result in a successful computational system that solves that problem, and vice versa. The case of visual perception is an obvious example where much was learned about human and animal perception and this knowledge was used to develop successful visual information processing algorithms. For the problem of signature verification, on the other hand, this identification has not yet proved to be useful — the most successful computer security systems are not signature verifiers, but public key encryption systems.

The study of minimal consequence undertaken in this thesis has been motivated by both of these objectives. Modelling the type of reasoning performed by people when faced with incomplete information relates to the first objective. Logic programming, which is founded on a closely related methodology, represents a stride towards accomplishing the second objective. Examples and speculations on the use of minimal consequence in human reasoning, on the one hand, and the successful use of minimal consequence in logic programming, on the other, suggest that a common approach to reasoning with incomplete information, based on minimal consequence, can be viable.

The main topic for further research issuing from this thesis is the investigation of alternative notions of minimality. In the previous section, we have discussed certain guidelines for the formulation and evaluation of alternative minimal consequence relations. The criteria that emerge are of two general types — naturalness and complexity. Naturalness can only be assessed in rudimentary terms, by means of examples or through very general (and often untestable) observations; the examples given in the AI literature to support the use of circumscription and the proposal in the previous section for alternative orderings of structures that are well founded can be seen as striving for naturalness. Note, however, that neither of these criteria addresses the problem of evaluating the success of a system for reasoning with incomplete information. We propose that this is the most crucial point to be addressed in future reasearch on this subject. To this end it will be most useful to consider reasoning with incomplete information (in both the human and machine case) as embedded in a *dynamic process of knowledge acquisition*. Although it is difficult to say whether the conjectures generated by a system in the presence of incomplete information are natural, there is certainly much to be said about the behaviour of that system over time. In particular, a clear measure of success for such a system is its "learning capacity." This suggests embedding the study and evaluation of proposals for reasoning with incomplete information in the context of machine inductive inference.

The earliest paradigm for machine inductive inference is that of *identification in the limit*, introduced by [Gold 67] and extended to the context of relational structures by [Osherson & Weinstein 86]. Under this paradigm, a minimal consequence relation can be construed as a learning function; the space of possibilities is a collection of structures; information about one of these structures is input to the learning function in a piecemeal fashion; at every stage the learning function conjectures a structure. Such a learning function would then be said to identify a collection of structures in the limit if and only if its output is eventually stable (after a finite number of stages) and its conjecture is in fact the structure that gives rise to its input.

Another paradigm for machine inductive inference, *polynomial learnability*, was introduced more recently by [Valiant 84]. Polynomial learnability represents a significant refinement of the ideas of machine inductive inference, by, first, imposing explicit polynomial complexity limits on the learning function, second, by accepting a stable conjecture that is "close" to the right answer, and, third, by taking into account a measure of how well the information was presented to the learner. Such refinements can be (and, in fact, have been) proposed for identification in the limit, but polynomial learnability, in addition, relates these to one another explicitly.

Viewing minimal consequence relations as learning functions in either of these paradigms immediately broadens their scope. Interesting representation issues arise and we are urged to rethink certain intuitions that led to the development of minimal consequence. In completing a theory, our intuitions tell us that we would choose among the "small" and "simple" models. Consider how one might, in general, go about completing a theory in sentential logic. Suppose that the input so far has been $\{p_2, p_4\}$. The minimal consequence relation studied in Chapter 3 would select the model that in this case is the same as the input. This model is certainly the smallest, in terms of cardinality. Suppose now that after a while, the input is $\{p_2, p_4, p_6, p_8, p_{10}, p_{12}\}$. Most people, by this stage, will begin conjecturing $\{p_{2i} | i \in \omega\}$. If this is in fact the model giving rise to this information, however, our notion of minimal consequence (now viewed as a learning function) will not identify it in the limit. This simple example suggests that the notion of "small model" could be better interpreted by considering additional representation issues, such as the size of a description of a model, rather than the size of the model itself.

