Notes Logic II.

IMS Stuttgart.
H. Kamp

These notes contain the material covered in the second level logic course which has been offered at the Institut für Maschinelle Sprachverarbeitung of the University of Stuttgart on an annual basis since 1992. The course is aimed at students who are familiar with the notation and use of the first order predicate calculus but have had little or no previous exposure to metamathematics.

Chapter I presents the syntax and model-theoretic semantics of classical first order logic and an axiomatic ("Hilbert style") characterization of first order deduction. The central aim of this Chapter is to establish the soundness and completeness of this deduction system, and thus the computability of the model-theoretic concepts of logical validity and logical consequence. The Chapter concludes with some easy corollaries of the Completeness Theorem (Compactness Theorem, Downward Skolem-Löwenheim Theorem) and the definition of the concepts of *model isomorphism*, *elementary equivalence* and of a *first order theory*. The Chapter closes with Robinson's preservation theorems for pure existential and for $\forall\exists$-sentences (sentences in which a quantifier-free formuila is preceded by a quantifier prefix consisting of a block of universal quantifiers followed by a block of existential quantifiers).

Chapter II presents a number of examples of first order theories - the theory of linear orderings, the first order theory of groups, the theories of Boolean Algebras and Boolean Lattices, the theory of first order Peano Arithmetic and the theory of real closed fields - and discusses some of their salient model-theoretic properties. The chapter also presents certain fragments of 1-st order predicate logic:, viz. Equational Logic (with a proof of Completeness for the equational Calculus and of Birkhoff's preservation theorem for equational sentences) and a version of feature logic. Thirdly, the Chapter contains a section on the theory of definitions (with Beth's Definability Theorem and Craig's Interpolation Theorem).

Chapter III is concerned with set theory. Set theory too is presented as a first order theory, more specifically, in the form of the so-called Theory of Zermelo-Fraenkel. But in this case the concern is not just to present yet another theory of first order logic, but also to develop, on the basis of the ZF axioms, those parts of set theory which are needed when set theory is used as framework for the formalisation of metamathematics - and more particularly those parts of metamathematics that are discussed in the two preceding chapters.

These three chapters are devoted exclusively to the classical first order predicate calculus. For anyone familiar with the history of symbolic logic over the past century this won't come as much of a surprise. In fact, many textbooks on mathematical logic have first order logic for their sole subject, and this is more or less the norm for introductions to symbolic logic. The reason for this is not only that most of the central results in formal logic pertain to first order logic, and that those pertaining to other systems often presuppose or build upon these; it is also a reflection of the mostly tacit but widespread belief that first order logic is the logical system *par excellence* - that it ithe best candidate we have for the position of 'the universal, all compassing logical formalism' - for the position of *characteristica universalis* in the sense of Leibniz' - a view that gets support from the fact that all other logical systems for which there exist precise definitions can be reduced, in some way or another, to the system of classical first order logic.

As a matter of fact the predominance of first order predicate logic is much less pronounced today than it was, say, thirty or forty years ago. There are several reasons for this, all connected with applications of formal logic in domains which forty years ago didn't even exist, or were still in their early development. Most important in this connection has been the use of mathematical logic in various branches of computer science, such as the theory of programming languages, the theory of communicating protocols that regulate parallel processing, programme verification and chip design validation. A second important domain of application is Artificial Intelligence (if, that is, AI is classified as a discipline that is distinct from Computer Science rather than as a branch of it). And lastly the variety of logical systems has grown through the use of formal logic in the semantics of natural language.

These developments have led to a rich landscape of logical formalisms. In this landscape classical first order predicate logic still holds a central place, but it is no longer one which dominates in quite the way it did in decades past.

In the light of this, exactly what place first order logic should be seen as occupying within this landscape has become a question that can no longer be ignored, and that has practical as well as purely philosophical implications. And even in an introductory text like this one it is appropriate that it should be asked at some point. But the further question that poses itself then is: When? On the one hand much could said for putting the discussion of this question up front; for after all it it is what can be said to this question which ultimately motivates the choice of the topics that will be dealt with. What speaks against this, however, is that many of the issues that should be raised in an exploration of the wider landscape are directly connected with the formal results that the text will present and so will be understandable only to a reader to whom the contents of bulk of the text (consisting of the first three chapters) are familiar. Believing that this last consideration far outweighs the first, I decided to postpone the discussion about hte relationship of classical first order logic to other logical systems till the very end. oIt has been made the subject of a separate chapter, Ch. 4.

[N.B. this chapter still needs to be added.]

Chapter I

1.1 Syntax, model theory and proof theory of classical first order predicate logic

It is assumed that the reader has some basic familiarity with the predicate calculus. There should be an awareness of how predicate logic is used in simple formalisation problems, e.g. the formalisation of mathematical structures such as orderings or Boolean algebras, and in the symbolisation of sentences and arguments from natural languages. Given this assumption it seems justified to proceed briskly with the presentation of the syntax and model theory of first order logic. In particular, we forego any informal explanation of what first order formulas 'mean'.

In fact, for a reader with antecedent exposure to the predicate calculus there won't be anything of substance in this presentation of the syntax and semantics of first order logic. Nevertheless, such a presentation cannot be dispensed with. Definitions of first order tend to vary in their details and for what is to come it must be clear which version is at issue. Moreover, it will be crucial for what follows that our characterisations of the syntax and semantics of our system are given with the formal rigour and precision none of the results that form the substance of these notes could be proved with the required logical rigour. For nearly all these results are results *about* the logical system itself. So exact proofs must be able to refer to exact definitions of the structures, objects and relations that are their targets.

One of the choices that have to be made in specifying the syntax and semantics of first order logic is the following: We can either (i) define a single formal system, with a fully fixed vocabulary and fully fixed sets of terms and formulas that can be built from it, or (ii) we can define first order logic as a family of 'first order languages', which will - while much like each other since they are all languages of first order logic - nevertheless differ form each other in one respect, viz. their so-called 'non-logical' vocabularies (roughly speaking; the part of their vocabularies which consists of their 'content words'). It has turned out that this second option has important conceptual and technical advantages over the first, which is why it is usually chosen when the focus is on the mathematical properties of first order logic. For this reason it is also the option that has been chosen here.

1.1.1 Syntax

The languages of first order predicate logic - or *first order languages*, as we will call them - differ from each other only in their non-logical vocabulary, in the predicates and functors which enable them to express contingent propositions about any particular subject matter. But they all share the *logical* vocabulary of first order logic, and with that the general rules for building complex expressions from simpler ones. We begin with the specification of this common logical vocabulary.

Def. 1 The *logical vocabulary of first order logic* consists of the following symbols:

(i) (individual) variables: v_1, v_2, v_3, ... (sometimes we also use the letters x, y, z, ... as symbols for variables)

(ii) connectives: $\neg$, &, v, $\rightarrow$, $\leftrightarrow$

(iii) quantifiers: $\forall$, $\exists$

(iv) identity: =

Each language of first order predicate logic includes the logical vocabulary listed in Def. 1. In addition it has a certain non-logical vocabulary, and as far as this vocabulary is concerned first order languages differ.

What exactly are the symbols that the non-logical vocabularies of first order languages consist of? Here there are two different policies we can follow. We can either specify a fixed stock of symbols in advance - enough to go around for any first order language we might wasnt to consider, and then define each individual language in terms of the subset of this total supply that constitutes its non-logical vocabulary. But we can also take a more liberal line. Instead of specifying one fixed stock of possible non-logical symbols in advance, we can leave it open what the non-logical symbols of any given first order are like.

This second option, which has certain advantages that cannot be properly explained at this point[1], is the one we adopt. This means however that we cannot assume that a symbol will tell us what kind of symbol it is - is it a predicate of the language or a function constant; and in either case, what is its *arity* (i.e. the number of its arguments)? - simply because of its form. So the information what kind of symbol it is must be supplied explicitly and separately: each symbol must come with a *signature,* as terminology has it, in which this information is supplied. There are various ways in which the information that signatures must provide could be encoded. For the case at hand, where we are only dealing with the first order predicates and functors, we have chosen the following encoding: A signature is a pair $<s,n>$, where s indicates whether the symbol of which it is the signature is a predicate or a functor and n is the constant's arity. This entails that the non-logical vocabulary of any first order language L can be specified as a function f whose domain is the set of non-logical constants of L and for each α in the domain $f(\alpha) = <s,n>$ is the signature of α. Furthermore, since it is only in regard of their non-logical vocabularies that first order languages can differ from each other, they are, as first order languages, fully identified by their non-logical vocabularies. Thus it is formally possible to actually *identify* them with their signatures. This identification proves very convenient in practice, and so we have adopted this stratagem.

The terms and formulas of any first order language L are built from on the one hand the symbols of their own non-logical vocabulary and on the other hand the logical symbols of first order logic, given in def. 1, that L shares with all other first order languages. It should be intuitively clear, therefore, that confusion might arise if there were overlaps

---

1 The point is this. In certain applications it is important not to have to put any upper bound on the size of the set of non-logical symbols of a language. This desideratum is incompatible with the first approach. For any set of symbols fixed in advance would impose an upper bound on the size of languages which would exclude some langauges that would be needed.

between the non-logical vocabulary of any language L and the vocabulary of Def. 1. We will therefore exclude this possibility.

These considerations lead us to the following definition:

Def. 2 A *language of first order predicate logic* is a function L from a set of "symbols" (the *non-logical constants* of L) to the *signatures* (or *logical types*) of those symbols, where a *signature* is a pair of the form $<\alpha,n>$, where

(i) α is either p (for "predicates") or f (for "functors") and
(ii) n is a natural number which specifies the *arity* (number of argument places) of the symbol.

The set of non-logical constants of L, DOM(L), must be disjoint from the logical vocabulary specified in Def. 1-

Terminology: if $L(\alpha) = <f,0>$, then α is an *individual constant of* L; if $L(\alpha) = <p,0>$, then α is a *propositional constant of* L.

Examples: {i) if $L(\alpha) = <p,2>$, then α is a 2-place predicate of L;
(ii) if $L(\alpha) = <f,1>$, then α is a 1-place functor of L; etc.

The well-formed expressions of a first order language L, its *terms* and its *formulas*, are built from its non-logical vocabulary together with the fixed logical vocabulary of Def. 1. We take it that the definitions of the terms and the formulas of L are familiar in substance and present them without further comment. The same goes for the distinction between free and bound occurrences of variables in terms and formulas.

Def. 3

1. The *terms of* a language L are defined as follows:

   (i) each variable is a *term*.
   (ii) if g is a functor with signature $<f,n>$ and $t_1,....,t_n$ are terms, then $g(t_1,....,t_n)$ is a *term of* L.

2. The *formulas of* L are defined thus:

   (i) If P is predicate of L with signature $<p,n>$ and $t_1,....,t_n$ are terms, then $P(t_1,....,t_n)$ is a *formula of* L.

(ii) If A, B are formulas of L, then $\neg A$, $(A \& B)$, $(A \vee B)$, $(A \rightarrow B)$ and $(A \leftrightarrow B)$ are *formulas of* L.

(iii) If A is a formula of L, then $(\forall v_i)A$ and $(\exists v_i)A$ are *formulas of* L.

(iv) If t_1 and t_2 are terms of L, then $t_1 = t_2$ is a *formula of* L.

N.B. For any occurrence of a formula $(\forall v_i)A$ $((\exists v_i)A)$ the corresponding occurrence of A is said to the *scope of* the corresponding occurrence of $(\forall v_i)$ $((\exists v_i))$.

<u>Def. 4</u> (Free and bound occurrences of variables)

(i) Every occurrence α of a variable v_i in a term τ is a *free* occurrence of v_i in τ.

(ii) Every occurrence of a variable in an atomic formula is *free* in that formula.

(iii) If α is a free occurrence of the variable v_j in A, then α is also a *free* occurrence in $\neg A$.

(iv) If α is a free occurrence of the variable v_j in A or in B, then α is also a *free* occurrence in $(A \& B)$, $(A \vee B)$, $(A \rightarrow B)$ and $(A \leftrightarrow B)$.

(v) If α is a free occurrence of the variable v_j in A, then it is *free* in $(\forall v_i)A$ and $(\exists v_i)A$, provided $i \neq j$.

(vi) No occurrence α in a formula A is free in A unless this follows from clauses (ii)- (v).

Every occurrence α in a term or A which is not free in A is called a *bound* occurrence of α in A.

Note that Def. 4 entails that an occurrence of v_j in A is always a bound occurrence in $(\forall v_j)A$ and in $(\exists v_j)A$.

<u>Def. 5</u> A *closed* expression of L is an expression (i.e. term or formula) of L which has no free occurrences of variables. The closed formulas of L are also called the *sentences of* L.

1.1.2 Models, Truth, Consequence and Validity

What was assumed in Section 1.1 regarding the syntax of first order languages - that the definitions are assumed to be familiar in substance - also goes for their semantics. Each first order language L determines a class of possible models for L. For each such model M we can define (i) the set of possible assignments of objects from M to the variables of first order logic and (ii) the value of any expression (term or formula) of L in M relative to any assignment **a** in M. (We say that the formula A is satisfied by in M if it gets the value 1 in M relative to **a**. (1 represents the truth value TRUE.) The values of closed terms and sentences are independent of what assignment is chosen. In particular, we can speak simply of the truth value of any sentence A of L in any model M for L: either A is true in M or A is false in M.

The definitions of satisfaction and truth in a model lead to the intuitively natural characterisations of *logical validity* and *logical consequence* (also sometimes referred to as *(logical) entailment* or as *logical implication*): the formula B of L is a *logical consequence of* the set Γ of formulas of L iff for every model M for L and every assignment **a** in M, if every C ε Γ is satisfied by **a** in M, then B is also satisfied by **a** in M. And B is logically valid when it is a logical consequence of the empty set of premises, i.e. if it is satisfied in all M by all **a**.

We take it that after this brief introduction the following definitions will be self-explanatory.

Def. 6

1. A *model for* L is a pair <U,F>, where

   (i) U is a non-empty set
   (ii) If L(g) = <f,n>, then F(g) is an n-place function from U into U
   (iii) If L(P) = <p,n>, then F(P) is an n-place function from U into {0,1}.

2. An *assignment in* a model <U,F> is a function from the set of variables into U.

Def. 7

1. The *value of* a term t of L *in* a model M = <U,F> *under* an assignment a, $[[t]]^{M,\mathbf{a}}$, is defined thus:

   (i) $[[v_i]]^{M,a} = \mathbf{a}(v_i)$

   (ii) $[[g(t_1,....,t_n)]]^{M,\mathbf{a}} = F(g)\ ([[t_1]]^{M,\mathbf{a}},...,[[t_n]]^{M,\mathbf{a}})$

2. The *truth value of* a formula A of L *in* model M *under* assignment **a**, $[[A]]^{M,\mathbf{a}}$, is defined as folows:

   (i) $[[P(t_1,....,t_n)]]^{M,\mathbf{a}} = F(P)\ ([[t_1]]^{M,\mathbf{a}},...,\ [[t_n]]^{M,\mathbf{a}})$

   (ii) $$[[\neg A]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,a} = 0 \\ 0 & \text{otherwise} \end{cases}$$

   (iii) $$[[A \;\&\; B]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,a} = [[B]]^{M,\mathbf{a}} = 1 \\ 0 & \text{otherwise} \end{cases}$$

   (iv) $$[[A \vee B]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,\mathbf{a}} = 1 \text{ or } [[B]]^{M,\mathbf{a}} = 1 \\ 0 & \text{otherwise} \end{cases}$$

   (v) $$[[(A \rightarrow B)]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,\mathbf{a}} = 0 \text{ or } [[B]]^{M,\mathbf{a}} = 1 \\ 0 & \text{otherwise} \end{cases}$$

   (vi) $$[[(A \leftrightarrow B)]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,\mathbf{a}} = [[B]]^{M,\mathbf{a}} \\ 0 & \text{otherwise} \end{cases}$$

   (vii) $$[[(\forall v_i)A]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,\mathbf{a}[u/vi]} = 1 \text{ for every } u \in U \\ 0 & \text{otherwise} \end{cases}$$

   (viii) $$[[(\exists v_i)A]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[A]]^{M,\mathbf{a}[u/vi]} = 1 \text{ for some } u \in U \\ 0 & \text{otherwise} \end{cases}$$

$$\text{(ix)}\quad [[t_i = t_j]]^{M,\mathbf{a}} = \begin{cases} 1 & \text{if } [[t_i]]^{M,\mathbf{a}} = [[t_j]]^{M,\mathbf{a}} \\ 0 & \text{otherwise} \end{cases}$$

Lemma 1: Suppose that X is a set of variables, that every variable that has free occurrences in the term or formula A is a member of X and that **a** and **b** are assignments in the model M such that for every variable v_i ε X, $\mathbf{a}(v_i) = \mathbf{b}(v_i)$. Then $[[A]]^{M,\mathbf{a}} = [[A]]^{M,\mathbf{b}}$

Proof: Although the proof of Lemma 1 is not difficult as proofs in mathematical logic go, it exemplifies some of the distinctive features of a great many proofs in this domain. In particular it provides a good illustration of the ubiquitous method of proof by induction, over well-founded but not necessarily linearly ordered domains. This is why I eventually decided to include a quite detailed proof, breaking with an earlier practice of leaving the proof as an exercise.

The task of the proof is to show that all members of an infinite set of objects - here the set of all terms and all formulas of a given first order language L - have a certain property. In the present case this is the property that a term or formula A of L has when it gets the same value in any model M under assignments **a** and **b** in M which coincide on a set of variables which includes all the free variables of A. The simplest way in which we might hope to establish this inductively is to proceed as follows:

We fix a particular model M for the language L in question as well as a given set of variables X and two assignments **a** and **b** in M such that for all x ε X $\mathbf{a}(x) = \mathbf{b}(x)$, and then prove that all terms and formulas of L have the following property (*):

(*) $[[A]]^{M,\mathbf{a}} = [[A]]^{M,\mathbf{b}}$.

To show that all terms and formulas have (*) we would then proceed inductively, i.e. by showing (i) to (iv):

(i) any atomic term A has (*);
(ii) any complex term A has (*) on the assumption that all the immediate constituent terms of A have (*),
(iii) any basic formula A has (*) on the assumption that all its constituent terms have (*); and

(iv) any complex formula has (*) on the assumption that its immediate constituent formula or fomulas has/have (*).

Unfortunately this will not work. The problem cases are the quantified formulas, i.e. formulas of the forms $(\forall v_i)$ B and $(\exists v_i)$ B. If we try to show that, say, $(\forall v_i)$ B has (*) on the assumption that B has (*), we run into the following difficulty: Our assumption is that the given variable set X contains all the free variables of $(\forall v_i)$ B. This, however, does not guarantee that X contains all variables that have free occurrences in B, for the variable v_i, which is bound in $(\forall v_i)$ B and thus need not belong to X, may well be free in B. So even if we assume that B has (*), this assumption may be of no use, since it does not tell us anything useful about B and the fixed X, **a** and **b**.

Therefore, as so often in proofs of induction, we need to "push through" the basic and recursive clauses of the definitions of *term of* L and *formula of* L, some property (**) other than (*), and which is such that once we know that all terms and formulas A have (**), we can conclude that all of them also have the property asserted in the theorem or lemma that is to be proved. In the present case the property which will do the trick is not all that different from the one which the Lemma requires us to show for all terms and formulas. (There are many inductive proofs where it is much more difficult to find the right property for which the induction can be made to go through; in fact, often finding this property is the real challenge of such proofs.) We get a property (**) which works simply by quantifying universally over the set X and the assignments **a** and **b**, rather than keeping them fixed ghroughout the inductive argument. In this way we obtain enough flexibility to deal with the quantifier cases . (The language L and the model M can still be kept fixed.)

Definition of (**). Let a language L and a moel M for L be given. (**) is the following property of terms and formulas A of L:

(**) For every set of variables X which contains all the free variables of A and every two assignments **a** and **b** in M such that for all $x \in X$, $\mathbf{a}(x) = \mathbf{b}(x)$, we have $[[A]]^{M,\mathbf{a}} = [[A]]^{M,\mathbf{b}}$.

The proof of (i)-(iv) above for the property (**) is for the most part uneventful-to-boring. The only slightly more interesting cases are those involving the quantifiers. (It is there where the difference between (**) and (*) will pay off.)