We can also reconsider other aspects of minimal consequence. For example, minimal consequence treats positive and negative information differently. This is also related to representation issues. As we saw in Chapter 4, language considerations play a much larger role with respect to minimal consequence than with ordinary logical consequence. Taking an even broader perspective, we may ask why minimal models in general should be favoured over other models of an incomplete theory. Apart from being small, they reflect a systematic way of completing (or partially completing) a theory. In addition, they give weight to the positive information present. Now we may ask, why is this a good thing to do. The answer to this would probably take us back to representation issues. Our intuitions tell us that we choose the way of representing and communicating information by making a preliminary judgement of what is important. Consider again the example given above, where the input has been $\{p_2, p_4\}$. There are many simple ways to complete this theory. For instance, we could choose the model $\{p_i | i \in \omega\}$, but somehow this seems to be missing the point. Moreover, the problem seems unrelated to the size of description of the model as discussed above. It appears, instead, that this has more to do with loss of information. Nevertheless, in the dual situation, where the information given consists of $\{\neg p_2, \neg p_4\}$, the unique minimal model is $\emptyset$. Are we to consider this problematic as well? Note that we are, in a sense, free to choose what we represent by the letters p_2 and p_4. For instance, p_2 could stand for the sentence "it is sunny" so that $\neg p_2$ would stand for "it is not sunny" (or "it is overcast"), or conversely, and the kind of input would depend on this (*i.e.*, whether the input is positive or negative), and similarly for p_4. Future research should aim to clarify questions such as these.

Summarizing, we conclude that learnability considerations are of primary importance to further research on reasoning with incomplete information, since these constitute the only means through which we can rigorously tackle the question of naturalness — to which this thesis and all past research on this subject frequently alludes.

Chapter 7

Conclusions

This thesis has presented a study of minimal consequence as a semantic approach to reasoning with incomplete information. Minimal consequence is a semantic notion implicit in past attempts to deal with the problem of incomplete information, in domains as disparate as common sense reasoning and logic programming. Our objectives have been twofold:

1. to formulate minimal consequence for sentential logic and provide a complete and detailed study of its model theoretic and recursion theoretic properties that bear on practical applications.

2. to investigate properties of minimal consequence for first order languages where these have a direct impact on the applicability and power of circumscription formalisms.

1 Sentential Logic

Regarding the first, we introduced minimal consequence in sentential logic and showed that the minimal consequence relation is not compact and is Π^0_2 and not Σ^0_2. The failure of compactness for minimal consequence, even in the case of sentential logic, rules out the possibility of elaborating a deductive calculus that captures this consequence relation. In this, we also observed that there is a break in the notions of compactness with respect to satisfiability and consequence. Although there is a definitional distinction between compactness of the consequence relation versus compactness of satisfiability in classical logic, the two can be shown to be equivalent and are, in fact, both true. This equivalence derives partly from the monotonicity of classical logic and naturally does not hold for a non-monotonic consequence relation such as minimal consequence. In our study of sentential logic we observe that minimal satisfiability is compact, whereas minimal consequence is not. For first order languages, however, neither relation is compact — for any reasonable construal of minimality.

The recursion theoretic results indicate that minimal consequence, even in the simplest case offered by sentential logic, is a rather complex relation: it is as hard as deciding whether a given program (in a Turing equivalent programming language) terminates on every input (a standard Π^0_2-complete problem). We thus conclude that for practical purposes it is essential to look at specific fragments where this difficulty does not arise. One such fragment, extensively studied in the literature, is provided by the Horn theories. Horn theories have unique minimal models and can be completed by the addition of every sentence that is not provable by the theory (negation as failure). We introduce another such fragment, the *subconditional theories* and show that it is the largest set of theories, up to logical equivalence, that has this property. We thus give a complete characterization of the class of theories in sentential logic that can be completed by negation as failure. We also show that the class of theories consistently completable by negation as failure is Π^0_2 and not Σ^0_2.

2 First Order Logic

In first order logic, minimal consequence is the semantic notion underlying circumscription formalisms. d-minimal, p-minimal, and f-minimal consequence are obtained by varying the type of minimization involved and correspond to domain, predicate, and formula circumscription, respectively. The results, again, are model theoretic and serve to clarify properties of minimal consequence in the predicate calculus which have an impact on the usefulness and applicability of circumscription formalisms (in fulfilment of the second objective above). The principal theme in the study of minimal consequence for first order languages presented here centres on yet another divergence between classical logical consequence and minimal consequence, namely, unusual phenomena which arise when considering the minimal consequences of a theory in an extended language. As was the case for the distinction between compactness of satisfiability and compactness of consequence, the minimal consequence relation differentiates between a theory in a language $\mathcal{L}$ and the same theory in an extension of $\mathcal{L}$, whereas in classical logic if the theory does not use any of the additional symbols, it is a conservative extension of the theory with respect to the original language. Using this fact, we were able to show that, although it is well known that there are satisfiable theories which are not minimally satisfiable, every satisfiable theory can be extended to a minimally satisfiable theory in an extended language. This may be a trivial extension, as in the case where the theory is already minimally satisfiable, or it may involve adding a finite number of symbols to the language and (in the case of p- and f-minimal consequence) a finite number of sentences to the theory.

By approaching non-monotonic reasoning semantically, specifically circumscription, we were able to clarify other questions in the recent literature of the subject, regarding subsumption relationships among the various circumscription formalisms. Using the notion of relativization, we established precise connections between domain, predicate, and formula circumscription.

Bibliography

[Apt & Blair 88] K. Apt and H. Blair. Arithmetic classification of perfect models of stratified programs. In *Proceedings of the Fifth International Logic Programming Conference*, pages 765–779, MIT Press, 1988.

[Apt et al 86] K. Apt, H. Blair, and A. Walker. Towards a theory of declarative knowledge. In J. Minker, editor, *Proceedings of the Workshop on Foundations of Deductive Databases and Logic Programming*, pages 546–623, Washington, August 1986.

[Chang & Keisler 73] C.C. Chang and H.J. Keisler. *Model Theory.* North Holland, 1973.

[Clark 78] K.L. Clark. Negation as failure. In *Logic and Data Bases*, pages 293–322, Plenum Press, 1978.

[Davis 80] M. Davis. The mathematics of non-monotonic reasoning. *Artificial Intelligence*, 13, 1980.

[Ebbinghaus et al 84] H. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic.* Springer-Verlag, 1984.

[Enderton 72] H.B. Enderton. *A Mathematical Introduction to Logic.* Academic Press, 1972.

[Etherington & Mercer 86] D. Etherington and R. Mercer. Domain circumscription revisited. In *Proceedings of CSCSI-SCEIO*, Montreal, Canada, May 1986.

[Etherington 86] D. Etherington. *Reasoning with Incomplete Information.* PhD thesis, University of British Columbia, 1986.

[Etherington et al 85] D. Etherington, R. Mercer, and R. Reiter. On the adequacy of predicate circumscription for closed world reasoning. *Computational Intelligence*, 1(1), February 1985.

[Gallaire et al 84] H. Gallaire, J. Minker, and J. Nicolas. Logic and databases: a deductive approach. *Computing Surveys*, 16(2):153–185, 1984.

[Gold 67] M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[Grice 75] H.P. Grice. Logic and conversation. In *Syntax and Semantics: Speech Acts*, pages 41–58, Academic Press, 1975.

[Halpern & Moses 85] J. Halpern and Y. Moses. Towards a theory of knowledge and ignorance: preliminary report. In K. Apt, editor, *Logics and Models of Concurrent Systems*, pages 459–476, Springer-Verlag, 1985.

[Hinman 78] P.G. Hinman. *Recursion-Theoretic Hierarchies.* Springer-Verlag, 1978.

[Hintikka 88] J. Hintikka. Model minimization — an alternative to circumscription. *Journal of Automated Reasoning*, 4:1–13, 1988.

[Kautz 86] H. Kautz. The logic of persistence. In *Proceedings of the AAAI*, Philadelphia, August 1986.

[Kowalski 79] R. Kowalski. *Logic for Problem Solving. Artificial Intelligence Series*, North Holland, 1979.

[Kreisel 67] G. Kreisel. Informal rigour and completeness proofs. In I. Lakatos, editor, *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam, 1967.