(i) According to Def. 3.1. i the atomic terms of L are the variables of first order logic. So suppose that A is the variable v_i. Let X, **a** and **b** be such that together with A they satisfy the conditions of (**) - i.e. $v_i \varepsilon X$ and **a** and **b** agree on the variables of X. So in particular $\mathbf{a}(v_i) = \mathbf{b}(v_i)$. By Def.7.1.i we have

$$[[v_i]]^{M,\mathbf{a}} = \mathbf{a}(v_i) \text{ and } [[v_i]]^{M,\mathbf{b}} = \mathbf{b}(v_i).$$

So we get: $[[A]]^{M,\mathbf{a}} = [[v_i]]^{M,\mathbf{a}} = \mathbf{a}(v_i) = \mathbf{b}(v_i) = [[v_i]]^{M,\mathbf{b}} = [[A]]^{M,\mathbf{b}}$.

(ii) Suppose that A is a complex term of L. Then, according to Def.3.1.ii, A is of the form $f^n{}_i(t_1, ..., t_n)$. Suppose that A is of this form and that $t_1, ..., t_n$ have (**). Again choose X, **a** and **b** as under (i). Since X contains all the free variables of A, X contains all the free variables of t_j, for j = 1,..., n. So since the t_j all have (**), and X, **a** and **b** fulfill together with t_j the conditons of (**), we have

(1) $[[t_j]]^{M,\mathbf{a}} = [[t_j]]^{M,\mathbf{b}}$ (for t_j = 1,..., n)

According to Def. 7.1.ii we have:

(2) $[[f^n{}_i(t_1, ..., t_n)]]^{M,\mathbf{a}} = F_M(f^n{}_i)([[t_1]]^{M,\mathbf{a}}, ..., [[t_n]]^{M,\mathbf{a}})$

Because of (1) the right hand side of (2) equals $F_M(f^n{}_i)([[t_1]]^{M,\mathbf{b}}, ..., [[t_n]]^{M,\mathbf{b}})$ and this is, by Def, 7.1.ii, the same as $[[f^n{}_i(t_1, ..., t_n)]]^{M,\mathbf{b}}$.

(iii) According to Def. 3.2.i the atomic formulas of L come in two varieties: (a) $P^n{}_i(t_1, ..., t_n)$ and (b) $t_1 = t_2$.

Suppose A is of the form $P^n{}_i(t_1, ..., t_n)$ and that (**) holds for $t_1, ..., t_n$. Then we get, just as in case (ii):

$[[P^n{}_i(t_1, ..., t_n)]]^{M,\mathbf{a}} = F_M(P^n{}_i)([[t_1]]^{M,\mathbf{a}}, ..., [[t_n]]^{M,\mathbf{a}}) =$
$F_M(P^n{}_i)([[t_1]]^{M,\mathbf{b}}, ..., [[t_n]]^{M,\mathbf{b}}) = [[P^n{}_i(t_1, ..., t_n)]]^{M,\mathbf{b}}$.

The case where A has the form $t_1 = t_2$ is just like the last one and is left to the reader.

(iv) A is a complex formula. Here there are quite a few possibilities for the form of A: A could be: a negation $\neg$ B, a conjunction B & C, a disjunction, an implication, a biconditional, a universally quantified formula or an existentially quantified formula. We consider three of these possibilities: (a) A is of the form $\neg$ B, (b) A is of the form B & C and (c) A is of the form $(\forall v_i)B$.

(a) Suppose that A is of the form $\neg$ B and that B has (**). Let X, **a**, **b** be chosen so that A, X, **a**, **b** satisfy the conditions of (**). Since the free variables of A are the same as the free variables of B, the conditions of (**) are also satisfied by B, X, **a**, **b**. So since B has (**), $[[B]]^{M,\mathbf{a}} = [[B]]^{M,\mathbf{b}}$. So we have, using Def. 7.2.ii:

$[[A]]^{M,\mathbf{a}} = 1$ iff $[[\neg B]^{M,\mathbf{a}} = 1$ iff $[[B]^{M,\mathbf{a}} = 0$ iff $[[B]^{M,\mathbf{b}} = 0$ iff $[[\neg B]^{M,\mathbf{b}} = 1$ iff $[[A]]^{M,\mathbf{b}} = 1$.

(b) Suppose that A is of the form B & C and that B and C both have (**). Again, let X, **a**, **b** be chosen so that A, X, **a**, **b** satisfy the conditions of (**). The free variables of B are among the free variables of A and thus included in X; and the same holds for C. So since b and C have (**), we have

(3) $[[B]]^{M,\mathbf{a}} = [[B]]^{M,\mathbf{b}}$ and $[[C]]^{M,\mathbf{a}} = [[C]]^{M,\mathbf{b}}$.

So we have:

$[[A]]^{M,\mathbf{a}} = 1$ iff $[[B \& C]^{M,\mathbf{a}} = 1$ iff $[[B]^{M,\mathbf{a}} = 1$ and $[[C]^{M,\mathbf{a}} = 1$ iff $[[B]^{M,\mathbf{b}} = 1$ and $[[C]^{M,\mathbf{b}} = 1$ iff $[[B \& C]^{M,\mathbf{b}} = 1$ iff $[[A]]^{M,\mathbf{b}} = 1$.

(c) Suppose that A is the formula $(\forall v_i)B$ and that B has (**). Again, let X, **a**, **b** be chosen so that A, X, **a**, **b** satisfy the conditions of (**). Suppose that y is a free variable of B. Then either y is the variable v_i or else y is a free variable of A and thus $y \in X$. So in either case $y \in X \cup \{v_i\}$. According to Def. 7.2.vii,

$[[A]]^{M,\mathbf{a}} = 1$ iff $[[(\forall v_i)B]]^{M,\mathbf{a}} = 1$ iff for all $u \in U_M$ $[[(B]]^{M,\mathbf{a}[u/v_i]} = 1$.

We observe the following:

(4) B, $X \cup \{v_i\}$ and the assignments $\mathbf{a}[u/v_i]$ and $\mathbf{b}[u/v_i]$ satisfy the conditions of (**)

To show (4) we first recall that $X \cup \{v_i\}$ contains all the free variables of B. Secondly we show that for any free variable y of B:
$\mathbf{a}[u/v_i](y) = \mathbf{b}[u/v_i](y)$. Recall that there are two possibilities for y. either $y = v_i$ or ($y \neq v_i$ and $y \varepsilon X$). In the first case we have:

$\mathbf{a}[u/v_i](y) = \mathbf{a}[u/v_i](v_i) = u = \mathbf{b}[u/v_i](v_i) = \mathbf{b}[u/v_i](y)$.

In the second case, since $y \neq v_i$, $\mathbf{a}[u/v_i](y) = \mathbf{a}(y)$ and $\mathbf{b}[u/v_i](y) = \mathbf{b}(y)$. Also, since **a** and **b** coincide on the variables in X and $y \varepsilon X$, $\mathbf{a}(y) = \mathbf{b}(y)$.
So we have: $\mathbf{a}[u/v_i](y) = \mathbf{a}(y) = \mathbf{b}(y) = \mathbf{b}[u/v_i](y)$. This concludes the proof of (4).

We are now in a position to complete the proof of (iv.c).

$[[A]]^{M,\mathbf{a}} = 1$ iff $[[(\forall v_i)B]]^{M,\mathbf{a}} = 1$ iff for all $u \varepsilon U_M$ $[[(B]]^{M,\mathbf{a}[u/v_i]} = 1$ iff (using (4) and the fact that B has (**)) for all $u \varepsilon U_M$ $[[(B]]^{M,\mathbf{b}[u/v_i]} = 1$ iff $[[(\forall v_i)B]]^{M,\mathbf{b}} = 1$ iff $[[A]]^{M,\mathbf{b}} = 1$.

The proofs of the other cases under (iv) are trivial variants of the proofs of cases (a), (b) and (c).

This completes the proof of Lemma 1. q.e.d.

<u>1.1.3 Interlude about Proofs by Induction</u>

It might be argued that strictly speaking the proof of Lemma 1 is not yet complete. For we are still left with the inference from all the basic and recursive steps of the proof to the conclusion that (**) is true of all terms and all formulas of L. This last step is normally left out in inductive proofs because it always rests on the same general principle. The principle is easiest to explain in connection with induction on the natural numbers (which incidentally is also the form of induction that tends to be familiar to non-mathematicians A well-known example of a proof by induction that all natural numbers have a certain property P is that where P is the property which the number n has if the sum of the numbers from 0 to n is equal to 1/2(n.(n+1):

(5) $$\sum_{i=0}^{n} i = 1/2(n.(n+1)$$

The typical way to prove this is to argue as follows:

(i) The statement (5) holds for n = 0. For in that case both sides are equal to 0.

(ii) Suppose that the statement (5) holds for n = k. Then (5) also holds for n = k+1. For

$$\sum_{i=0}^{k+1} i = \sum_{i=0}^{k} i + (k+1) = 1/2(k.(k+1)) + (k+1) =$$

$$1/2((k.(k+1)) + 2.(k+1)) = 1/2(k^2 + 3k + 2) = 1/2(k+1)(k+2)$$

From (i) and (ii) we can infer that (#) holds for all n. Why? Well, one way to argue is as follows: (i) shows that (#) holds for the first natural number 0. Combining this information with (ii) leads to the conclusion that (#) holds for 1. Combining that information with (ii) we conclude that (#) holds for 2; and so on.

We can also turn this argument upside down: Suppose that (#) does not hold for all natural numbers n. Then there must be a smallest number n_o for which (#) fails. Because of (i), n_o must be different from 0. So there must be a number m such that $n_o = m+1$. But then, since n_o is the smallest number for which (#) does not hold, (#) holds for m. So by (ii) it must hold for m+1, that is for n_o: contradiction. So we conclude that (#) holds for all n.

The case of our proof of Lemma 1 is somewhat more complex, but it is in essence like the one just considered. In the case of Lemma 1 the task is to show that all terms and formulas of L satisfy a certain condition (our condition (**)). That the basic and inductive clauses (of which we proved a representative selection) together entail that all terms and formulas A have (**) can be argued along similar lines as as we followe in proving (5). Suppose that there was a term or formula A for which (**) does not hold. Then among those terms and/or formulas there must be at least one that is minimal w.r.t. (**), i. e. a term or formula A_o which itself does not have (**) but which is such that all its immediate constituent terms or formulas have (**). But then we get a contradiction, just as in the natural number case: A_o can't be an atomic term, for that would contradict the base case of the proof. So A_o must have immediate constituents, all of which do have (**). But then we have a contradiction with that part of the proof which concerns the particular form of A_o.

A more abstract way of stating the validity of the method of proof by induction is this: Suppose that Y is a set of objects and that there is a partial ordering < of Y which is *well-founded*, i.e. which has the property that if Z is a non-empty subset of Y, then Z must contain at least one <-*minimal* element; that is, there must be at least one element z of Z such that for all y ε Y such that y < z, it is the case that y ε Y\Z. To establish that all members of Y have a certain property P it is then enough to show the following:

(6) Let z ε Y and suppose that for all y < z, P(y). Then P(z).

It is easy to see that the binary relation which holds between two between terms and/or formulas A and B of L iff A is a constituent of B is a well-founded partial ordering of the set of all tems and formulas of L. So what our proof of Lemma 1 amounts to is that (6) holds for the case where < is the constituent relation between terms and formulas of L and P is the property (**).

1.1.4 Continuation of 1,1,2

The most important consequence of Lemma 1 is that the values of closed terms and closed formulas (i.e. sentences) are independent of the assignment.

Def.8 A sentence A of L is said to be *true in* a model M iff for all assignments a in M, $[[A]]^{M,a} = 1$.

Notation. It follows from Lemma 1 that when A is a sentence, then for all assignments **a** and **b**, $[[A]]^{M,\mathbf{a}} = [[A]]^{M,\mathbf{b}}$. So in this case we may, without risk of confusion, suppress mention of the assignment. We will often do this and write " M ⊧ A" instead of "$[[A]]^{M,\mathbf{a}} = 1$ for some **a**". More generally, when the free variables of A are among v_1, ..., v_k, and a_1, ..., a_k are elements of the model M, we will write "M ⊧ A[a_1, ...,a_k]" in stead of " $[[A]]^{M,\mathbf{a}} = 1$ for some assignment **a** in M such that $\mathbf{a}(v_i) = a_i$ for i = 1, ..., k". Again the intuitive justification is given by Lemma 1, which guarantees that if A is as described and **a** and **b** are assignments which both assign a_1, ..., a_k to v_1, ..., v_k, then $[[A]]^{M,\mathbf{a}} = [[A]]^{M,\mathbf{b}}$.

Even more generally than this, in a case where the free variables of A have been specified as x_1, ..., x_n, (where the x_i may be any variables from the list v_1, v_2 , ... of all variables of first order logic) we will sometimes write " $M \models A[a_1, ..., a_k]$" in stead of " $[[A]]^{M,a} = 1$ for some assignment a in M such that $a(x_i) = a_i$ for i = 1, ..., n".

<u>Def.9</u>

1. A set of sentences Γ of a language L *semantically entails* a sentence A of L (or: A *is a (logical/semantic) consequence of* Γ; in symbols: $\mathbf{\Gamma \models A}$) iff for every model M for L the following is true:

   If every member B of Γ is true in M, then A is true in M.

   More generally, a set of formulas Γ of L (*semantically*) *entails* a formula A iff for every model M for 1 and every assignment **a** in L, if for all sentences B in Γ $[[(B]]^{M,\mathbf{a}} = 1$, then $[[A]]^{M,\mathbf{a}} = 1$.

2. A formula A is *valid* iff $\varnothing \models A$.

N. B. According to Def. 9.2 a formula A of L is valid iff for every model M for L and every assignment a in L, $[[A]]^{M,\mathbf{a}} = 1$.

Exercise: Show this!

The following Lemma 3 states an important relation between the value of a term t or formula B with free occurrences of a certain variable v_i and the value of the result of substituting a term t' for the free occurrences of v_i in t or B. In order to formulate the second part of the Lemma we need a further definition.

<u>Def. 10.</u> (i) Let B be a formula, α some particular free occurrence of the variable v_i in B and let t be some term. Then α is said to be *free for* t *in* B iff no variable occurring in t becomes bound in B when t is substituted for α in B.

(ii) Let B, t be as under (i). Then the variable v_i is said to be *free for* t *in* B iff every free occurrence of v_i in B is free for t in B.

Lemma 2 (i) Let t, t' be any terms of L, let M be any model for L and **a** an assignment in M. Then:

$$[[\ t[t'/v_i]\ ]]^{M,\mathbf{a}} = [[t]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i\ ]}$$

(ii) Let B be a formula of L, let M, t' and **a** be as under (i) and suppose that v_i is free for t' in B. Then

$$[[\ B[t'/v_i]\ ]]^{M,\mathbf{a}} = [[\ B\ ]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i\ ]}$$

Proof. We first prove (i) by induction on the complexity of terms.

(a). Let t be a variable v_j. First suppose that j = i. Then $t[t'/v_i] = v_i[t'/v_i] = t'$. So we have:

$[[\ t[t'/v_i]\ ]]^{M,\mathbf{a}} = [[t']]^{M,\mathbf{a}} = [[v_i]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i\ ]}$.

Now suppose that $j \neq i$. Then $t[t'/v_i] = v_j[t'/v_i] = v_j$. So

$[[\ t[t'/v_i]\ ]]^{M,\mathbf{a}} = [[v_j]]^{M,\mathbf{a}} = \mathbf{a}(v_j)$. Moreover, if $j \neq i$, then

$\mathbf{a}(v_j) = (\mathbf{a}[[[t']]^{M,\mathbf{a}}/v_i])(v_j)$. So

$[[\ t[t'/v_i]\ ]]^{M,\mathbf{a}} = [[v_j]]^{M,\mathbf{a}} = [[v_j]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i\ ]} =$
$[[t]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i\ ]}$

(b) Suppose that t is the term $g(t,...,t_n)$ and suppose that for k = 1,..., n. (i) holds with t_k instead of t,. It is easily seen that $(g(t_1, ..., t_n))[t'/v_i] = g(t_1[t'/v_i], ..., t_n[t'/v_i])$. So

$[[\ t[t'/v_i]\ ]]^{M,\mathbf{a}} = [[\ (g(t1, ..., tn))[t'/v_i]\ ]]^{M,\mathbf{a}} =$
$[[\ g(t_1[t'/v], ..., t_1[t'/vi])]]^{M,\mathbf{a}} =$
$(F_M(g))([[t_1[t'/v_i]]]^{M,\mathbf{a}}, .., [[t_n[t'/v_i]]]^{M,\mathbf{a}}) =$
$(F_M(g))([[t_1]]^{M,\mathbf{a}'}, .., [[t_n]]^{M,\mathbf{a}'})$,
where $\mathbf{a}'$ is the assignment $\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i]$. But
$(F_M(g))([[t_1]]^{M,\mathbf{a}'}, .., [[t_n]]^{M,\mathbf{a}'}) =$
$[[g(t_1, ..., t_n)]]^{M,\mathbf{a}'}$.
This concludes the proof of (i)

We now prove (ii) by induction on the complexity of formulas.

(a) Let B be the formula $P((t_1, ..., t_n)$. We proceed essentially as under (i.b):
$[[\ B[t'/v_i]\ ]]^{M,\mathbf{a}} = [[\ (P(t_1, ..., t_n))[t'/v_i]\ ]]^{M,\mathbf{a}} =$
$[[\ P(t_1[t'/v], ..., t_1[t'/vi])]]^{M,\mathbf{a}} =$
$(F_M(P))([[t_1[t'/v_i]]]^{M,\mathbf{a}}, .., [[t_n[t'/v_i]]]^{M,\mathbf{a}}) =$
$(F_M(P))([[t_1]]^{M,\mathbf{a}'}, .., [[t_n]]^{M,\mathbf{a}'}) =$
$[[\ (P(t1, ..., tn))]]^{M,\mathbf{a}'}$, where **a**' is as above.

(b) Suppose that B is a formula whose main operator is a sentence connective. We consider just one case, that where B is a negation, i.e. B = $\neg$C for some C. We assume that (ii) holds for C. Clearly we have that $B[t'/v_i] = (\neg C)[t'/v_i] = \neg(C[t'/v_i])$. So $[[\ B[t'/v_i]\ ]]^{M,\mathbf{a}} = 1$ iff $[[\ \neg(C[t'/v_i])]]^{M,\mathbf{a}} = 1$ iff $[[\ C[t'/v_i]\ ]]^{M,\mathbf{a}} = 0$ iff (by the induction assumption) $[[C]]^{M,\mathbf{a}[[[t']]^{M,\mathbf{a}}/v_i]} = 0$ iff $[[\neg C]]^{M,\mathbf{a}[[[t']]^{M,\mathbf{a}}/v_i]} = 1$ iff $[[B]]^{M,\mathbf{a}[[[t']]^{M,\mathbf{a}}/v_i]} = 1$.

(c) Now suppose that B begins with a quantifier. We only consider the case where B is of the form $(\exists v_j)C$. Once more we have to distinguish between the case where j = i and that where j ≠ i. When j = i, then $((\exists v_j)C)[t'/v_i] = (\exists v_j)C$ since in that case v_i has no free occurrences in $(\exists v_j)C$. But for this very same reason we have that $[[(\exists v_j)C]]^{M,\mathbf{a}} =$ $[[(\exists v_j)C]]^{M,\mathbf{a}[[[t']]^{M,\mathbf{a}}/v_i\ ]}$ (by Lemma 1, since **a** and $\mathbf{a}[\ [[t']]^{M,\mathbf{a}}\ /v_i]$ coincide on the free variables of $(\exists v_j)C$ (because any free occurrences of v_j in C are bound by the initial quantifier $(\exists v_j)$). This concludes the argument for the case that j = i.

The second case is that where j ≠ i. This case has to be subdivided once more into two subcases, (i) v_i has no free occurrences in C and (ii) v_i has at least one free occurrence in C. In case (i) we have, as in the case already considered that $((\exists v_j)C)[t'/v_i] = (\exists v_j)C$. Again **a** and $\mathbf{a}[\ [[t']]^{M,\mathbf{a}}\ /v_i]$ coincide on the free variables of $(\exists v_j)C$, since in fact they already coincide on all the free variables of C. So the conclusion follows as above.