[Lifschitz 84] V. Lifschitz. Some results on circumscription. In *Workshop on Non-Monotonic Reasoning*, New Paltz, NY, October 1984.

[Lifschitz 86a] V. Lifschitz. On the satisfiability of circumscription. *Artificial Intelligence*, 28:17–25, 1986.

[Lifschitz 86b] V. Lifschitz. Pointwise circumscription: preliminary report. In *Proceedings of the AAAI*, Philadelphia, August 1986.

[McAllester 78] D. McAllester. *A three valued truth maintenance system*. Technical Report AI Lab Memo 473, MIT, 1978.

[McAllester 80] D. McAllester. *An outlook on truth maintenance*. Technical Report AI Lab Memo 551, MIT, 1980.

[McCarthy 77] J. McCarthy. Epistemological problems in artificial intelligence. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 1138–1144, Cambridge, Mass., 1977.

[McCarthy 80] J. McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13, 1980.

[McCarthy 84] J. McCarthy. Applications of circumscription to formalizing common sense reasoning. In *Workshop on Non-Monotonic Reasoning*, New Paltz, NY, October 1984.

[McDermott & Doyle 80] D. McDermott and J. Doyle. Non-monotonic reasoning I. *Artificial Intelligence*, 13, 1980.

[Minsky 75] M. Minsky. A framework for representing knowledge. In Winston P., editor, *The Psychology of Computer Vision*, McGraw-Hill, 1975.

[Osherson & Weinstein 86] D. Osherson and S. Weinstein. Identification in the limit of first-order structures. *Journal of Philosophical Logic*, 15:55–81, 1986.

[Osherson & Weinstein 88] D. Osherson and S. Weinstein. Paradigms of truth detection. *Journal of Philosophical Logic*, to appear, 1988.

[Papalaskari & Bundy 84] M. A. Papalaskari and A. Bundy. Topics for circumscription. In *Workshop on Non-Monotonic Reasoning*, New Paltz, NY, October 1984.

* See page 124.

[Perlis & Minker 86] D. Perlis and J. Minker. Completeness results for circumscription. *Artificial Intelligence*, 28, 1986.

[Przymusinski 88] T. Przymusinski. On the declarative semantics of deductive databases and logic programs. In J. Minker, editor, *Foundations of Deductive Databases and Logic Programming*, pages 193–216, Morgan Kaugmann, 1988.

[Reiter 78] R. Reiter. On closed world databases. In *Logic and Data Bases*, pages 56–76, Plenum Press, 1978.

[Reiter 80a] R. Reiter. Equality and domain closure in first order databases. *JACM*, 27:235–249, 1980.

[Reiter 80b] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13, 1980.

[Rogers 67] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967.

[Schank 77] R Schank. *Script, Plans, Goals and Understanding.* Erlbaum, 1977.

[Schlipf 87] J. S. Schlipf. Decidability and definability with circumscription. *Annals of Pure and Applied Logic*, 35:173–191, 1987.

[Shoham 88] Y. Shoham. *Reasoning about Change. Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, 1988.

[Stockmeyer 87] L. Stockmeyer. Classifying the computational complexity of problems. *The Journal of Symbolic Logic*, 52(1):1–43, 1987.

[Tarski 35] A. Tarski. Der Wahrheitsbegriff in der formalisierten Sprachen (English translation: "The concept of truth in formalized languages", in *Logic Semantics and Metamathematics*, Oxford 1956). *Studia Philos.*, 1, 1935.

[Valiant 84] L. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[Van Emden & Kowalski 76] M. Van Emden and R. Kowalski. The semantics of predicate logic as a programming language. *J. ACM*, 23(4):733–742, 1976.

[Van Gelder 86] A. Van Gelder. Negation as failure using tight derivations for general logic programs. In J. Minker, editor, *Proceedings of the Workshop on Foundations of Deductive Databases and Logic Programming*, pages 712–732, Washington, August 1986.

[Papalaskari & Weinstein 90] M.A. Papalaskari and S. Weinstein. Minimal consequence in sentential logic. *Journal of Logic Programming*, to appear, Spring 1990.