Now suppose that v_i has free occurrences in C. Since $j \neq i$. the freee occurrences of $_i$ in C are also free occurrences in B. By assumption v_i is free for t' in B. This means that the variable v_j cannot occur in t', for if it did, then its occurrences in t' would be bound in B (viz. by B's initial quantifier $(\exists v_j)$) when t' is substituted for the free occurrences of v_i in B.
Furthermore we observe that $((\exists v_j)C)[t'/v_i] = (\exists v_j)(C[t'/v_i])$. From the Truth Definition clause for $\exists$ we get:

$[[B[t'/v_i]\ ]]^{M,\mathbf{a}} = 1$ iff $[[\ ((\exists v_j)C)[t'/v_i]\ ]]^{M,\mathbf{a}} = 1$ iff $[[(\exists v_j)(C[t'/v_i])]]^{M,\mathbf{a}} = 1$ iff

for some $d \in U_M$ $[[C[t'/v_i]\ ]]^{M,\mathbf{a}[d/vj]} = 1$ (*)

By the induction assumption,

$[[C[t'/v_i]\ ]]^{M,\mathbf{a}[d/vj]} = [[C]]^{M,\mathbf{a}''}$,

where $\mathbf{a}''$ is the assignment $\mathbf{a}[d/v_j]\ [\ [[t']]^{M,\mathbf{a}\ [d/vj]}/v_i]$. We now make use of the fact that v_j does not occur in t'. Because of this $[[t']]^{M,\mathbf{a}\ [d/vj]} = [[t']]^{M,\mathbf{a}}$. So $\mathbf{a}'' =$ $\mathbf{a}[d/v_j]\ [\ [[t']]^{M,\mathbf{a}}\ /v_i] = \mathbf{a}[\ [[t']]^{M,\mathbf{a}}\ /v_i]\ [d/v_j]$, since the order in which the assignment changes in $\mathbf{a}$ to, respectively, v_i and v_j are carried out is immaterial. (These changes are independent from each other.) This means that we can rewrite (*) as:

for some $d \in U_M$ $[[C]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/vi][d/vj]} = 1$ (**)

By the Truth Definition clause for $\exists$ (**) is equivalent to

$[[(\exists v_j)C]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/vi]} = 1$. In other words, $[[B]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/vi]} = 1$.

Since the above transformations are all reversible, we have thus shown that

$[[\ B[t'/v_i]\ ]]^{M,\mathbf{a}} = 1$ iff $[[B]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/vi]} = 1$.

This concludes the proof for the case where B is of the form $((\exists v_j)C$, and therewith of part (ii) of Lemma 2.

q.e.d.

Below we will need in particular a special case of Lemma 2, stated in Corollary 1, in which the term t' is an individual constant c. (The proof of this special case is somewhat simpler, because there is no need to worry about proper substitution (i.e. about v_i being free in B for the term that is to be substituted for it in B); since c contains no varaibles, v_i will be free for c in b no matter what.)

Corollary 1 (i) Let t be any term of L, c any individual constant of L, M any model for L and **a** any assignment in M. Then:

$$[[\ t[c/v_i]\ ]]^{M,\mathbf{a}} = [[t]]^{M,\mathbf{a}[F_M(c)/vi]}$$

(ii) Similarly, if B is a formula of L, and M, c and **a** as under (i), then

$$[[\ B[c/v_i]\ ]]^{M,a} = [[\ B\ ]]^{M,a[F_M(c)/vi]}$$

Suppose that the free variables of the formula A of L are v_{i1},.., v_{in}, listed in some arbitrarily chosen order. Let m be a model for L. Then according to Lemma 2, any two assignments **a** and **b** which assign the same objects u_1,.., u_n of M to v_{i1},.., v_{in} will assign to A the same truth value in M. We can make this explicit by displaying the free variables of A, in the chosen order, as 'arguments' of A by including them in parentheses behind A, and then fixing the truth values of A in M by mentioning just the objects u_1,.., u_n of M that these assignments assign to the free variables v_{i1},.., v_{in}.
With these specifications A turns into the expression

$$A(v_{i1},.., v_{in})[_1,.., u_n\ ].$$

Since the iinformation encoded in this expression determines a unique truth value for A in M, we can write

$M \models A(v_{i_1},.., v_{i_n})[u_1,.., u_n]$ to indicate that the assignment of $u_1,.., u_n$ to satisfies A in M (i.e. that hte truth value of A under any such assignment is 1). This notation is quite useful in prctice and we will make use of it occasionally.

When A is a sentence, i.e. when the set of its free variables is empty, then, as Cor. 1 makes explicit, any two assignments in M will assign it the same truth value. In this case we can speak simply of 'the truth value of A in M' and of A 'being true in M' or ' being false in M'. We express this formally by writing '$M \models A$' for 'A is true in M'.

1,1,5 Axioms, Rules. Proofs and Theorems.

Def.10

1. An *axiom of* L is any formula of L that has one of the forms A1 - A13:

A1. $A \rightarrow (B \rightarrow A)$

A 2. $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

A3. $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$

A4. $(\forall v_i)(A \rightarrow B) \rightarrow (A \rightarrow (\forall v_i)B)$, provided v_i has no free occurrences in A

A5. $(\forall v_i)A \rightarrow A[t/v_i]$, provided v_i is free for t in A[2]

A6. $(A \rightarrow B) \rightarrow ((B \rightarrow A) \rightarrow (A \leftrightarrow B))$

A7. $(A \leftrightarrow B) \rightarrow (A \rightarrow B)$

A8. $(A \leftrightarrow B) \rightarrow (B \rightarrow A)$

A9. $(A \& B) \leftrightarrow \neg(A \rightarrow \neg B)$

A10 $(A \vee B) \leftrightarrow (\neg A \rightarrow B)$

A11 $(\exists v_i)A \leftrightarrow \neg(\forall v_i) \neg A$

A12. $v_i = v_i$

A13. $v_i = v_j \rightarrow (A \rightarrow A')$, where A' results from replacing one occurrence of v_i in A by v_j and the new occurrence of v_j in A' is free in A'

In the formulation of A5 there is reference to the notion of "v_i being free for t in A". Intuitively this means that t can be substituted for each of the free occurrences of in A without this leading to free variables of t (other than v_i) being captured by quantifiers in A.
To define the concept (of v_i being free for t in A) correctly, we must (a) distinguish between the different occurrences of expressions - variables, terms, subformulas, quantifiers - within a given formula B, and then (b) define the notion of the *scope of* a quantifier occurrence in B.

The notion of an occurrence in a formula presupposes that different occurrences of the same expression type - for instance, two occurrences of the variable v_1 - must be somehow distinguishable so

---

2 For the definition of "v_i is free for t in A" see Def. 10 below.

they must be indexed, or labeled, in some way. There are all sorts of ways to accomplish this, some fancy, others homely. Here we will simply assume that each formula B can be identified as a finite string of symbols, that is, as a function which maps some initial segment {1, ... , n} of the positive integers into the set of symbols of the given language L to which B belongs. In this way each of the symbol occurrences in B will be assigned an identifying integer, and each larger constituent of B can be identified with the subset of {1, ... , n} which consists of those integers that are associated with the symbol occurrences in B that belong to that constituent. Among other things, identification of the different symbol occurrences in B enables us to refer to individual quantifier occurrences, i.e. particular occurrences of the symbol strings "$(\forall v_i)$" and "$(\exists v_i)$".

The definition of the notions *free* and *bound* rests on the fact that the well-formed expressions (terms and formulas) of predicate logic are *syntactically unambiguous*: For each symbol string that is syntactically well-formed (that is, each string that can be derived as an expression of a language L by using the clauses of Def. 3.1 und 3.2) there is *only one* syntactic analysis - only one way in which these clauses can be applied to put the string together. (Strictly speaking this is a property of Def. 3 that can and ought to be proved. But the prooof is rather tedious and has been omitted here.)

It is a familiar feature of definitions of syntactic structure (or "grammars", as they are usually called, when the language in question is a natural language) that expressions which are well-formed according to these definitions have syntactic analyses (by virtue of the given definition) that can be represented in the form of a tree. In the case of formal languages (though not as a rule for natural languages) the analysis of any well-formed expression will as a rule be unique.

<u>Exercise</u>:

Construct syntactic derivation trees for the formulas:

(a) $(\exists v_1)((\exists v_1)P(v_1) \rightarrow P(v_1))$;
(b) $((\exists v_1)P(v_1)\ \&\ Q(v_1)) \rightarrow (\exists v_1)P(v_1)\ \&\ (\exists v_1)Q(v_1))$;
(c) $(\forall v_1)(\forall v_2)(\forall v_3)((R(v_1,v_2) \leftrightarrow (R(v_2,v_3) \leftrightarrow R(v_1,v_3)) \leftrightarrow ((R(v_1,v_2) \leftrightarrow R(v_2,v_3)) \leftrightarrow R(v_1,v_3))$.

Let Q be an occurrence in B of the existential quantifier expression "$(\exists v_j)$" (the scope of an occurrence of a universal quantifier expression is defined in the same way.). Then the *scope of* Q *in* B is that formula occurrence A such that the transition from A to the string QA (using clause (iii) of the definition of well-formedness) is part of the unique parse of B.

We can now define (i) what it is for a term t to be *free for* a particular free occurrence *v* of the variable v_i *in* the formula B, and (ii) what it is for t to be *free for* v_i *in* B:

<u>Def. 10</u>:

(i) t is *free for v in* B iff t contains no variable v_j such that *v* belongs to the scope of any occurrence of either "$(\exists v_j)$" or "$(\forall v_j)$" in B;

(ii) t is *free for* the variable v_i *in* B iff t is free in B for all free occurrences in B of v_i.

2. The *Inference Rules (of* L) are given by the following two schemata:

$$\text{(i)}\quad \frac{\vdash A \qquad \vdash A \rightarrow B}{\vdash B} \qquad\qquad \frac{\vdash A}{\vdash (\forall v_i)A}$$

(Modus Ponens) (Universal Generalization)

3. A *proof in* L *of* a formula A of L *from* a set of formulas Γ of L is a sequence $A_1, \ldots, A_n$ of formulas of L such that

1. $A_n = A$, and
2. for each A_i with $i \leq n$ either:
   - (i) A_i is an axiom of L, or
   - (ii) $A_i \in \Gamma$, or
   - (iii) there are j, k < i such that $A_k = A_j \rightarrow A_i$, or
   - (iv) $A_i = (\forall v_m)B$, there is a $j < i$ such that $A_j = B$ and v_m does not occur free in any member of Γ which occurs as a line A_r with $r \leq j$.

We write: $\boldsymbol{\Gamma \vdash_L A}$ for "there exists a proof in L of A from Γ".

<u>Lemma 3</u>: Supppose that $L \subseteq L'$ (i.e. the function L' extends the f unction L; in other words, that each non-logical constant of L is also a non-logical constant of L' and with the same signature), that A is a formula of L and Γ a set of formulas of L and that there is a proof of A from Γ in L'. Then there is a proof of A from Γ in L.

<u>Proof.</u> Suppose that A is a sentence of L and Γ a set of sentences of L and that P is a proof of A from Γ in some language L'. Take some fixed sentence B of L, e.g. $(\forall v_1)\ v_1 = v_1$, and replace every atomic formula occurring in P which contains a non-logical constant that belongs to L' but not to L by the sentence B. It is easily verified that the sequence of formulas P' into which P is converted by these transformations is a proof of A from Γ in L. q.e.d.

Lemma 2 justifies dropping the subscript "$_L$" from the expression "$\Gamma \vdash_L A$". So henceforth we will write simply "$\Gamma \vdash A$" to express that there exists a proof of A from Γ.

The central theoretical result about first order predicate logic is that semantic consequence can be captured by a notion of provability such as the one defined here. (This is one of several fundamental results that logic owes to the greatest logician of the 20-th century, the Czech-Austrian mathematician Kurt Gödel). The equivalence has two sides, usually referred to as the *soundness* and the *completeness* (of the concept of proof in question):

<u>1.2 Soundness and Completeness of the axiomatic proof system of Section 1.1.3</u>

<u>Theorem 1 (Soundness)</u>: If $\Gamma \vdash A$, then $\Gamma \models A$

<u>Theorem 2. (Completeness)</u>: If $\Gamma \models A$, then $\Gamma \vdash A$

<u>Proof of Theorem 1.</u> Soundness is proved by showing:

(i) every formula B which has the form of one of the axioms has the property (*)

(*) for any model M for L and any assignment a in M, $[[B]]^{M,a} = 1$

and

(ii) if P is a proof of A from Γ, then all lines A_i of P have the following property (**):

(**) if M is a model, then for every assignment a in M such that $[[B]]^{M,a} = 1$ for all B ε Γ which occur as a line A_r in P with $r \leq i$, then $[[A_i]]^{M,a} = 1$.

The proof of (i) is straightforward for all axioms other than A4 and A5. An exact proof of (*) for formulas of the form of A4 requires Lemma 1 the proof for formulas of of the form of A5 requires Lemma 2.

Exercise: Show the validity (i.e. condition (*) above) for each of the Axioms A1 - A13.
(Hint: Use Lemma 1 in the proof for A4 and Lemma 3 in the proofs for A5.)

Proof of (**): The proof of (**) is by induction on the length of the proof. More precisely, fix L, Γ and M and suppose that (**) holds for all proofs from Γ of length < n. We then have to show that (**) also holds for proofs of length n.

Let P be a proof $<C_1, ..., C_{n-1}, C_n>$ be a proof from Γ of length n. Let a be any assignment in M and assume that for all lines C_j in P which belong to Γ, $[[C_j]]^{M,a} = 1$.

There are four possibilities for C_n:

(i) C_n is an instance of one of the axioms A1 - A13;
(ii) C_n ε Γ;
(iii) C_n comes by Modus Ponens from earlier lines C_j and C_k (where C_k is the formula $C_j \rightarrow C_n$);
(iv) C_n comes by Universal Generalisation from an earlier line C_j; in this case C_n will be of the form $(\forall v_i)A$, whereas C_j is the formula A.

The only interesting case of the proof is (iv), which the one we consider.

We must show that $[[C_n]]^{M,a} = [[(\forall v_i)A]]^{M,a} = 1$. To this end we must show that $[[A]]^{M,a[u/v_i]} = 1$ for every $u \in U_M$. Let $u \in U_M$. Because of the constraint on the application of UG we know that for every C_k preceding C_j in P which is a member of Γ, v_i does not occur free in C_k. Since by assumption $[[C_k]]^{M,a} = 1$ for each of these C_k, we conclude by Lemma 2 that $[[C_k]]^{M,a[u/v_i]} = 1$. By assumption the induction hypothesis (**) holds for C_j (since C_j belongs to a proof from Γ of length < n). So $[[C_j]]^{M,a[u/v_i]} = [[A]]^{M,a[u/v_i]} =1$. Since this holds for all $u \in U_M$, $[[(\forall v_i)A]]^{M,a} = 1$.

1.2.1 Proof of the Completeness Theorem.

Proof of Theorem 2. Proving completeness is a good deal more involved than proving soundness. The proof relies among other things on showing that for certain consequence relations - i.e. relations of the form "$\Gamma \models A$" for certain formulas A and formula sets Γ - there exists a proof of A from Γ using our axioms and rules. To build up the needed stock of such results it is necessary to proceed in the right order. Here follows a sequence of useful results about provability which (with the exception of T2) can be established without too much difficulty so long as one proceeds the indicated order. It will be useful to distinguish between provability simpliciter and provability without use of the rule UG (Universal genralisation). Provability in this latter, restricted sense we indicate by " $\vdash'$ ". Thus " $\Gamma \vdash' B$" means that there is a proof of B from Γ in which UG is not used.

T1. $\vdash' A \rightarrow A$

T2. For all formulas A, B and sets of formulas Γ,
$\Gamma \vdash' A \rightarrow B$ iff $\Gamma \cup \{A\} \vdash' B$

T3. $\vdash' (A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$

T4. $\vdash' (A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$

T5. If $\Gamma \vdash' A$ and $\Delta \cup \{A\} \vdash' B$, then $\Gamma \cup \Delta \vdash' B$

T6. $\neg B \rightarrow \neg (A \rightarrow A) \vdash' B$

We abbreviate the formula $\neg (A \rightarrow A)$ as $\perp_A$. In the following it will also be useful to have a name for one particular formula of this form, in which A is some single sentence. The sentence chosen involves only logical vocabulary and thus belongs to every first order language. So we let $\perp$ be short for the following formula:

(Def. $\perp$) $\neg((\forall v_1)(v_1 = v_1) \rightarrow (\forall v_1)(v_1 = v_1))$.

T7. $\vdash' \perp \rightarrow B$

T8. $\neg\perp \rightarrow \perp \vdash' \perp$

T9. $\neg\perp \rightarrow \neg\neg\perp \vdash' \perp$

T10. $\neg\neg\perp \vdash' \perp$

T11. $\vdash' \neg\perp$

T12. $B, \neg B \vdash' \perp$

T13. $\neg\neg B \vdash' B$

T14. $\neg\neg\neg B \vdash' \neg B$

T15. $B \vdash' \neg\neg B$

T16. $\neg B \rightarrow A, \neg A \vdash' B$

T17. $B \rightarrow A \vdash' \neg A \rightarrow \neg B$

T18. $\Gamma \vdash' B$ iff $\Gamma \cup \{\neg B\} \vdash' \perp$

T19. $\neg B \rightarrow B \vdash' B$

T20 $\Gamma \cup \{A\} \vdash' B$ and $\Gamma \cup \{\neg A\} \vdash' B$ iff $\Gamma \vdash' B$

T21. $\vdash' (\forall v_i)(A \rightarrow B) \rightarrow ((\forall v_i) A \rightarrow (\forall v_i) B)$

T22. $\vdash' B \rightarrow (\forall v_i) B$, provided v_i does not occur free in B

T23. $\vdash' (\forall v_i) B \rightarrow (\forall v_k) B[v_k/v_i]$,
provided v_k does not occur free in B and every occurrence of v_k in $B[v_k/v_i]$ which is not an occurrence of v_k in B is free in $B[v_k/v_i]$.

T24. $\vdash' [B]\, {}^{t}/v_i \rightarrow (\exists v_i)B$

T25. $\vdash' t = t' \rightarrow t' = t$, provided t is free for v_i in B

T26. $\vdash' (t = t' \,\&\, t' = t'') \rightarrow t = t''$

T27. $\vdash (\forall v_i)(A \rightarrow B) \rightarrow ((\exists v_i) A \rightarrow (\exists v_i) B)$

T28. $\vdash ((\exists v_i) A \rightarrow A$, provided v_i does not occur free in A.

T29. $\vdash (\exists v_i) A \rightarrow (\exists v_k) A[v_k/v_i]$, provided v_k is free for v_i in A.

T30. $\vdash (\exists v_i)\, t = v_i$, provided v_i does not occur in t.

T31. For all sentences A, formulas B and sets of sentences Γ,
$\Gamma \vdash A \rightarrow B$ iff $\Gamma \cup \{A\} \vdash B$

T32. $\neg A \rightarrow \perp \vdash A$

T33. $\vdash' A \leftrightarrow A$

T34. If $\vdash' A \leftrightarrow A'$ and $\vdash' B \leftrightarrow B'$, then $\vdash' (A \,\&\, B) \leftrightarrow (A' \,\&\, B')$, $\vdash' (A \vee B) \leftrightarrow (A' \vee B')$, $\vdash' (A \rightarrow B) \leftrightarrow (A' \rightarrow B')$, $\vdash' (A \leftrightarrow B) \leftrightarrow (A' \leftrightarrow B')$

The theorems T1-T31 have been arranged so that the earlier ones may be used in the proofs of later ones. (Though some other orderings would work just as well.) We leave the proofs as exercises in all cases except for those of T2 and T31.

Proof of T2:

$\Rightarrow$ Suppose that P is a proof of $A \rightarrow B$ from Γ. Append to P the new lines: (i) A and (ii) B. The first of these is justified as a member of the premise set $\Gamma \cup \{A\}$, the second as an application of M.P. Thus this extension will be a proof of B from $\Gamma \cup \{A\}$.

$\Leftarrow$. Suppose $P = \langle C_1, ..., C_n \rangle$ is a proof of B from $\Gamma \cup \{A\}$ in which there are no applications of UG. Note that for each $i < n$, the initial segment $\langle C_1, ..., C_i \rangle$ is a proof (without UG) of C_i from $\Gamma \cup \{A\}$.
We transform P into a proof $\langle D_1, ..., D_{f(n)} \rangle$ of $A \rightarrow B$ from Γ in which for each line C_i of P there is a corresponding line $D_{f(i)}$ of the form $A \rightarrow C_i$. (f is a monotone increasing function from $\{1, ..., n\}$ into $\{1, ..., f(n)\}$.) We do this by (i) constructing a proof P_1 of $A \rightarrow C_1$ from Γ, and (ii) extending successively for $i = 1, ..., n-1$ the already obtained proof P_i of $A \rightarrow C_i$ from Γ to a proof P_{i+1} of $A \rightarrow C_{i+1}$ from Γ.

(i) In this case the proof $\langle C_1 \rangle$ consists of the single line C_1. There are three possibilities regarding C_1:

(i) C_1 is the formula A;
(ii) C_1 is an axiom;
(iii) C_1 is a member of Γ.

In case (i) we take for P_1 a proof of $A \rightarrow A$ from the empty premise set (see T1).

In cases (ii) and (iii) we take for P_1 the three lines:

(1)	C_1	(Axiom or member of Γ)
(2)	$C_1 \rightarrow (A \rightarrow C_1)$	(Axiom A1)
(3)	$A \rightarrow C_1$	(MP from lines (1) and (2))

Clearly this is a proof of $A \rightarrow C_1$ from Γ.

Now suppose that $1 \leq i < n$ and that a proof $P_i = \langle D_1, ..., D_{f(i)} \rangle$ of $A \rightarrow C_i$ from Γ with the desired properties has already been constructed. For the line C_{i+1} of $\langle C_1, ..., C_n \rangle$ there are the following possibilities:

(i) C_1 is the formula A;
(ii) C_1 is an axiom;
(iii) C_1 is a member of Γ;
(iv) there are j, k < i such that $C_k = C_j \rightarrow C_{i+1}$.

In cases (i) - (iii) we cosntruct P_{i+1} by appending to P_i the proof P_1 which we constucted for these respective cases under (1). It is clear that in each of these cases this does give us a proof of the intended kind. For the remaining case (iv) we extend with the following lines:

((f(n) + 1)	$(A \rightarrow (C_j \rightarrow C_{i+1})) \rightarrow ((A \rightarrow C_j) \rightarrow (A \rightarrow C_{i+1}))$	(Axiom A2)
((f(n) + 2)	$((A \rightarrow C_j) \rightarrow (A \rightarrow C_{i+1}))$	(MP, from lines f(k), ((f(n) + 1))
((f(n) + 3)	$(A \rightarrow C_{i+1})$	(MP, from lines f(j), ((f(n) + 2))

In this manner we obtain eventually a proof of $A \rightarrow C_n$ from Γ.
This concludes the proof of T2. q.e.d.

T2 is a special case of the more general equivalence:

(*) $\Gamma \vdash A \rightarrow B$ iff $\Gamma \cup \{A\} \vdash B$

The proof of this equivalence is considerably more complicated than the one just given. Since our immediate need is in connection with the "propositional calculus" theorems T3-T20, T25, T26, all of which can be proved without the use of UG, the more restricted version T2 suffices. In the central part of the Completeness Proof we will need another special case of (*), in which A, B and the members of Γ are sentences. In the above lits this is T31, hte proof of which follows presently.

In its full generality the equivalence (*) will follow as a corollary to the Completeness Theorem, given that the semantic equivalent (**) of (*) holds:

(**) $\Gamma \models A \rightarrow B$ iff $\Gamma \cup \{A\} \models B$

That (**) does hold is easily shown. (Exercise: Prove this!)
A proof of (*) along the lines of the proof of T2 is given in the Appendix.

<u>Proof of T31.</u>

$\Rightarrow$ As in the proof of T2.

$\Leftarrow$ Again we assume that there is a proof P = < C_1, ..., C_n> is a proof of B from $\Gamma \cup \{A\}$ and construct for i = 1,..., n proofs P_i of $A \rightarrow C_i$ from Γ. The construction of P_1 is as in the proof of T2, and the extension of P_i to P_{i+1} is also as in the earlier proof for the four cases considered there. The one additional case that is to be considered now is that where C_{i+1} is the result of an application of UG. In that case C_{i+1} has the form $(\forall v_j)D$ for some j while there exists a k < i+1 such that C_k is D. We P_i with the lines

(f(i) + 1)	$(\forall v_j)(A \rightarrow D)$	(UG)
(f(i) + 2)	$(\forall v_j)(A \rightarrow D) \rightarrow (A \rightarrow (\forall v_j) D)$	(A4)
(f(i) + 3)	$A \rightarrow (\forall v_j D$	(MP, from (f(i) +1), (f(i) + 2))

Note that the application of UG in line (f(i) + 1) is unproblematic since all members of Γ are sentences. Moreover, since A is a sentence, and thus has no free occurrences of v_j, (f(i) + 2) is a correct instance of A4.

Would that this were all the equipment we need for the proof of the Completeness Theorem. But alas, it appears that there is one further property of our axiomatic deduction system that we must verify in order to be able to carry through the construction that the completenes proof involves. This is the property that our deduction system enables us to prove the equivalence of *alphabetic variants*. Roughly speaking ,two formulas are alphabetic variants of each other if they differ only in that one can be obtained from the other merely by "renaming bound variables". It is a well-known and intuitively obvious fact that if this is the only difference between two formulas, then they are logically equivalent. The "name" of a bound variable doesn't matter; or, more correctly put, which variable symbol we use to play the role of a particular bound varuiable in a formula makes no difference to the semantics and logic of the formula. For instance, the sentences

$$(\forall v_1)(\exists v_2)(P(v_1,v_2) \;\&\; P(v_2,v_1)),$$
$$(\forall v_1)(\exists v_3)(P(v_1,v_3) \;\&\; P(v_3,v_1))$$

are alphabetic variants; and so are the free variable formulas

$$(\forall v_1)(\exists v_2)(Q(v_1,v_2,v_4) \;\&\; Q(v_2,v_1,v_4)),$$
$$(\forall v_1)(\exists v_3)(Q(v_1,v_3,v_4) \;\&\; Q(v_3,v_1,v_4)).$$

But we have to be careful about unwanted variable bindings. For instance, the formulas

$$(\forall v_1)(\exists v_2)(Q(v_1,v_2,v_4) \;\&\; Q(v_2,v_1,v_4)),$$
$$(\forall v_1)(\exists v_4)Q(v_1,v_4,v_4) \;\&\; Q(v_4,v_1,v_4))$$

are not alphabetic variants, as the occurrences of v_4 that are free in the first formula are bound by the quantifier $(\exists v_4)$ in the second. This means that we have to be careful to define the relation of alphabetic variance in such a way that such cases are excluded. The best way to accoplish this is by defining the relation inductively on the complexity of formulas.

<u>Def. 10'</u> (alphabetic variants)

(i) Suppose A is atomic. Then A' is an *alphabetic variant of* A iff A' = A.

(ii) Suppose that A' is an alphabetic variant of A and B' is an alphabetic variant of B. Then $\neg$A' is an *alphabetic variant of* $\neg$A, (A' & B') is an *alphabetic variant of* (A & B), and likewise for the other connectives

(iii) Suppose that A' is an alphabetic variant of A and that v_i, v_j and v_k are variables such that:

a. v_i is free for v_k in A and A has no free occurrences of v_i;
b. v_j is free for v_k in A' and A' has no free occurrences of v_j.

Then $(\forall v_j)A'[v_j/v_k]$ is an *alphabtic variant of* $(\forall v_i)A[v_i/v_k]$.

Likewise for $(\exists v_i)A[v_i/v_k]$ and $(\exists v_j)A'[v_j/v_k]$.

<u>Remark</u> Note that the only way in which two alphabtic variants can differ is by having different bound variables subject to the restrictions imposed in clause (iii). This means in particular that if the alphabetic variants A and A' have any free variables at all, they have exactly the same free variable occurrences. (For instance, if A has a free occurence of the variable v_i, then A' has a free occurrence of that same variable v_i, in exactly the same position.)

<u>Lemmma. 3'</u> Let L be a language.

(i) The relation of alphabetic variance is an equivalence relation on the set of formulas of L.

(ii) Let A be a formula with 0 or more free occurrences of the variable v_i and let v_r be a variable that is "fresh" to A, i.e. which does not occur anywhere in A (neither bound nor free). Then $(\forall v_i)A$ and $(\forall v_r)A[v_r/v_i]$ are alphabetic variants; and so are $(\exists v_i)A$ and $(\exists v_r)A[v_r/v_i]$.

<u>Exercise</u>: Prove the two parts of this proposition.

Hint: (i) should be proved by induction along the clauses of Def. 10'. (ii) follows from clause (iii) of Def. 10', if one uses the fact that A is an alphabetic
variant
of itself.

<u>Lemma 3''</u>. Whenever A and A' are alphabetic variants, then $\vdash A \leftrightarrow A'$.

Proof: We prove the result by induction along the clauses of Def, 10'.

(i): We have |- $A \leftrightarrow A$ by T33.

(ii) Suppose that |- $A \leftrightarrow A'$ and |- $B \leftrightarrow B'$. Then by the first two theorems liset under T34 |- $\neg A \leftrightarrow \neg A'$ and |- $(A \& B) \leftrightarrow (A' \& B')$. For the other connectives the result can be proved similarly, while making use of the other theorems listed under T34

(iii) Suppose that $(\forall v_i)A[v_i/v_k]$ and $(\forall v_j)A'[v_j/v_k]$ are as in clause (iii) of Def. 10'. By induction assumption |- $A \leftrightarrow A'$. Because of the restrictions on v_i, we have that v_k is free for v_i in $A[v_i/v_k]$ and that v_k has no free occurrences in $A[v_i/v_k]$. This entails that $A = (A[v_i/v_k])[v_k/v_i]$ and from that it follows that $(\forall v_i)A[v_i/v_k] \rightarrow A$ is a legitimate instance of axiom A5. So we have:

|- $(\forall v_i)A[v_i/v_k] \rightarrow A$.

Since we also have |- $A \leftrightarrow A'$, it follows that

|- $(\forall v_i)A[v_i/v_k] \rightarrow A'$.

By UG we can infer from this:

|- $(\forall v_k)((\forall v_i)A[v_i/v_k] \rightarrow A')$

We now note that v_k has no free occurrrences in $(\forall v_i)A[v_i/v_k]$, since all its free occurrrences in A have been replaced by free occurrences of v_i. If $i \neq k$, then all free occurrences of v_k are gone from $A[v_i/v_k]$; and if $i = k$, then the free occurrences of v_k are bound by $(\forall v_i)$. From this it follows that the following is an instance of A4.

$(\forall v_k)((\forall v_i)A[v_i/v_k] \rightarrow A') \rightarrow ((\forall v_i)A[v_i/v_k] \rightarrow (\forall v_k)A')$

Since the antecdent of this conditional is provable, and the conditional as a whole is too (since it is an axiom), the consequent of hte conditional is provable as well:

$\vdash ((\forall v_i)A[v_i/v_k] \rightarrow (\forall v_k)A'$ (*)

We now make use of the fact that v_j is free for v_k in A' and that v_j has no free occurrences in A'. From the first assumption it follows that $(\forall v_k)A' \rightarrow A'[v_j/v_k]$ is an instance of A5. So this formula is provable and by UG we can get from it a proof of $(\forall v_j)((\forall v_k)A' \rightarrow A'[v_j/v_k])$. Since $(\forall v_k)A'$ has no free occurrences of v_j,

$(\forall v_j)((\forall v_k)A' \rightarrow A'[v_j/v_k]) \rightarrow ((\forall v_k)A' \rightarrow (\forall v_j)A'[v_j/v_k])$

is an instance of A4, so that we get:

$\vdash (\forall v_k)A' \rightarrow (\forall v_j)A'[v_j/v_k])$.

Combining this with (*), we get:

$\vdash ((\forall v_i)A[v_i/v_k] \rightarrow (\forall v_j)A'[v_j/v_k])$

The converse of this implication is proved in exactly the same way.

The equivalence of $(\exists v_i)A[v_i/v_k]$ and $(\exists v_j)A'[v_j/v_k]$ can be obtained from the equivalence between $(\forall v_i)A[v_i/v_k]$ and $(\forall v_j)A'[v_j/v_k]$ by making use of axiom A11.

<u>1.2.2 The core of the Completeness Proof.</u>

We now turn to the construction which will yield the proof of Theorem 2.

The method we will use to prove completeness is that developed by Leon Henkin (1950). As Gödel (1929) noticed, to prove completeness it suffices to show that every consistent set of formulas has a model, where a *consistent* set of formulas is a set Δ from which no explicit contradiction can be proved: not-($\Delta \vdash \perp$). We prove this by (i) extending the given consistent set Δ to a maximal consistent set Δ_ω and (ii) using Δ_ω to construct a model which verifies all members of Δ_ω. In the present proof we confine ourselves to the case where Δ and Δ_ω are sets of sentences.

Assume that Γ is a consistent set of sentences of some language L. Let $c_1, c_2, \ldots$ be an infinite sequence of new individual constants and let L' be the language L $\cup$ $\{c_1, c_2, \ldots\}$.[3] Let $A_1, A_2, \ldots$ be an enumeration of all the sentences of L'. We define the sets Δ_i as follows:

(i) $\Delta_o = \Gamma$

(ii) $\Delta_{i+1} = \begin{cases} \Delta_i \cup \{A_{i+1}\} & \text{if } \Delta_i \cup \{A_{i+1}\} \text{ is consistent and } A_{i+1} \text{ is not of the form } (\exists v_j)B \\ \Delta_i \cup \{A_{i+1}, B[c_k/v_j]\} & \text{if } \Delta_i \cup \{A_{i+1}\} \text{ is consistent, } A_{i+1} \text{ is of the form } (\exists v_j)B \text{ and } c_k \text{ is the first new constant which does not occur in } \Delta_i \cup \{B\} \\ \Delta_i \cup \{\neg A_{i+1}\} & \text{otherwise} \end{cases}$

Let $\Delta_\omega = \bigcup_{i \varepsilon \omega} \Delta_i$. The Δ_i and Δ_ω have the following properties:

(P1) Δ_i is consistent.
(P2) Δ_ω is consistent.

---

[3] This is not directly possible, of course, in case L already contains all but a finite number of the individual constants which our formalism makes available. However, since the set of all individual constants of our formalism is infinite, it is always possible to make an "isomorphic copy" L' in which some infinite subset of this set is not included. For this language L' we can then proceed as indicated. Each consistent set of sentences of L translates into a consistent set of sentences of L' and the model for L' in which all the sentences of this second set are true can be straightforwardly converted into a model for L in which the sentences of the original set are true.

(P3) Δ_ω is complete in L', i.e. for each sentence B of L' either B ε Δ_ω or $\neg$B ε Δ_ω.

(P4) If $\vdash$ B, then B ε Δ_ω.

(P5) If B $\vdash$ C and B ε Δ_ω, then C ε Δ_ω.

(P6) $(\exists v_j)B$ ε Δ_ω iff $B[c/v_j]$ ε Δ_ω for some individual constant c.

(P7) For each closed term t of L' there is an individual constant c such that the sentence t = c belongs to Δ_ω.

Here follow proofs of the propositions P1 and P3. The others are left to the reader:

Exercise: Prove the propositions P2, P4 - P7!

Proof of P1. (By induction on n.)

(i) $\Delta_0 = \Gamma$ is consistent by assumption.

(ii) Suppose Δ_n is consistent. We show that Δ_{n+1} is consistent.

(a) Suppose that $\Delta_n \cup \{A_{n+1}\}$ is consistent. If A_{n+1} is not of the form $(\exists v_j)B$, then $\Delta_{n+1} = \Delta_n \cup \{A_{n+1}\}$ and thus consistent. So suppose that A_{n+1} is of the form $(\exists v_j)B$. Suppose that $\Delta_{n+1} = \Delta_n \cup \{(\exists v_j)B, B[c_r/v_j\}$ is inconsistent, where c_r is a new constant which occurs neither in Δ_n nor in $(\exists v_j)B$. Thus

$$\Delta_n \cup \{(\exists v_j)B, B[c_r/v_j]\} \vdash \bot \qquad (1)$$

So by T2 (the Deduction Theorem),

$$\Delta_n \cup \{(\exists v_j)B\} \vdash B[c_r/v_j] \rightarrow \bot \qquad (2)$$

That is, there is a proof (3)

C_1
C_2
.
.
.
C_{n-1}
$B[c_r/v_j] \rightarrow \bot$

all premises in which are from $\Delta_n \cup \{(\exists v_j)B\}$. Now let v_k be a variable that does not occur anywhere in the proof (3). Then it is easy to verify that

$$\begin{array}{l} C'_1 \\ C'_2 \\ . \\ . \\ . \\ C'_{n-1} \\ B[v_k/v_j] \rightarrow \perp \end{array} \tag{4}$$

is also a correct proof (which now derives the free variable formula $B[v_k/v_j]$ from the premise set $\Delta_n \cup \{(\exists v_j)B\}$. Since the premises are all sentences, we can apply UG to this last line, obtaining as next line

$$(\forall v_k)(B[v_k/v_j] \rightarrow \perp) \tag{5}$$

Using T27 and T28 we can extend this proof further to one whose last line is

$$(\exists v_k)B[v_k/v_j] \rightarrow \perp \tag{6}$$

At this point we make use of our Lemmata about alphabetic variants. From Lemma 3'.ii it follows that $(\exists v_k)B[v_k/v_j]$ is an alphabetic variant of $(\exists v_j)B$. So by Lemma 3'' $(\exists v_j)B$ and $(\exists v_k)B[v_k/v_j]$ are provably equivalent. From this it is easy to see that the proof can be further extended to noe whose last line is (7).

$$(\exists v_j)B \rightarrow \perp \tag{7}$$

We now have a proof of $(\exists v_j)B \rightarrow \perp$ from $\Delta_n \cup \{(\exists v_j)B\}$. So by T31 we have a proof of $\perp$ from $\Delta_n \cup \{(\exists v_j)B\}$. So $\Delta_n \cup \{(\exists v_j)B\}$ is inconsistent, which contradicts our assumption.

(b) Now assume that $\Delta_n \cup \{A_{n+1}\}$ is inconsistent. Then $\Delta_{n+1} = \Delta_n \cup \{\neg A_{n+1}\}$. Suppose Δ_{n+1} is inconsistent. Then we have

$$\Delta_n \cup \{A_{n+1}\} \vdash \perp \tag{8}$$

and

$\Delta_n \cup \{\neg A_{n+1}\} \vdash \perp$ (9)

From (12) we get, by T31 and T6

$\Delta_n \vdash A_{n+1}$ (10)

From (10) und (8) we conclude that $\Delta_n \vdash \perp$, but this contradicts the assumption that Δ_n is consistent. So once more our assumption that Δ_{n+1} is inconsistent has been disproved, and Δ_{n+1} is consistent.

This concludes the proof of P1.

Proof of P3.

Suppose that B is a sentence of L' such that neither $B \in \Delta_\omega$ nor $\neg B \in \Delta_\omega$. Let B be the formula A_{n+1} of our enumeration of the sentences of L' and $\neg B$ the formula A_{m+1}; and let us suppose, without loss of generality, that $n < m$. Since A_{n+1} does not belong to Δ_ω, we can conclude that

$\Delta_n \cup \{A_{n+1}\} \vdash \perp$. (1)

For if not, then A_{n+1} would have been a member of Δ_{n+1} and thus of Δ_ω. By the same reasoning we conclude that $\Delta_m \cup \{A_{m+1}\} \vdash \perp$. Moreover, since by assumption $n < m$, and so $\Delta_n \subseteq \Delta_m$, it follows from (1) that
$\Delta_m \cup \{A_{n+1}\} \vdash \perp$. So we have

$\Delta_m \cup \{B\} \vdash \perp$ (2)

and

$\Delta_m \cup \{\neg B\} \vdash \perp$ (3)

But then we infer as in the last part of the proof of P1 that Δ_m is inconsistent, which contradicts P1. So our assumption that there is a sentence B such that such that neither $B \in \Delta_\omega$ nor $\neg B \in \Delta_\omega$ has been disproved. This concludes the proof of P3. q.e.d.

We define the folllowing relation $\sim$ between constants of L':

$c \sim c'$ iff$_{def}$ the sentence $c = c'$ belongs to Δ_ω.

(P8) $\sim$ is an equivalence relation.

(P9) if $c \sim c'$ and $P(t_1,.., c,..,t_n) \varepsilon \Delta_\omega$, then $P(t_1,.., c',..,t_n) \varepsilon \Delta_\omega$.

Exercise: Prove P8 and P9!

From Δ_ω we define a model M = <U,F> as follows:

(i) U is the set of all equivalence classes $[c]_\sim$ for individual constants c of L'.

(ii) for each n-place functor g, F(g) is that n-place function from U into U such that for any members $[c_1]_\sim$, ..., $[c_n]_\sim$ of U, F(g) = $[c]_\sim$, where c is some individual constant from L' such that the sentence $g(c_1,...,c_n) = c$ belongs to Δ_ω.

(iii) for each n-place predicate P, F(P) is that n-place function from U into {0,1} such that for any members $[c_1]_\sim$, ..., $[c_n]_\sim$ of U, F(P) = 1 iff the sentence $P(c_1,...,c_n)$ belongs to Δ_ω.

N.B Note that clause (ii) entails that if g is a 0-place functor (i.e. an individual constant), then F(g) = $[g]_\sim$, since g = g will belong to Δ_ω.

We now prove by induction on the complexity of sentences B of L':

$$M \models B \text{ iff } B \varepsilon \Delta_\omega. \qquad (*)$$

Proof of (*)

Before we can turn to the proof of (*) itself we first need to say something about terms. We start by recalling that for each closed term t (i.e. each term t not containing any variables) the sentence $(\exists v_1)\ t = v_1$ is a logical theorem. (See T30.)

$$\vdash (\exists v_1)\ t = v_1 \qquad (1)$$

So $(\exists v_1)\ t = v_1\ \varepsilon\ \Delta_\omega$. This means also that if $(\exists v_1)\ t = v_1$ is the sentence A_{n+1} in our enumeration, then $\Delta_n \cup \{A_{n+1}\}$ is consistent and thus $\Delta_{n+1} = \Delta_n \cup \{(\exists v_1)\ t = v_1,\ t = c_r\}$, for some new constant c_r. So there is at least one constant c such that the sentence $t = c$ belongs to Δ_ω.

We now show that what we have made true by definition for "simple" terms of the form $g(c'_1, \ldots, c'_n)$ holds for closed terms in general:

Let a be any assignment in M. Then we have for any individual constant c of L' and any closed term t: (2)

$$[[t]]^{M,a} = [c]_\sim \text{ iff } t = c \ \varepsilon\ \Delta_\omega$$

The proof of (2) is by induction on the complexity of t. If t is an individual constant, then the result follows from clause (ii) of the definition of M. (See remark follwoing the def.)

So suppose that t is a complex term of the form $g(t_1, \ldots, t_n)$ and that (2) holds for the terms t_i. First suppose that $[[t]]^{M,a} = [c]_\sim$. Let c'_i (i = 1, ..., n) be constants such that the sentences $t_i = c'_i \ \varepsilon\ \Delta_\omega$. So by induction hypothesis,

$$[[t_i]]^{M,a} = [c'_i]_\sim \qquad (3)$$

Since $[[t]]^{M,a} = F(g)\,(<[[t_1]]^{M,a}, \ldots, [[t_n]]^{M,a}>)$, by the def. of F, we get from (3):

$$F(g)\,(<[c'_1]_\sim, \ldots, [c'_n]_\sim>) = [c]_\sim \qquad (4)$$

As we have seen (def. of F!), this is equivalent to

$$g(c'_1, \ldots, c'_n) = c \ \varepsilon\ \Delta_\omega \qquad (5)$$

Since also $t_i = c'_i \ \varepsilon\ \Delta_\omega$ for i = 1,...,n, we infer with the help of A13 that $g((t_1, \ldots, t_n) = c \ \varepsilon\ \Delta$.

Now suppose that $t = c \ \varepsilon\ \Delta_\omega$. Again choose c'_i (i = 1, ..., n) such that $t_i = c'_i \ \varepsilon\ \Delta_\omega$. Once more we have (3) because of the Induction Hypothesis. Also, by A13. etc. we may infer that (5). So, by the def. of F we get (4). (3) and (4) allow us to infer that

$$F(g)\,(<[[t_1]]^{M,a}, \ldots, [[t_n]]^{M,a}>) = [c]_\sim \qquad (6)$$

So by the definition of $[[\ .\ ]]^{M,a}$, $[[t]]^{M,a} = [c]_\sim$

We now start with the proof of (*) itself. We begin with the case where

(i) B is an atomic sentence $P(t_1, \ldots, t_n)$, in which the t_i are closed terms of L'. In this case we have, for any assignment a, $[[B]]^{M,a} = 1$ iff $<[[t_1]]^{M,a}, \ldots, [[t_n]]^{M,a}> \varepsilon F(P)$. But for each t_i we have that $[[t_i]]^{M,a} = [c'_i]_\sim$ and by definition F(P) consists precisely of those tuples $<[c'_1]_\sim, \ldots, [c'_n]_\sim>$ such that
$P(c'_1, \ldots, c'_n)\ \varepsilon\ \Delta_\omega$. Thus we conclude that $[[P(c'_1, \ldots, c'_n)]]^{M,a} = 1$ iff $P(c'_1, \ldots, c'_n)\ \varepsilon\ \Delta_\omega$.

(ii) B is of the form t = t'. Let c and c' be constants such that $t = c$ and $t' = c'\ \varepsilon\ \Delta_\omega$. First suppose that $t = t'\ \varepsilon\ \Delta_\omega$. Then, given the assumption just made, also $c = c'\ \varepsilon\ \Delta_\omega$. Sob by Def. of M, $[c]_\sim = [c']_\sim$. From the first part of the proof it follows that $[[t]]^{M,a} = [c]_\sim$ and $[[t']]^{M,a} = [c']_\sim$. So $[[t = t']]^{M,a} = 1$. If conversely $[[t = t']]^{M,a} = 1$, then reasoning as above, we infer that $[c]_\sim = [c']_\sim$, and hence that $c = c'\ \varepsilon\ \Delta_\omega$. Since also $t = c$ and $t' = c'\ \varepsilon\ \Delta_\omega$, it follows with A13 that $t = t'\ \varepsilon\ \Delta_\omega$.

(iii) B is of the form $\neg A$. Then $[[B]]^{M,a} = 1$ iff $[[A]]^{M,a} = 0$ iff (by induction hypothesis) not $(A\ \varepsilon\ \Delta_\omega)$ iff (by P2 and P3) $\neg A\ \varepsilon\ \Delta_\omega$.

The cases where B is of one of the forms A & C, A v C, $A \rightarrow C$ or $A \leftrightarrow C$ are handled similarly to (iii).

(iv) B is of the form $(\exists v_j)A$. This case requires a special case of Lemma 3, which we will state here as Lemma 3'. We also add, somewhat superfluously, a separate proof of this case.

<u>Lemma 3'.</u> (i) Let t be any term of L, c an individual constant of L, M any model for L and a an assignment in M. Then:

$$[[\ t[c/v_i]\ ]]^{M,a} = [[t]]^{M,a[F(c)/v_i]} \qquad (7)$$

(ii) Similarly, if B is a formula of L, M, c and a as under (i), then

$$[[\ B[c/v_i]\ ]]^{M,a} = [[\ B\ ]]^{M,a[F(c)/vi]} \quad (8)$$

Proof. (i) is proved by induction on the complexity of t, (ii) by induction on the complexity of B. We consider a few of the steps of these two proofs.

(i) (a) if t is a constant or a variable distinct from v_i, then $t[c/v_i]$ is the same as t, and t is assigned the same value by a and by $a[F(c)/v_i]$. So $[[\ t[c/v_i]\ ]]^{M,a} = [[\ t\ ]]^{M,a} =$
$[[t]]^{M,a[F(c)/vi]}$.

(b) Suppose that t is the term $g(t_1, \ldots, t_n)$ and that (7) holds for $t_1, \ldots, t_n$. Then

$[[\ t[c/v_i]\ ]]^{M,a} = [[\ g(t_1[c/v_i], \ldots, t_n[c/v_i])]]^{M,a} =$

$F(g)(<[[\ t_1[c/v_i]]]^{M,a}, \ldots, [[\ t_n[c/v_i]]]^{M,a}>) =$

$F(g)(<[[\ t_1]]^{M,a[F(c)/vi]},.., [[\ t_n]]^{M,a[F(c)/vi]}>) =$
$[[t]]^{M,a[F(c)/vi]}$

(ii) (a) B is the atomic formula $P((t_1, \ldots, t_n)$. This case is just like (i.a) above.

(b) B is of the form $\neg$ A while (9) is assumed for A. Then
$[[B[c/v_i]]]^{M,a} = [[\neg(A\ [c/v_i])\ ]]^{M,a} = 1$ iff
$[[A\ [c/v_i]\ ]]^{M,a} = 0$ iff (ind. hyp.) $[[A]]^{M,a[F(c)/vi]}$
$= 0$ iff $[[\ B\ ]]^{M,a[F(c)/vi]} = 1.$

(c) B is of the form $(\exists v_j)A$, with j /= i, while (9) is assumed for A. Then $[[\ B[c/v_i]\ ]]^{M,a} = 1$ iff for some u ε U_M
$[[A[c/v_i]\ ]]^{M,a[u/vj]} = 1$ iff (ind. hyp.) for some
u ε U_M $[[\ A\ ]]^{M,a[u/vj]\ [F(c)/vi]} = 1$ iff
$[[\ (\exists v_j)A]]^{M,a[F(c)/vi]} = 1.$

We now proceed with case (iv) of the proof of (*), in which B is of the form $(\exists v_j)A)$. The case where B is of the form $(\forall v_j)A$ is proved analogously. First suppose that B ε Δ_ω. Then, by the construction of Δ_ω, $A[c_r/v_i]$ ε Δ_ω for some constant c_r. So, by induction hypothesis,
$[[A[c_r/v_i]\ ]]^{M,a} = 1$. So, by Lemma 3',
$[[A]]^{M,a[F(cr)/vi]} = 1$. So there is some u in U_M such that

$[[A]]^{M,a[u/vi]} = 1$ and so by the Truth Definition,
$[[(\exists v_j)A]]^{M,a[F(c)/vi]} = 1$.

Now suppose that $[[B]]^{M,a[F(c)/vi]} = 1$. Then, by the truth definition, there is some u in U_M such that $[[A]]^{M,a[u/vi]} = 1$. but if $u \in U_M$, then there is some constant c such that $u = [c]_{\sim}$. But then, because of the way M has been defined, $[c]_{\sim} = F(c)$. So by Lemma 3' we infer that $[[A[c/v_i]]]^{M,a} = 1$. So by induction hypothesis $A[c/v_i] \in \Delta_\omega$. So, since $\vdash A[c/v_i] \rightarrow (\exists v_j)A$, $(\exists v_j)A \in \Delta_\omega$.

q.e.d.

1.3 Interlude on Set Theory and the Role of Logic in the Foundations of Mathematics

The completeness theorem has a number of almost immediate but independently important corollaries. In order to state these, however, it is necessary to make use of a number of concepts and theorems from the theory of sets. Since these go beyond the (very basic) set-theoretic knowledge which these Notes presuppose, they must be introduced before the corollaries of the completeness theorem can be presented.[4] It would have been preferable to leave these set-theoretical matters until Ch. 3, where set theory is developed in detail and in the rigorous way in which it should be in a course on formal logic and metamathematics. But waiting that long would have the disadvantage that the mentioned corollaries and a number of issues related to them would have to wait until Ch. 3 as well, instead of being discussed here and now, in immediate juxtaposition to the completeness theorem and its proof, from which they follow. That would be unnatural too, so I have settled for a compromise: The concepts and theorems we need for our immediate purposes will be introduced informally in this

[4] The only set theory presupposed here is that which can be found in the lecture notes for the first semester introduction to logic that is offered at the IMS ("Institut für Maschinelle Sprachverarbeitung") of the University of Stuttgart. (See Hans Kamp's web page, Lecture Notes/Introductory Logic (ps.file).)
The part of these notes that is devoted to set theory merely covers the basic information that will be known to any mathematician (including those who have no traffic with formal logic): set-theoretical notions such as that of 'set', 'set membership', 'set inclusion', 'union', 'intersection', 'subtraction', 'relation' and 'function' as well as the standard devices of set-theoretical notation.

interlude. A more formal treatment - of these set-theoretical concepts and results, together with many others - will then follow in Ch. 3.

Since the general tenor of this interlude is less formal and more discursive than the rest of the notes, this seems a suitable point to raise a number of other issues which are important for an understanding of the role and place of predicate logic within a wider setting of mathematical and philosophical logic, and, beyond that, within the general context of the foundations of mathematics, science und human knowledge. So before we proceed with the informal presentations of the set-theoretical notions and results we need at this point, I will begin with a few observations on these more philosophical aspects of formal logic and of the predicate calculus as its principal manifestation.

1.3.1 Predicate Logic and the Analyticity of Arithmetic.

The first observation is largely historical, and concerns the origins and motives of symbolic logic as we know it today. As noted in the introductory remarks to this chapter, the father of modern formal logic is Gottlob Frege (1848-1925). To Frege we owe the first precise formulation - in the form of his *Begriffsschrift* - of the predicate calculus. Frege's principal motive for developing his *Begriffsschrift* was a larger project, that of refuting Kant's claim that the truths of arithmetic are *synthetic a priori*. An essential ingredient to this refutation was a rigorous formulation of a symbolic language expressive enough to permit a formalisation of arithmetic, together with an (equally rigorous) formulation of a system of *inference principles* - rules for inferring from any given formulas of this language those other formulas that are logically entailed by them.

Kant (1724-1804) presented his doctrine that arithmetical truths are synthetic *a prori* in his *Kritik der Reinen Vernunft*. The theorems (or "laws") of arithmetic, he observed, present us with two connected epistemological puzzles:

(i) We can come to know the truth of arithmetical propositions - such as that 5 plus 7 equals 12, that there are infinitely primes and so on - without recourse to information about the outside world;

and

(ii) The method we have for obtaining such knowledge - that of "arithmetical proof", as it is normally called - provides us with a knowledge that is apparently 'proof' against all possible doubt or refutation.

The explanation which Kant proposed for these two observations was that the truths of arithmetic are *synthetic truths a priori*: They are truths that can be known with certainty, he surmised, and without any appeal to information about the outside world, because what they express are aspects of the nature of consciousness itself: Consciousness is constituted in such a way that it forces all our experiences of what goes on in the world outside us (as well as our experiences of our own inner life, but in this brief expose we will not speak explicitly of these any more) into a certain mould. As a consequence, the actual form in which our experiences are accessible to us when we are aware of them or reflect on them, is as much a product of the moulding which consciousness imposes on information which reaches it from the outside world as of the external facts or events which are the source of this information. Kant thought that it was possible for consciousness to detect the nature of its own constitution, and, more particularly, the general effects of that constitution on the form in which its contents are represented. In this way consciousness can recognise certain statements as true, because what they say follows from the contraints that it itself imposes on representational form.

Kant called such statements, which consciousness can irecognise as true because they pertain to its own structure, *synthetic a priori*. He saw them as truths *a priori* because they are true independently of any contingencies concerning the outside world and hence can be recognized as true without consultation of the outside world, but solely on the strength of looking into the nature and "boundary conditions" of consciousness itself. He regarded them as *synthetic* because they tell us something of substance, viz. in that they reveal the effects of the structure of human consciousness on mental representation. In this last respect they are different, he held, from purely "logical" or *analytic* truths, statements which are vacuously true by virtue of the way in which they arrange the concepts they involve: In an analytic statement the arrangement of concepts is such that the statement just could not be false - the arrangement 'pre-empts' the statement as it were, preventing it from making any meaningful statement about what its concepts refer to and thus depriving it from any opportunity to say something that could be false. Kant believed, like the vast majority of philosophers and scientists of his day, that the range of analytic truths was very limited: Analytic truths are not only vacuous but they can also be quite easily recognized as such. For understanding any statement necessarily involves recognizing the concepts it contains and the way in which they are arranged in it; so in those cases where this arrangement reduces the statement to vacuity, our understanding should be able to see that right off. Thus understanding an analytic truth would have to be tantamount to seeing that it must be, vacuously, true. Indeed, the comparatively few examples of analytic truths which Kant cites seem to confirm this judgement. They are either sentences involving predicates which stand in some obvious relation of subsumption, such as "Bachelors are unmarried.", or they are straightforward "trivialities" like the Law of Identity: "a = a".)

One aspect of the moulding force which consciousness cannot help exerting, Kant thought, is the temporal structure which it necessarily imposes on experience: We experience events as *temporally ordered*, i.e. as arranged in what he saw as an essentially discrete linear ordering. He further saw arithmetic, the theory of the natural number sequence 0, 1, 2, ..., as a reflection of this temporal dimension of the structure of consciousness. And that, he claimed, explains our ability to establish the truths of arithmetic without reference to external reality. The basis of arithmetical proof is consciousness' capacity for self-reflection.[5]

Contrary to Kant, Frege was persuaded that the truths of arithmetic are truths of logic - or analytic truths. They are truths of pure logic, he maintained, because when analyzed correctly, they can be shown to be about purely logical concepts: about the ("second order ") concept of being a concept, and, closely related to that, about an unending sequence of second order concepts n_{Co}, for n = 0, 1, 2, ... where 0_{Co} is the conept that is true of a concept C iff C has no instantiations, 1_{Co} is the conept that is true of a concept C iff C has exactly one instantiation, 2_{Co} is the conept that is true of a concept C iff C has exactly two instantiations, and so on.

It is these second order concepts, Frege held, - those of being a concept C that has exactly n instances, for n = 0, 1, ... - that should be seen as the entities that arithmetic is really about, viz. as the 'true natural numbers'. And he took these concepts to be purely logical concepts,

5 Kant held similar views about the statements of pure geometry and about certain propositions about causation (such as that every event has a cause): These statements too, he maintained, reflect intrinsic features of consciousness, which force the relevant kinds of experience into a predetermined mould. However, in the present context it is only his views on arithmetic which are at issue, for it was only in relation to those that Frege meant to challenge him.

since they can be defined in purely logical terms. (In present day terminology: each n_{C_0} can be defined by a formula of predicate logic which contains apart from the predicate symbol C only logical vocabulary; thus as defining formula for 0_{C_0} we can choose: "(C falls under C_0 iff) $\neg$ ($\exists$x) C(x)". Moreover, Frege realized that when the natural numbers 0, 1, 2, ... are identified with the concepts C_0, C_1, C_2, ..., then the familiar arithmetical operations, such as addition and multiplication, can also be defined in purely logical terms.[6]

Along these lines Frege succeeded in reducing all of standard arithmetic in an intuitively plausible way to concepts and statements that he had good reasons to regard as belonging to pure logic. To show that the *truths* of arithmetic are *logical truths*, however, something more is needed than just this: One also has to show that the true statements of arithmetic, when recast in these logical terms, *can be shown to be true for purely logical reasons*. The traditional way to go about this kind of task, and the one Frege chose, is to show that arithmetical truths can be derived by a series of infallible logical steps from a set of equally infallible basic logical laws, or 'logical axioms'. The infallible truth of these axioms must be established independently. It was primarily to this end that Frege developed the system of logic part of which has survived as the first order predicate calculus. It was also in this context that he committed the fatal error that flawed his reduction of arithmetic to logic and that to this very day noone has succeeded in repairing in a way which does full justice to Frege's original intentions.

Notwithstanding this error (about which more below), Frege's development of predicate logic has removed once and for all the misconception which Kant shared with his contemporaries, according to which analyticity is a marginal phenomenon within both language and thought, and according to which analytic statements are easily identified for what they are. Even though Frege's reduction of arithmetic to logic does not go through in the way in which he intended, he nevertheless pointed the way to a method for translating arithmetical statements into formulas of pure logic such that the latter are truths of logic when the former are truths of arithmetic, and where discovering the logical truth of the latter is in essence just as hard as discovering the "arithmetical" truth of the former. We all know how

---

6 For instance, addition of two numbers n and m can now be defined as the operation which when applied to the "numbers" n_{C_0} and m_{C_0} forms the second order concept of being a concept whose *extension* (= the set of things instantiating it) can be split into two parts one of which is the extension of a concept of which n_{C_0} is true while the otherc is the extension of a concept that n_{C_0} is true of..

hard that can be, something that even the more elementary books on number theory will make plain to anyone who might harbour any doubts on this point. Moreover, that this is not just a matter of subjective judgement was shown definitively about half a century after the publication of Frege's *Begriffsschrift* through the work of Kurt Gödel (1906-1978) and Alonzo Church (1903-1997). Following up on Gödel's Undecidability Theorem, Church proved the undecidability of predicate logic, which states in essence that there can be no algorithm (or "abstract machine") which decides for arbitrary formulas of predicate logic whether or not they are logical truths. If an argument was needed that mathematics can be genuinely difficult, this surely will be it: No formal task which is even beyond the most sophisticated calculating devices could be an easy task for any of us.

That arithmetic cannot be reduced to logic in the way Frege wanted was the great tragedy of his intellectual career. The flaw in his reduction was discovered by Bertrand Russell (1873-1970) at the very time when Frege's *Grundgesetze der Arithmetik*, the magnum opus in which his reduction of arithmetic to logic was carried out in full detail and which contained the fruits of more than two decades of assiduous work - was completed and had already gone to press.[7] Like the Fregean programme to which it dealt such a devastating blow at the time, Russell's discovery has been of enormous importance to subsequent developments in the foundations of logic and mathematics. It is known as *Russell's Paradox*.

To understand the gist of Russell's Paradox it is necessary to say a little more about Frege's attempt to reduce arithmetic to logic. Frege made an essential use of the systematic conceptual relation that exists between concepts and sets (or 'classes', the distinction between sets and classes, which will be explained in Ch. 3, doesn't matter at this point): Every concept determines a certain set (or class, but we won't mention classes any further in the following considerations), its so-called *extension*, consisting of those and only those things which *fall under* the concept (or to which, as one also says, the concept *applies*).

---

[7] Frege attempted to correct the mistake that Russell had discovered in the galley proofs of the *Grundgesetze*, which reached him at more or less the same time as Russell's letter. Unfortunately, the correction didn't improve matters: The resulting system was still inconsistent, while some of the derivations presented in the book did no longer go through as given. Nevertheless, the basic ideas of Frege's reduction of arithmetic to logic have proved enormously influential and have become a central ingredient of the philosophy of mathematics since the beginnings of the 20-th century. Russell himself developed an alternative implementation of Frege's programme in his monumental *Principia Mathematica*, written jointly with A. N. Whitehead (1861-1947).

Conversely, with each set there is associated the concept of being an element of this set (and of course, the extension of that concept is the very set from which one started). Frege's reduction of arithmetic to logic makes crucial use of what at face value appears as the obvious and uncontroversial formal version of the first of these principles. This is his so-called *Comprehension Principle*. The Comprehension Principle says that for any formula A with free variable x (A is here to be thought of as characterising the concept of being a thing such that A is true when that thing is assigned as value to x) there exists the set consisting of just those objects of which A is true. Since sets are assumed to be entirely determined by what elements they contain, this set is unique: Each concept can have only one extension (This is the so called Extensionality Principle, another fundamental principle connected with the concept 'set (and likewise with the concept 'class'.)

Exactly what the Comprehension Principle amounts to will depend on the properties of the system over all, for it is these which determine what free variable formulas the system contains. As it turned out, the expressive power of Frege's system was such as to allow instances of the Principle which lead to a contradiction; this is what Russell's Paradox showed. In modernised and somewhat simplified terms, the problem which the Paradox brings to light is the following. Among the possible values that the variables in Frege's system can take there are in particular the sets themselves. (This is a consequence of the fact that according to Frege any bound variable must range over the totality of all entities there are.) Moreover, the system makes it possible to say of two entities x and y that the former is an element of the latter; let us assume that this statement takes the form "x ε y", with ε being a 2-place predicate symbol denoting the relation "is an element of". As in any current system of predicate logic, this formula can be negated, and the two variables x and y can be identified. The result is the formula "$\neg$ (x ε x)". When we apply the Comprehension Principle to this formula, it returns the existence of a uniquely determined set X, consisting of all things which do not contain themselves as elements. The existence of X now leads directly to a contradiction: Suppose that X is an element of X. Then X does not instantiate the formula "$\neg$ (x ε x)", so it does not fall under the concept which that formula defines and so doesn't belong to its extension. In other words, X is not an element of X. This contradicts our assumption. So the assumption has been refuted and we may conclude that it is false, i.e. that X is not an element of X. This, however, amounts to saying that X does fall under the concept defined by "$\neg$ (x ε x)". That is, X does belong to the extension of that concept;

so X is an element of X after all.[8] So we have arrived at hte conclusion that X is not an element of itself and also that it is.

In other words, we have derived a logical contradiction simpliciter. In order to remove this contradiction Frege made the last minute correction in the proofs of *Grundgesetze* already referred to in fn. 7. The correction meant to restrict the applications of the Comprehension Principle to non-paradoxical cases. As noted in fn. 7, this attempt was not successful. It was the first of a number of such attempts, generally undertaken with the aim of saving the substance of Frege's reduction of arithmetic to logic while eliminating the deficiencies of its original implementation. One of the first of these, we also noted in fn. 7, was the logical system which Russell & Whitehead developed in *Principia Mathematica*. This system does away with Frege's assumption that the value ranges of variables must consist of all entities at once. In the so-called *Theory of Types* of *Principia Mathematica* this is never the case. Instead each variable belongs to some particular type, which restricts its possible values to just the entities that are of that type. Thus the Theory of Types presupposes a complex ontology of different logical types of entities, and these are reflected in the types of the variables of the formal system.

Today the Type Theory proposed by Russell & Whitehead is hardly used. But it is still with us in modified and streamlined form, viz. as the so-called *Typed λ-Calculus*, a system designed originally for the description of functions that was developed in the thirties by Church (and used by him among other things to prove the undecidability of first order predicate logic). To most linguists and computational linguists this formalism will be known primarily known through its use in Montague Grammar and other theories of formal semantics.

A conceptually quite different way of tackling the problem exposed by Russell's Paradox is the one first explored by Ernst Zermelo (1871-1953). The central idea here is that the paradoxical applications of the Comprehension Principle arise in cases where the extension of the concept to which it is applied is too large. The goal of this approach is accordingly to allow use of the Comprehension Principle only in cases

8 (N. B. The reason for calling this argument a "Paradox" is that it leads from what appear to be valid principles - the Comprehension Principle together with the other assumption used here, viz that there is such a concept as that of non-self-membership, which falls within the scope of the Principle - to a contradiction.)

where there is a previously established bound on the extension of the concept to which it is applied.

The actual form which this approach took eventually is that of a theory of sets formalised within first order predicate logic. This theory is developed as a formal theory of the basic relation of set theory, the relation of an entity x being an element of a set y. (The symbol commonly used for this purpose is the Greek letter ε, as we did just now in our proof of Russell's Paradox) The most familiar formalisations of set theory along these lines have been carried out in the predicate-logical language {ε}, in which ε is the only non-logical symbol. These formalisations are committed to the assumption that the totality of entities described by the theory consists exclusively of sets. (I.e. all entities in the universe of a model for the axioms of such a formalisation are sets.) This is an assumption that goes against the intuitions of many people, professional logicians and mathematicians no less than people outside these professions. These sensibilities can be accommodated by formalising the theory of sets in a form which also leaves room for entities which are not sets. To this end one needs a way of distinguishing sets from non-sets. Minimally this need can be met by adopting besides ε one further non-logical constant: a 1-place predicate S, which serves to distinguish the sets from those entities which are not. (Those who want to may extend the vocabulary further by introducing additional predicates and functors which make it possible to say more about entities that are not sets.) For the deeper logical and foundational issues connected with set theory as a theory of first order logic it turns out to matter little which of these two options - the one with or the one without S, etc. - one chooses. In these Notes (that is, in Chapter 3) we follow the more common practice within mathematical logic of formalising set theory as a first order theory within the language {ε}.

Even when the decision has been made to formalise set theory in this language, a further decision is needed: What set-theoretical axioms should one adopt? The set theory which is most widely used today (and the one that is presented in Chapter 3) is the so-called *Theory of Zermelo-Fraenkel*, so-called after the two mathematicians to whom the theory is due, Zermelo and the somewhat younger Abraham Fraenkel (1891 - 1965).[9]

---

9 Usually the theory of Zermelo-Fraenkel is referred to simply as "ZF". At first glance ZF closely resembles the theory that was proposed by Zermelo in 1908. The contribution made by Fraenkel consists of just one axiom, which to a casual observer might look like a minor addition. As a matter of fact Fraenkel's axiom makes an absolutely crucial difference. For details we refer to Ch.3.

All currently accepted formalisations of set theory have a feature that must worry someone who would like to maintain a sharp distinction between the truths of pure logic and those which make substantive claims about non-logical matters (in other words, the distinction between *analytic* truths and *contingent* truths, often referred to as the *analytic-synthetic* distinction). The reason is that the claims which the axioms of these formalisations make about the nature of sets appear to detract from the "purely logical" notion of a set as the extension of a concept. Rather, sets now appear as one category of mathematical objects among many others - numbers, straight lines, vectors, manifolds, and so on and so forth. In view of this the theory of sets - and this holds in particular for formalisations such as ZF - takes on a rather different character than what Frege had in mind: Not that of a (formal) theory of pure logic, but rather that of one mathematical theory among others, dealing with its own province of the mathematical universe. True, the specifically set-theoretic part of a formal theory like ZF rests on a foundation (provided by the axioms and rules of the first order predicate calculus) which we can still accept as "purely logical". But what is made to rest on this fundament seems to pertain just to the special province.

There is a tension between this view of set theory, and the fact that it is possible to develop essentially all of mathematics within it (thereby 'reducing' all of mathematics to set theory). This possibility largely confirms the intuitions of Frege, Russell, Whitehead and others that set theory (in combination with an underlying system of logic) has a universal status, which sets it apart from other branches of mathematics (such as number theory, geometry or functional analysis). This tension - between set theory as one mathematical theory among many and set theory as a general framework for the formalisation of mathematics - is one of the central unresolved issues in the philosophy of mathematics. And it is one which may well prove to be beyond resolution forever. We will turn to issues related to this question in Ch. 4.

1.3.2 Set Theory and the Formalisation of Mathematics

To fully appreciate the implications of this (admittedly informal) conclusion we must take account of another motivation for the formalisation of logic. This motivation was not so much a philosophical one - like that of Frege, who wanted to correct what he took to be Kant's misconception of the nature of arithmetic truth - but rather one which relates directly to serious problems that had arisen within mathematics itself. Roughly at the same time when Frege developed the *Begriffsschrift*, a crisis had developed within mathematics as it was practiced and understood by the professional mathematical community, and which affected some of the actual work that mathematicians were doing at the time. This crisis had its roots in the spectacular advances that had been made during the two preceding centuries in various branches of mathematics, and most strikingly in functional analysis (i.e. the theory of functions on the real an the complex numbers). Progress in that domain had led to theorems and proofs of an increasingly abstract nature - theorems and proofs which often dealt with whole classes or types of functions, rather than with particular functions for which explicit definitions could be given with the means then available. On the whole the abstract concepts that these theorems made use of were without a proper foundation. Missing in particular was a proper definition of 'function', as well as of the related concepts of 'set' and 'relation'. In some instances this unsatisfactory state of affairs led to paradoxes, in the sense elucidated above: contradictions obtained through apparently impeccable derivations from what were thought to be sound assumptions and unobjectionable definitions.

Within a discipline which until then had been regarded as the paradigm of intellectual soundness and certainty - and as the only remaining bulwark against the ever growing scepticism that had made its entry into western philosophy through the work of Descartes (1596-1650) - the discovery of these paradoxes came as a real shock; and it was felt to be of the utmost importance that the sources of these paradoxes be discovered and eliminated, so that the trustworthiness of mathematical argument would be restored. One of the ways in which mathematicians hoped to achieve this was to develop a logical formalism so rigorous and transparent that its inference principles could not possibly lead one astray, and to formalise all of mathematics (or at any rate all the parts where trouble brewed) within it. In this way, it was hoped, the paradoxical arguments would be forced to reveal their hidden assumptions and could then be banned from the new transparent formal framework within which mathematics was to be redeployed.

It is important to distinguish between this second motivation for developing systems of formal logic and the one we described as the primary motive for Frege. For one thing, the desire to put mathematics on a surer footing through formalisation within a system of symbolic logic is not confined to just arithmetic. In principle it concerns all branches of mathematics. And the branch that seemed to be most seriously in need of such an overhaul was that where the paradoxes had most glaringly appeared, viz. functional analysis. As noted, the basic ontological domain of analysis, however, is not that of the natural numbers, but that of the real numbers (of which the natural numbers form a proper, but in an important sense inseparable subset).[10]

The two motives that we have discussed for wanting to formalise the principles of logic are thus quite different; and on the basis of the little that has been said here one could well have imagined that since they seem to impose quite different requirements on formalisation, they might have led to quite different results. But in fact this is not so. In both cases the need is for a system of formal logic that

(i) correctly captures the basic constructs that are indispensable for the representation of information - including predication, sentence connectors and quantification - and gives the correct inference principles for those structures;

and

(ii) provides a suitable formalisation, on the basis provided by (i), of the notions of 'set', 'relation', 'function' and certain others that are connected with these.

It is these combined requirements which proved decisive and led to formal systems such as ZF, which on the one hand permit the formalisation of mathematics and on the other enable us to evaluate philosophical claims like Frege's thesis about the logical nature of arithmetical truth in ways not previously available.

It has to be admitted, however, that for either of these problems the solutions that ZF and like systems make available fall short of what was initially hoped for. In either case this has to do with the nature of sets

---

[10] We will see in Ch. 2 that the relationship between arithmetic and the theory of the real numbers is complicated and surprising. Connected with the mentioned inseparability of the subset of the natural numbers from the set of all real numbers is that as collectives the real numbers and the natural numbers behave very differently; as mathematical totalities they have strikingly different properties, and the same is true of the theories which describe those properties. Russell & Whitehead's *Principia Mathematica*, which we mentioned in fn. 7 in connection with Frege's project to reduce arithmetic to logic, targeted the logical formalisation of mathematics in general - a truly monumental endeavour, of which the formalisation of aritmetic is but one aspect taking up only a comparatively small part of the work as a whole.

and with what the set-theoretic axioms one adopts have to say about them. We already made the observation that what theories such as ZF have to say about sets tends to make sets look like mathematical entities - on a par with numbers, geometrical figures and so on - rather than entities belonging to the realm of pure logic. This has the effect that a development of arithmetic within a theory such as ZF looks much less like a confirmation of Frege's view of arithmetic as a part of pure logic than he probably would have found acceptable. Rather than a reduction of arithmetic to logic we seem to have a reduction of one branch of mathematics, number theory, to another, the theory of sets. Perhaps this can still be seen as a refutation of Kant, but that doesn't make it a corroberation of what Frege really wanted.

For this very same reason a system like ZF leaves room for doubt when used as a framework for sanitizing mathematics through formalisation. We noted that one of the problems in the design of these systems is to decide which set-theoretical axioms to adopt. On the one hand these axioms must be powerful enough to make fomalisation of a given part of mathematics possible. For such a formalisation requires (a) that we find a general schema for translating the statements from that part of mathematics into formulas of our formalism (e.g. into formulas of the language $\{\varepsilon\}$), and, furthermore, (b) that the translations of those statements that are theorems can be shown to be valid by formally deriving them (using the logical inference rules of the system, such as for instance MP and EG) from (logical and) set-theoretical axioms. On the other hand, however, we want our set-theoretical axioms to be *true* - that is, true of our pretheoretically given notion of set, to the extent that such a notion exists. And that not only because truth is desirable for its own sake, but also because the truth of a set of axioms guarantees their consistency. For it is consistency that we need most if our formalisation of mathematics is to provide us with the much wanted certainty that mathematics (in this new formalised guise) is free from contradiction.

One might well have thought that consistency could be established without any appeal to truth. After all, there have been in the history of mathematics and science many occasions where "axioms" that were proposed at one time were subsequently shown to be false, but where nevertheless the axiom system of which they were part was demonstrably consistent. (Within the natural sciences, whose aim it is to chart truthful accounts of aspects of the empirical world, and which make extensive use of quantitative axioms coined in mathematical language, there are instances galore of this.) In such cases it is often possible to show consistency to everyone's satisfaction but by way of arguments which do not rely on actual truth, something which would of course be impossible, since by assumption the axioms aren't all true!

Unfortunately, however, a formal proof of the consistency of the axioms of ZF - or, for that matter, of other formal systems of comparable power - is not to be had. This is one of the consequences of Gödel's famous Incompleteness Theorems, which he proved in conjunction with his already mentioned Undecidability Theorem. The only hope we have for bolstering our confidence in the consistency of a system like ZF is therefore to convince ourselves that the system is consistent because all its axioms say things that are true of what they talk about - i.e. about sets. But how and where do we get the knowledge that is extensive and solid enough to ascertain the truth of these axioms, given that it is knowledge about a realsm that is almost as elusive to is now as it must have been to those who were confronted, more than a century ago, with the bewilderingly paradoxical properties which made ist closer exploration such an urgent necessity?

1.3.3 Formalisation of Formalisations?

One of the central purposes of formalisation, we noted, is to guard against the dangers that are lurking in the shadows when mathematics is pursued without proper clarification of its basic concepts and principles. Only when these have been suitably clarified - and, in particular, when an explicit formulation has been given of the rules of mathematical proof - can we be reasonably confident that mathematical arguments, when formulated in accordance with those rules, will not lead to trouble (i.e. won't yield wrong conclusions starting from correct premises). This consideration applies not only to arguments in parts of mathematics like analysis, where the foundational crisis of the nineteenth century had its origin, but also for arguments in the realm of *metamathematics* - i.e. of that branch of mathematics which studies the mathematical properties of formal systems. In fact, for metamathematical arguments the issue of reliability is especially important. For it is on these arguments that our trust in the method of formalisation - as a method for avoiding error and inconsistency in mathematics - is partly based.

Does this mean that what we should really strive for is yet a further formalisation - a formalisation of metamathematics (i.e. of the science of formal systems) itself? The complexity of metamathematical arguments is often such that the need for a further formalisation, which turns these arguments into formal derivations, can be keenly felt. The question must be asked, however, what could really be gained by such a "secondary" formalisation. Aren't we, when we engage in such a further formalisation, setting out on a path that is circular, or that leads to an infinite regress?

Let us retrace the initial segments of this path: It starts with our need for greater reliability of mathematical arguments than informal mathematics can give us; therefore we want to develop methods of formalisation which will reveal the hidden assumptions and errors of informal arguments; to this end we want to develop formal systems within which these methods can be made explicit; however, to convince ourselves that these formal systems really do serve the purpose for which they have been developed, we want to prove that they behave in the ways we want them to.

So far so good. But is this good enough? How much trust are we entitled to place in our proofs - which as we said are often quite involved - that these systems do live up to our expectations? Shouldn't we formalise *these* proofs in their turn, in order to make sure that *they* are sound? But then, should we? For if we do, what better grounds could we find to trust this second formal system, needed for this second formalisation, than can be found for the first one?

The answer to this question is anything but straightforward. On the one hand we have to take this into consideration: The subject matter of metamathematics is different from that of the traditional branches of mathematics such as number theory, analysis or geometry. Metamathematics' topics of investigation are formal systems - systems consisting of symbols, structures built from symbols, such as strings or trees, and rules for manipulating such structures (i.e. turning some such structures by purely syntactic transformations into others). It is quite conceivable that a formal theory about such symbol systems could be proved correct or consistent in ways that are not available for formal theories about more traditional mathematical domains (such as, for instance, the natural number sequence, the continuum or the Euclidean plane, etc). For a consistency proof for such a formal theory would only have to deal only with finite structures such as strings and trees of symbols, and their formal manipulations. Such objects and operations are, one might be inclined to think, much easier to control than mathematical objects in general.

It was from such a conviction - that formal theories of formal systems are special in that their correctness (and therewith their consistency) can be demonstrated conclusively - that in the course of the first three decades of the 20-th century David Hilbert (1862-1943) developed an approach to the problem of certainty in mathematics known as *finitism*. In order to place mathematics on a certifiably sound foundation one should, he proposed, proceed in three steps:

(i) Formalise the different branches of mathematics using in each case some suitable formal system, consisting of a formalism with a precisely defined syntax and a set of axioms characterising the branch of mathematics that is being formalised.
(ii) Develop a formal system FS for the formalisation of these formal systems; FS in its turn will consist of a well-defined syntax together with formal axioms describing the general properties of the symbolic systems used in these formalisations
(iii) Demonstrate the consistency of FS.

Hilbert's hope that the correctness of such a theory FS could be established by simple and unquestionably sound methods was destroyed by the cluster of results - culminating in the famous Incompleteness Theorem - that were obtained by Gödel around 1930. These results entail that for almost any of the established domains of mathematics a formal system suitable for the formalisation of that area can be proved consistent only in systems which are more powerful than the system itself. This entails that a proof of a formal system which allows for its own formalisation - and surely the theory FS would have to be such a system - is not possible using the resources which the system itself provides.

One consequence of these general results is that since the first order predicate calculus, with the syntax, axioms and inference rules defined in Sections 1.1-1.3, is a formal system of the kind in question it cannot be proved consistent by the means that it provides. What is needed in addition are certain non-logical principles. There are various ways in which these can be made available. One of these is to add a certain compendium of axioms of set theory, like the axioms of ZF which we will discuss in Ch. 3. Note however, that in order to prove the consistency of this system an even more powerful system will be required and so on - the regress *is* infinite.

As far as the first order predicate calculus is concerned, this is no ground for serious worry. By now, after 125 years during which predicate logic has been used in uncounted applications and its formal properties have been investigated in depth, and from many different angles, the circumstantial evidence for its consistency is such as to leave little room for suspicions that the system might be inconsistent after all. In particular, the proofs of the Soundness Theorem make, in view of all the different variations in which they have been given, the possibility that the deduction systems to which they pertain might YET be found to be inconsistent appear extremely remote. But the matter is quite different for a system such as ZF, in which the logical axioms of predicate logic have been extended with a powerful set of axioms which concern the notion of set. The realm of sets, and the properties of that realm which the axioms of ZF articulate, are so complex that the fact that no inconsistency has been uncovered in the course of the century during which the system has now been in use doesn't seem to entitle us to believe in its consistency with anything near the degree of confidence that appears justified in the case of the predicate calculus as such. Here a formal consistency proof would be very welcome indeed; but Gödel's results tell us that all such proofs must in a certain sense be self-defeating, since they require formal systems more powerful than the ones that they are about, for which the consistency problem rises once again, and with a vengeance.

This is not to say, however, that the formalisation of metamathematics is necessarily pointless. Even if the formal system needed in the formalisation of the notion of a formal system cannot be proved consistent in a way that raises no further questions, the formalisation may still help us to get a firmer grip on the metamathematical concepts that have been formalised, and this may help to bolster our confidence that the formal systems targeted in the formalisation - those used in the formalisation of various branches of mathematics - do indeed have the desirable properties of consistency and correctness which these proofs are meant to establish.

1.3.4 Some Concepts and Results of the Theory of Sets.

The remarks of Section 1.3.3 were meant to give a glimpse of the complex conceptual and formal relationship between logic and mathematics, and especially of the crucial and at the same time delicate role that is played within that relationship by the concept of set.

When compared with these sweeping vistas the few set-theoretical notions and theorems which we need at this point - and which will be presented in this section - will seem to be but a small matter. But actually this is misleading. As only a thorough discussion of the aims and methods of metamathematics could reveal more clearly, it is the very notions and results that will be introduced below which are at the heart of the conceptual and technical difficulties inherent in the concept of 'set' and its precarious position on the borderline between mathematics and pure logic.

The set-theoretical concepts and facts that will be needed in the next sections of this Chapter, and which will be reused in several parts of Ch. 2 ar the following:

(i) The notions of *finite* and *infinite* sets and the difference between them.

(ii) The concept of the *cardinality of* a set. Cardinality is a way of assessing the size of a set. For finite sets it amounts simply to the number of elements the set contains. But for infinite sets the notion of the "number" of elements of a set has no unambiguous meaning. Here, a careful analysis of the notions of "number" and "size" is needed. The upshot of tghis analysis is that we must distinguish between (at least) two different notions of size, 'cardinality' and 'ordinality'.

The latter notion, ordinality, applies only to sets whose elements are given in a certain order. In contrast, cardinality does not presuppose any arrangement of the elements of the set, and therefore is applicable to any set, irrespective of whether its presentation involves any kind of order. The notion of cardinality we will present below is a simplified version, but one which reveals all the most important features of the notion of cardinality.

Both the distinction between finite and infinite we will define here and the characterisation of cardinality (which differs somewhat from the 'official' definition which will be given in Ch. 3, are both based on the concept of a 1-1 function from one set X to another set Y. We begin with the notion of cardinality.

A. Comparative Cardinality.

In Chapter 3 we will be in a position to develop this notion in such a way that it will be possible to speak properly of "the cardinality of" any set X. That is, we will then be able to assign to each X a set-theoretical object which can be identified with the cardinality of X. For the time being, however, we will have to be content with something less than that. What we will introduce now are (i) the relation of two sets X and Y *being of the same cardinality* and (ii) that of X *being of greater cardinality than* Y.

The basic idea is that Y has cardinality at least as large as X iff there is a 1-1 function from X into Y.

Def.11 (i) Y *is of cardinality at least as large as* X, $X \preccurlyeq Y$, iff there exists a 1-1 function from X into Y.

(ii) X *is of greater cardinality than* Y, $Y \prec X$, iff $Y \preccurlyeq X$ and not $X \preccurlyeq Y$.

Prop.1 (Obvious properties of the relations $\preccurlyeq$ and $\prec$)

(i) $\preccurlyeq$ is reflexive; (ii) $\preccurlyeq$ is transitive.
(iii) $\prec$ is irreflexive; (iv) $\prec$ is transitive.

Perhaps the historically most important theorem of set theory says that for any set X the corresponding power set $P(X)$ is of greater cardinality than X. (The power set $P(X)$ of a set X is the set $\{Y: Y \subseteq X\}$ consisting of all subsets of X.)

Thm. 12 (Cantor) $X \prec P(X)$

Proof. We have to show (i) $X \preccurlyeq P(X)$ and (ii) not $P(X) \preccurlyeq X$. (i) is easy. The function Si which maps each element x of X onto the singleton set $\{x\}$ is a 1-1 function from X into $P(X)$.

The proof of (ii) is more interesting. (It is one of classical examples of a proof by reduction ad absurdum.) Suppose there was a 1-1 function f from $P(X)$ into X. Then we can distinguish between those $Y \subseteq X$ such that $f(Y) \in Y$ and those Y for which this is not so. Let A be the set of all Y for which this condition does not hold, and let Z be the set of all corresponding values f(Y):

(*) $A = \{Y \subseteq X: f(Y) \notin Y\}$.
(**) $Z = \{f(Y): Y \in A\}$.

Then the question whether f(Z) is an element of Z leads to a contradiction. First suppose that $f(Z) \in Z$. Then by the definition of Z, $Z \in A$. So by the definition of A, $f(Z) \notin Z$. So we have arrived at a contradiction from the assumption that $f(Z) \in Z$. So this assumption is false and we have $f(Z) \notin Z$. So by the definition of Z, $Z \notin A$. So by the definition of A, $f(Z) \in Z$, and now we have reached a contradiction which only depends on the assumption that there is a 1-1 function from $P(X)$ into X. So this assumption has been refuted.

q.e.d.

Given our definition of "Y has cardinality at least as large as that of X" there appear to be two natural definitions of the notion: "X and Y have the same cardinality": (i) $X \preccurlyeq Y$ & $Y \preccurlyeq X$; and (ii) there exists a 1-1 function from X onto Y (also called a *bijection*, or *1-1 correspondence*, between X and Y). Clearly (ii) entails (i): if f is a bijection between X and Y, then f is also a 1-1 function from X into Y and f^{-1} is a 1-1 function from Y into X. What is not obvious is that the entailment also holds in the opposite direction. This is the content of the next theorem. First we define:

Def.13 $X \sim Y$ (X is *equipollent* with Y) iff there is a bijection between X and Y

Thm. 3 (Schröder-Bernstein)

If $X \preccurlyeq Y$ and $Y \preccurlyeq X$, then $X \sim Y$.

Proof. Suppose that $X \preccurlyeq Y$ and $Y \preccurlyeq X$. Then there exists (i) a 1-1 function f from X into Y and (ii) a 1-1 function g from Y into X. Our task is to construct on the basis of these two functions a bijection h between X and Y.

The construction makes use of a lemma due to Tarski, according to which any monotonic function F from the subsets of a given set Z to subsets of Z has a fixed point (i.e. an argument of F such that $F(x) = x$):

Lemma 4. (Tarski).

Let F be a monotone function from $P(Z)$ into $P(Z)$, i.e. a function such that for all $U \subseteq V \subseteq Z$, $F(U) \subseteq F(V)$. Then there exists a $W \subseteq Z$, such that $F(W) = W$.

We will prove Lemma 4 below. But first we will use it to carry through the proof of the Schröder-Bernstein Theorem.

Let U be any subset of X. By f[U] we understand the set {f(u): u ε U}.

Consider the set Y\ f[U]. This is a subset of Y, so, using the same notation, we can form g[Y\ f[U]]. This is a subset of X. So we may define the function H from P(X) to P(X) as follows:

(*) H(U) = X\g[Y\f[U]].

Claim: H is monotone. For suppose U $\subseteq$ V $\subseteq$ X. Then f[U] $\subseteq$ f[V]; so Y\ f[V] $\subseteq$ Y\ f[U]; so g[Y\ f[V]] $\subseteq$ g[Y\ f[U]]; so X\g[Y\ f[U]] $\subseteq$ X\g[Y\ f[V]]. So, by Tarski's Lemma, H has a fixed point W. Using W we can define the bijection that we are looking for as follows:

(**) Let x ε X. Then:

(i) if x ε W, h(x) = f(x)
(ii) if x $\notin$ W, then h(x) = $g^{-1}(x)$

That h is indeed a bijection is easily verified.

(h.1) We first show that h is properly defined for all of X. Let x ε X. If x ε W, then h(x) is obviously well-defined (since f is defined for all of X). Suppose x $\notin$ W. Then x ε X\W = X\H(W) = X\(X\g[Y\ f[W]]) = g[Y\ f[W]]. So there is a y ε Y\f[W] such that x = g(y). Since g is 1-1, also y = $g^{-1}(x)$ = h(x) (by (ii) from the definition of h). So once again h(x) is defined.

(h.2) We next show that h is onto Y. Let y be any member of Y. Then we have that either y ε f[W] or y ε Y\f[W]. In the first case y = f(x) for some x ε W, and so y = f(x) = h(x). In the second case, g(y) ε X\W, so h(g(y)) = $g^{-1}(g(y))$ = y. So each y ε Y is in the Range of h, and h is onto Y.

(h.3) Finally we show that h is 1-1. Suppose that x, x' are arbitrary members of X such that x $\neq$ x'. We must show that h(x) $\neq$ h(x'). If x, x' ε W, then h(x) $\neq$ h(x') follows from the fact that f is 1-1. If x, x' ε X\W, then by the proof of (h.1) there are y, y' such that x = g(y) and x' = g(y'). Since g is 1-1, h(x) = $g^{-1}(y)$ and h(x') = $g^{-1}(y')$, it follows that h(x) $\neq$ h(x'). Lastly suppose x ε W, x' ε X\W. Then h(x) ε

$f[W]$ and $h(x') \in Y \setminus f[W]$, so again $h(x) \neq h(x')$. So h is 1-1. This concludes the proof of the Schröder-Bernstein Theorem.

q.e.d.

Proof of Tarski's Lemma:

Let F be a monotone function from $P(X)$ to $P(X)$. Let $Z = \bigcup \{Y \in P(X): Y \subseteq F(Y)\}$. We show that Z is a fixed point of F.

First note that since $\varnothing$ is a member of the set $\{Y \in P(X): Y \subseteq F(Y)\}$, this set is not empty. Second, we show that $Z \subseteq F(Z)$. Suppose $z \in Z$. Then there is a V in $\{Y \in P(X): Y \subseteq F(Y)\}$ such that $z \in V$. Since $V \in \{Y \in P(X): Y \subseteq F(Y)\}$, $V \subseteq F(V)$. Since F monotone and $V \subseteq Z$, $F(V) \subseteq F(Z)$. So $V \subseteq F(Z)$ and consequently $z \in F(Z)$.
Third, we argue that $F(Z) \subseteq Z$. Since $Z \subseteq F(Z)$, it follows by the monotonicity of F that $F(Z) \subseteq F(F(Z))$. So $F(Z)$ belongs to the set $\{Y \in P(X): Y \subseteq F(Y)\}$ and so $F(Z)$ is included in the union of that set, i.e. $F(Z) \subseteq Z$.

q.e.d.

Let us take stock of what we have so far established about the relations $\preccurlyeq$, $\prec$ and $\sim$. The Schröder-Bernstein Theorem tells us that $\sim$ is equivalent to the intersection of $\preccurlyeq$ and its converse. Moreover, $\preccurlyeq$ is reflexive and transitive, and Cantor's Theorem tells us that there is no upper bound to the sizes of sets in the sense of $\preccurlyeq$: For any set X, the cardinality of $P(X)$ is bigger than that of X. So $\preccurlyeq$ is a partial ordering without a largest element.

What we do not know yet is whether $\preccurlyeq$ is a linear order. As a matter of fact $\preccurlyeq$ is a linear order, but this is a fact that at this point we can only state. We will show that it is a fact in Chapter 3.

Thm 4. For all sets X and Y, $X \preccurlyeq Y$ or $Y \preccurlyeq X$.

B Finite and Infinite.

We now turn to the notions "finite" and "infinite" set. We have a fairly good intuitive grasp of this distinction: A finite set is one whose members can be counted and thereby shown to add up to some finite number n, an infinite set is one for which this is not possible - one can keep on counting elements without ever getting to the end. However, exactly how this intuitive idea is to be captured in formal terms is not altogether straightforward. In fact, the set-theoretical literature contains several definitions of the notions, "finite set" and "infinite set". and not all of these are based on the same conception what the difference consists in. Even so, the definitions turn out to be equivalent given sufficiently strong set-theoretic assumptions. But the assumptions that are needed for this are not entirely self-evident. In Chapter 3 we will see what these assumptions are. For now what we will do is give just one of the possible definitions. It is one for which the intuitive support appears to me to be particularly strong.

The definition of a *finite* set (and, with it, of the complementary notion of an *infinite* set) which we will adopt is based on the following consideration: If X is a finite set and Y is a proper subset of X then there can exist no bijection between X and Y. Intuitively this seems obvious: If X is finite, there must be some natural number n such that X has n members. But then, if Y is a proper subset of X, then Y has at most n-1 members, so no function which has Y as its Domain can exhaust the members of X. For infinite sets this consideration does not apply. Take for instance the set $\mathbb{N}$ of natural numbers $\{0, 1, 2, ...\}$. The function $f(n) = n-1$, defined on the proper subset $\{1, 2, ...\}$ of $\mathbb{N}$ has $\mathbb{N}$ for its range. So here we do have a bijection between $\mathbb{N}$ and a proper subset of it.

Of course this last consideration doesn't prove that bijections between a set X and a proper subset of it will exist for all sets X which we have reason to regard as infinite. But closer consideration makes this equation - a set is infinite iff there exists a bijection between it and some proper subset of it - seem very plausible. The equation comes to look compelling in particular when we think of an infinite set as one which must of necessity include a subset which can be regarded as a copy of $\mathbb{N}$. And that idea is very plausible too: If a set's being infinite is to mean intuitively that when you start counting its members, you don't get to the end of it in a finite number of steps, then that would seem to be tantamount to the set containing a (potential) copy of $\mathbb{N}'$ which gets "created" in this (unending and thus abortive) act of counting the set. (To make this assumption formally precise is not quite so easy. We will see in Ch. 3 how this can be done.)

Returning to our equation: As soon as a set X includes as one of its subsets an "isomorphic copy" $\mathbb{N}'$ of $\mathbb{N}$, the existence of a bijection with a proper subset seems warranted: Let $\mathbb{N}''$ be the subset of $\mathbb{N}'$ which we get by taking away one element 0' of $\mathbb{N}'$ (which we may think of as the "copy" of 0 under an isomorphism g between $\mathbb{N}$ and $\mathbb{N}'$). Let f be the function which maps $\mathbb{N}'$ 1-1 onto $\mathbb{N}''$ and which maps all other elements of X onto themselves. Then f is a bijection between X and its proper subset X\{0'}.

This much will have to do for now as motivation for the following definition.

<u>Def. 14</u> (i) A set X is *infinite* iff there exists a bijection between X and a proper subset of X.

(ii) X is *finite* iff x is not infinite.

Nothing that has been said so far entails that any infinite sets exist[11]. When systems for the formalisation of mathematics were first developed, there seems to have been an expectation that their existence could be proved from some more fundamental logical principles. But in the meantime this hope has had to be abandoned. The current systems of axiomatic set theory acknowledge this necessity in that they all contain an axiom which asserts the existence of some infinite set more or less directly.

The form in which this axiom is often stated is that there exists a set X which (i) contains the empty set as a member, and (ii) contains, for any set x which is a member of it, also the set $x \cup \{x\}$ as a member. (This is one way of saying that X contains all the "natural numbers", with $\varnothing$ playing the role of the number 0, $\varnothing \cup \{\varnothing\}$ (= $\{\varnothing\}$) that of the number 1, $\{\varnothing\} \cup \{\{\varnothing\}\}$ (= $\{\varnothing,\{\varnothing\}\}$) that of the number 2, etc.)

<u>Postulate.</u> (Axiom of Infinity)

There exists a set X such that:

(i) $\varnothing \in X$; and

(ii) for any x, if $x \in X$, then $x \cup \{x\} \in X$.

From the Axiom of Infinity we can easily derive that there is a smallest set satisfying the conditions (i) and (ii). For let X be as postulated. Let

[11] I am referring here to the introduction to informal set theory which we gave in the Introductory course ot which the present one is the sequel. (Notes: Logik & Mathematische Methoden I & II, University of Stuttgart, 1998/1999.

Z be the set $\{Y: Y \subseteq X \ \& \ \varnothing \ \varepsilon \ Y \ \& \ (\forall x)(x \ \varepsilon \ Y \rightarrow x \cup \{x\} \ \varepsilon \ Y\})$. Since X ε Z, Z is non-empty. So its intersection $\cap$ Z is well-defined. We will call this intersection 'ω'. Suppose that V is any set satisfying the conditions (i) and (ii) of the Axiom of Infinity. Then $V \cap X$ also satisfies these conditions and since this set is included in X it belongs to Z. So $\omega \subseteq V \cap X$ and consequently $\omega \subseteq V$. So ω is included in all sets satisfying the conditions of the Infinity Axiom and thus is the smallest among them (in the strong sense of "smaller than" as "properly included").

In Ch. 3 we will adopt a principle that will allow us to show that ω is indeed as small as any infinite set can be. More precisely, we will then be able to show that if X is any infinite set in the sense of Def. 3, then $\omega \preccurlyeq X$. For the time being, however, it is enough to observe two things:

(i) ω is the starting point of an infinite sequence of sets of ever larger cardinality: ω, $P(\omega)$, $P(P(\omega))$, $P(P(P(\omega)))$, ...

(ii) ω belongs to the category of those infinite sets that are of smallest infinite cardinality. Sets of this cardinality - i.e. sets equipollent with ω - are called "denumerable", "denumerably infinite or "countably infinite". The distinction between the countable and uncountable infinite plays an important role in many branches of mathematics and in particular in mathematical logic. One instance of its importance in logic we have already encountered: the models constructed in the completeness theorem are either finite or countably infinite. Furthermore, the way in which completeness was proved made use of the fact that the set of formulas of any first order language L (containing either a finite or a countably infinite set of non-logical symbols) is countable and thus can be enumerated as a sequence indexed by the natural numbers. In Ch. 2 we will see other instances in which the fact that certain sets are countable is important.

1.4 Corollaries to the Completeness Proof. Model Isomorphims and Elementary Equivalence.

After this set-theoretical interlude we return to the point where we left the Completeness Theorem and its proof. Corollaries 1 and 2 are some of their immediate consequences.

<u>Def. 15</u> Let L be a first order language. We say that a set Γ of sentences of L is *satisfiable* iff there is a model M for L such that for every $C \varepsilon \Gamma$, $M \models C$.

Corollary 1 consists of two simple restatements of the Correctness and the Completeness Theorem.

<u>Cor.1</u> Let L be a first order language.

a. A set Γ of sentences of L is satisfiable iff it is consistent.
b. A set Γ of sentences of L is inconsistent iff it is not satisfiable.

The next corollary is known as the Compactness Theorem. The proof, which makes an essential use of the Correctness and Completeness Theorems, is left to the reader.

<u>Cor. 2.</u> (Compactness)

Let L be a first order language. A set Γ of sentences of L is satisfiable iff every finite subset of Γ is satisfiable.

A brief remark about the term 'compactness'. The (to my knowledge) earliest use of this term occurred in connection with one of the most important theorems of *Analysis*, i.e. of the theory of the field of real numbers. This is the so-called Theorem of Heine-Borel-Lebesgue, which says: any closed bounded set of real numbers (i.e. every set that can be written as a finite union of closed intervals) which is included within the union of an infinite set Y of open intervals is already included within the union of a finite subset of this set Y. Here the term "compact" makes good intuitive sense: closed bounded sets of reals are "compact" in the sense that their points are so much "heaped together" that they cannot be spread out over an infinity of different open sets (and so in particular not over an infinity of different open intervals.

In the meantime compactness has become a central notion in *Topology*; and in fact it has had an almost unparalleled number of applications in all sorts of branches of mathematics.

The HBL Theorem can be seen as stating that a certain property P - that of being a set whose union covers a given closed bouned set Y is "finite": iff some infinite set U has P, then so does some finite subset V of U has P. This is the general form of compactness. In many instances the property P is such that if any set V has it, then every superset of v has it too. When such a property P is finite, then the implication holds both ways:

U has P iff some finite subset V of U has P.

This is so for the HBL Theorem in its original form: if the union of a finite subset of U already covers a given closed bounded subset, then surely the union of all sets in U will do too. But the substance of the compactness claim is the implication also holds in the opposite direction.

In the application of compactness that is given by the Compactness Theorem for first order predicate logic, which is stated here as Cor. 2, the infinite set U is a set Γ of sentences of some language L of first order pedicate logic and P is the property of being not satisfiable. The Compactness Theorem says this property is finite: A set Γ has P iff some finite subset of Γ has P. Taking the negations of both sides of this biconditioal gives us the Compactness as stated.

Cor. 2 follows from the *statement* of the Correctness and Completeness Theorems. This is different for the *Downward Skolem-Löwenheim Theorem*, given here as Cor. 3. The Downward Skolem-Löwenheim Theorem follows not simply from the statement of the Correctness & Completeness Theorem, but from the way in which we have proved completeness.

Cor. 3. (Downward Skolem-Löwenheim Theorem)

> If a set Γ of sentences of some first order language L has any model at all, then it has a model whose universe is at most denumerably infinite.

The Downward Skolem-Löwenheim, Cor. 3.a, follows from the proof of the Completeness Theorem. This is because for any consistent set of sentences Γ the model of Γ which is constructed in the completenesss proof is at most denumerable. For the proof given above this is so because the language L' for which a maximal consistent set is constructed, which then gives us the model M = <U,F> of Γ, is of the form $L \cup \{c_1, .., c_n, ..\}$, where $c_1, .., c_n, ..$ is a countable sequence of individual constants not occurring in L, while U consists of equivalence classes of constants each of which will contain at least one member from the sequence c_1, ..,c_n,.. . It follows that U will be at most countable.[12]

A companion theorem to the Donward Skolem-Löwenheim Theorem is the *Upward Skolem-Löwenheim Theorem*:

> Let κ be any inifinite cardinal. If a set Γ of sentences of some first order language L has a denumerably infinite model, then it has a model whose universe is of cardinality κ.

The Upward Skolem-Löwenheim Theorem doesn't follow from the proof of the Completeness Theorem as we have given it. What we need in addition is (i) a proper definiition of cardinals (especially infinite cardinals) and (ii) a generalisation of the Completeness proof for languages with arbitrarily large infinite sets of individual constants (more precisely:, with sets of individual constants of any given infinite cardinality κ). (We can, for the sake of *stating* the Upwards Skolem-Löwenheim Theorem, identify cardinalities with equivalence classes of sets under the equipollence relation $\sim$ given in Def. 13 in Section 1.3.4. But to *prove* the Theorem we need a somewhat different notion of cardinal. See XCh. 3 for details, as well as certain set-theoretical methods that are connectec with that definition.

We will return to the Upward Skolem- Löwenheim Theorem there.

<u>Exercise</u>. Prove the following statement: Suppose that L is a first order language and that Γ is a set of sentences of L which has an infinite model. Then Γ has a denumerably infinite model.

(Hint: For each natural number n there is a sentence D_n of First Order Predicate Logic which says that there are at least n different things. Let M be an infintie model of Γ. Then all D_n are true in M. So $\Gamma \cup \{D_n\}_{n = 1,2,..}$ is consistent.)

---

12 In the Appendix to this Chapter Correctness and Soundness are proved not for the axiomatic proof method described in 1.1.5, but for the method of proof by construction of a semantic tableau. This completeness proof also entails the Downward Skolem-Löwenheim Theorem as an easy corollary. The point in this case is that when an argument is valid, then there is a closed semantic tableau for the argument. Since a closed tableau is always a finite object, involving finitely man tree nodes and finitely man formulas associated with those nodes, a closed tableau for the argument $<\Gamma,B>$ will involve only fnitely many premsies from Γ, So the argument $<\Delta,B>$, where Δ is the set of those finitely many premises will also be valid.

The Downward Skolem-Löwenheim Theorem shows, in quite general terms, that first order languages are unable to "fully describe" certain structures which we should like to be able to characterise in terms of first order logic. Take e.g. the structure $\mathbb{R}$ of the real numbers, with the operations of addition, multiplication, the relation of *less than* and 0 and 1 as distinguished elements. This structure is non-denumerable. (There are as many real numbers as there are subsets of the natural numbers, so the non-denumerability follows from Cantor's Theorem.) Let Γ be any set of sentences from some first order language chosen for the purpose of describing this structure. (A common choice is the language whose non-logical constants are the two 2-place functions + and $\times$, the 2-place relation < and the individual constants 0 and 1.) According to the Skolem-Löwenheim Theorem, if the sentences in Γ are all true in $\mathbb{R}$, Γ will also be satisfied by certain denumerable models, and thus by models which differ importantly from $\mathbb{R}$. To be precise, Γ will have models which are not *isomorphic* to the intended structure $\mathbb{R}$. This intuition can be made precise as follows:

Def. 16 Let L be a language and let M = <U,F> and M' = <U',F'> be models for L.

1. We say that the function h from U into U' is an *isomorhism from* M *to* M' iff

(i) h is onto U' (h is a surjection);
(ii) h is 1-1 (h is an injection);
(iii) if α is an n-place predicate constant of L, then for all $u_1,..,u_n$ from U, $F'(\alpha)(h(u_1),..,h(u_n)) = 1$ iff $(F(\alpha)\ (u_1,..,u_n) = 1$;
(iv) if α is an n-place function constant of L, then for all $u_1,..,u_n$ from U, $F'(\alpha)(h(u_1),..,h(u_n)) = h(F(\alpha)\ (u_1,..,u_n)$.

2 M and M' are called *isomorphic*, in symbols $M \cong M'$, iff there exists an isomorphism from M to M'.

Prop. 3 For any first order language L, $\cong$ is an equivalence relation on the class of all models for L.

Evidently no sentences of any language can distinguish between isomorphic structures; for obviously such structures behave in exactly the same way with respect to truth. Indeed, we have the following theorem:

<u>Thm. 5.</u> Let M and M' be models for L and let h be an isomorphism from M to M'. Then we have for every formula A of L and every assignment **a** in M: $[[A]]^{M,\mathbf{a}} = [[A]]^{M', h.\mathbf{a}}$, where h.**a** is the *composition* of h and a, i.e. that function which assigns to each variable v_i the value $h(\mathbf{a}(v_i))$.

<u>Exercise:</u> Prove Theorem 5.

Theorem 5 has the following obvious corollary: If M and M' are isomorphic models for L and A is a sentence of L, then M ⊨ A iff M' ⊨ A. We will state this corollary using the concept of elementary equivalence:

<u>Def. 17</u> Let M and M' be models for the language L. M and M' are said to be *elementarily equivalent*, in symbols $M \equiv M'$, iff for every sentence A of L, M ⊨ A iff M' ⊨ A.

<u>Prop.</u> 4 Let M and M' be models for L. If $M \cong M'$, then $M \equiv M'$.

Cor. 3 makes explicit that there is no hope of using first order sentences to distinguish between two isomorphic structures. Arguably that is no real draw-back, since from a mahtematical point of view two ismorphic structures are essentially the same - they are the same as far a their relevant mathematical properties are concerned. One might hope, however, that it should be possible to use first order logic at least to describe structures up to isomorphism. But we already have evidence that that is not the case either. This is one of the implications of the Skolem-Löwenheim Theorems. Take for instance the Downward Skolem-Löwenheim Theorem. It entails that an uncountable structure can never be fully characterised (i.e. characterised up to isomorphism) by a set of first order sentences. For any set of sentences that is true in this structure will also be true in some denumerably infinite model, and thus in a model that is not isomorphic to the original structure. And the Downward and Upward Skolem-Löwenheim Theorems taken together netail that this negative conclusion applies to all infinite structures, countable and uncountable alike.

For finite models the situation is different. Whenever M is a finite model for some language L, then all models which are elementarily equivalent to M are isomorohic to it. We give a slightly more elaborate version of this claim in the next theorem.

Thm 6. Let M be a finite model for some language L.

1. If M' is any model for L such that $M' \equiv M$, then $M' \cong M$.
2. If L is finite, then there is a single sentence A_M of L such that for any model M' for L, if $M' \models A_M$, then $M' \cong M$.

Proof. We first prove 2. Suppose that L is finite and that $M = \langle U,F \rangle$ is a finite model for L. Since U is finite, we may assume that $U = \{u_1,..,u_n\}$ for some number n. Let $v_1,..,v_n$ be n distinct variables which we choose to correspond 1-1 to the objects $u_1,..,u_n$. (As a matter of fact, $v_1,..,v_n$ are the first n variables from the infinite list in the original definition of the syntax of predicate logic, which is fine, if not essential to the following argument.) For each k-place predicate P of L let D_P be the set consisting of all formulas $P(v_{i_1},..,v_{i_k})$, such that $F(P)(\langle u_{i_1},..,u_{i_k}\rangle) = 1$, where $u_{i_j} \varepsilon \{u_1,..,u_n\}$ for $j = 1,..,k$, and all formulas $\neg P(v_{i_1},..,v_{i_k})$, such that $F(P)(\langle u_{i_1},..,u_{i_k})\rangle) = 0$. Similarly, where g is a k-place function constant of L, let D_g be the set consisting of all formulas $g(v_{i_1},..,v_{i_k}) = v_j$, such that $F(g)(\langle u_{i_1},.., u_{i_k}\rangle) = u_j$ and all formulas $\neg (g(v_{i_1},..,v_{i_k}) = v_j)$, such that $F(g)(\langle u_{i_1},.., u_{i_k}\rangle) \neq u_j$. Let B be the conjunction of all the formulas in the sets D_P and D_g for arbitrary P and g in L. Since M is finite, each of the sets D_P and D_g is finite. Further, since by assumption L is finite, there are only finitely many such sets D_P and D_g. Therefore there are only finitely many formulas in all the sets D_P and D_g together. So we can form the conjunction B of all these formulas. B is a formula of L and can be turned into a sentence A_M in the way shown in (1).

(1) $(\exists v_1) \ldots (\exists v_n)((\Lambda_{i \neq j}\ v_i \neq v_j) \ \&\ (\forall v_{n+1})\ V_i(v_{n+1} = v_i) \ \&\ B)$

We will refer to the part of A_M which follows the initial block of existential quantifiers $(\exists v_1) \ldots (\exists v_n)$ as A^*_M.

Claim: A_M describes M up to isomorphism. That is,

(2) For any model M' for L we have: M' is a model of A_M iff $M' \cong M$.

The proof of (2) consists of two parts. First, we have to show that M is a model of A_M. This is more or less obvious from the way in which A_M has been constructed. Second, we have to show that if $M' \models A_M$, then $M' \cong M$. We observe first that if M' satisfies A_M, then there are $w_1,..,w_n$ such that $M' \models A^*_M$, i.e.

(3) $M' \models (\Lambda_{i \neq j}\ v_i \neq v_j\ \&\ (\forall v_{n+1})\ V_i(v_{n+1} = v_i)\ \&\ B)[w_1,..,w_n]$[13].

It is easily seen that because of the part of the formula in (3) which precedes B, $w_1,..,w_n$ are all the elements of $U_{M'}$. So M' has cardinality n. Moreover, the function f: $\{u_1,..,u_n\} \rightarrow \{w_1,..,w_n\}$ defined by "$f(u_i) = w_i$" is an isomorphism from M to M'. For instance, supose that P is a k-place predicate of L and $<u_{i_1},..,u_{i_k}>$ is some k-tuple of elements from $\{u_1,..,u_n\}$. Then B will contain either the conjunct $P(v_{i_1},..,v_{i_k})$, or the conjunct $\neg P(v_{i_1},..,v_{i_k})$, depending on whether $F(P)(u_{i_1},..,u_{i_k}) = 1$ or $F(P)(u_{i_1},..,u_{i_k}) = 0$. In the first case we will have, because of (3), that $M' \models P(v_{i_1},..,v_{i_k})\ [w_{i_1},..,w_{i_k}]$. This means that $F_{M'}(P)(w_{i_1},..,w_{i_k}) = 1$. i.e.

(4) $F_{M'}(P)(<f(u_{i_1}),..,f(u_{i_k})>) = F_{M'}(P)(<w_{i_1},..,w_{i_k}>) =$
$F(P)\ (<u_{i_1},..,u_{i_k}>)$.

In the second case $M' \models \neg\ P(v_{i_1},..,v_{i_k})\ [w_{i_1},..,w_{i_k}]$. So $F_{M'}(P)(w_{i_1},..,w_{i_k}) = 0$ and again we have (4) and thus satisfaction of the requirement. Since this holds for arbitrary argument sequences $u_{i_1},..,u_{i_k}$, the isomorphism requirement for P is satisfied. The case of other predicates of L and also that of any function constant of L are handled in the same way. This concludes the proof of Part 2. of the Theorem.

To prove Part 1 of thee Theorem we only need to consider the case where L is infinite, as the case where L is finite has already been dealt with. If L is infinite, we may assume that L is the union of an infinite chain of ever more inclusive finite languages L_j : $L = \cup \{L_j: j = 1, 2, ..\}$, where $L_j \subseteq L_{j+1}$ and all L_j are finite. Let M = <U,F> be a finite model for L with universe $U = \{u_1,..,u_n\}$. For each language L_j let M_j be the *reduction of* M *to* L_j, i.e. that model M_j which we obtain when we

---

13 For the notation with the objects from the model M' in squarwe brackets see the remark following Corollary 1 to Lemma 2 on p. 21.

"throw away" the specifications $F_M(\alpha)$ of the extensions in M for all those non-logical constants α of L which do not belong to L_j.[14] So M_j = $<U,F_j>$, where F_j is the restriction of F to L_j. For each j we can find a sentence A_{M_j} of the form (1) such that for any model M'_j for L_j, $M'_j \models A_{M_j}$ iff $M'_j \cong M_j$.

Let M' be any model such that $M' \equiv M$. Then $M' \models A_{M_j}$ for all j. As in the proof of Part 2, this entails (for any j whatever) that $U_{M'}$ consists of n elements $w_1,..., w_n$. Furthermore we can construct for each j, just as in the proof of 1., an isomorphism h_j between M_j and M'_j.

Now we observe the following: Since U_{M_j} (= $\{u_1,...,u_n\}$) and $U_{M'_j}$ (=$\{w_1,...,w_n\}$) are both finite, there are only finitely many different bijections from the universe of M_j to the unvierse of M'_j (i.e. only finitely many bijections from $\{u_1,...,u_n\}$ to $\{w_1,...,w_n\}$). So one of these must occur infinitely often among the infinite sequence of bijections h_1, h_2, , .. . Let h be such a bijection. We show that h is an isomorphism between M' and M. Consider any non-logical constant α of L. Suppose (without loss of generality) that α is a 2-place predicate P. There exists a number j_P such that P belongs to L_j for $j \geq j_P$. Since h = h_j for infinitely many j, there is a $j_1 \geq j_P$ such that h = h_{j_1}. Therefore f maps the extension P_M of P in M onto the extension $P_{M'}$ of P in M'. For suppose that $<u_r,u_s> \varepsilon P_M$. Then $P(v_r,v_s)$ is a conjunct of $A_{M_{j1}}$. So by the form of $A_{M_{j1}}$ specified in (1), $<w_r,w_s> \varepsilon P_{M'}$. Similarly, if it is not the case that $<u_r,u_s> \varepsilon P_M$, then $\neg P(v_r,v_s)$ occurs as a conjunct of $A_{M_{j1}}$. So by the same reasoning it is not the case that $<w_r,w_s> \varepsilon P_{M'}$.

q.e.d.

1.5 First Order Theories and Modeltheoretic Relations.

We conclude this chapter with:

(i) a discussion of the notion of a (*formal*) *theory* (*of* some first order language L), and

[14] For an explicit formal definition of model reduciton see Def. 21 below.

(ii) the definition of two fundamental relations between models:

(a) the relation of one model for a language L being a *submodel* of some other model for L, and

(b) the reduction relation between models - that relation which holds between a model M for a language L and a model M' for some more inclusive language L' iff M is *the reduction of* M'.

The first of these relations will then be applied in what will be the last significant theorem of this Chapter. This theorem is a so-called *preservation theorem*. In general, preservation theorems say that a logical formula has a certain model-theoretic property P iff it is logically equivalent to a formula with a certain syntactic form. The model-theoretic property is typically of the form: if the given formula A is true in a model M then it is also true in any model M' that stands in a certain relation R to M; in other words, P says that the truth of A is preserved going from models M to models M' standing in the relation R to M, In the theorem we will consider here, R will be the submodel relation.

We have already made a few very simple uses of the reduction relation between models, viz. in those cases where we extended a language L to a language L' with additional individual constants and then "expanded" models M for L to models M' for L' by adding inerpretations for those new constants. In each such case M is the reduction of M' to the language L. More interesting applications of the reduction relation will not be given in this Chapter. But we will encounter the relation again in the next section, in the logical theory of definitions that we will discuss in 2.5. and where it will play a central role.

1.5.1 Deductive Closure and First Order Theories.

The notion of a first order theory which we will define shortly is motivated by the use of logic in the formalisation of scientific knowledge. The formalisation of science - not only of pure mathematics but also of the empirical sciences, especially sciences like physics, chemistry, astronomy, etc. in which mathematics plays an important role - became one of the central goals of the philosophy of science in the first half of the twentieth century. This, it was thought by many, would be the one and only way to make scientific knowledge truly precise and thus to make unequivocally clear what empirical predictions would follow from any given set of scientific hypotheses. The thesis that this is the proper way to develop scientific theories is known as the *Deductive-Nomological Model* (or, abbreviated, the 'DN Model') of theory formation and scientific discovery, and the method of theory development that is implied by this model as the *D(decutive)-N(omological) Method*. The general formulation of the DN Model of scientific theory formation is due to Carl Hempel (1905-1997) and Paul Oppenheim (1885-1977).

We will have more to say about the history and the implications of the DN-Method in the last section of this chapter (Section 1.5.3). Here we will confine ourselves to just one observation, which has been of central importance in the history of scientfic methodology and the role that logic plays in it.

The assumption of the DN model that every scientific theory can be formulated as an axiomatic theory of predicate logic implies that the relation of entailment - the relation that holds between B and A when B follows from A - is the same for all scientific domains: There is just one, universally applicable entailment relation and that is the relation of logical consequence as we have defined it in these notes - B is a logical consequence of Γ if truth is preserved from Γ to B in all possible models. The Completeness Theorem for first order logic, noreover, adds to this the computability of this universal entailment relation. It tells us that there exist formal deduction methods which are correct and complete for the consequence relation of for first order logic. Any such deduction method can be used to derive the theorems of any theory formalised as an axiomatic fiirst order theory.

According to the DN Model, then, both the question: "What the entailment relations ifor different scientific domains?" and the question: "How can the entailments defined by those relations be actually computed?" are solved in one fell swoop: There is just one such relation ans any complete proof procedure for that relation can be used to compute its instances.[15]

At the time when the DN Method was first applied to particular scientific theories, and then, not long after, formulated as a general canon of scientific methodology, the logical uniformity it implied - that

---

[15] In the course of the history of formal logic since Frege and Peirce a considerable variety of such correct and complete proof methods for first order logic have been developed. Some of these look quite different from each other at least on the surface, even though they produce tghe same output. Theory engineers can take their pick.

all sciences can be seen as making use of one and the same logic - came as a revelation (or as a shock, depending on methodological or philosohical persuasions). Until then it had been widely believed that many different sciences are governed by their own, domain-specific logics, and that it was one of the important tasks of any branch of science to discover the special properties of the logic determined by its domain.

The most salient example of a science with which people associated such a belief was plane geometry. For plane geometry an axiomatic formulation had been in existence since Euclid (300 B.C.). Until the very end of the 19th century it was thought that geometry was distinguished by a special form of "geometrical reasoning", which manifests itself in the use of diagrams (of "arbitrary triangles" and so forth) and in the drawing of auxiliary lines as part of the demonstration that iseems to be making an essential use of the diagram.[16] This feature of geometry was seen as distinct from the content of Euclid's postulates as such. It took well over two thousand years before this belief in the special nature of geometrical reasoning was shown to be without a proper foundation. The demonstration was given by Hilbert in his monograph *Grundlagen der Geometrie* (1900)[17] In order to demonstrate this Hilbert had to do what noone had done before him throughout the long history of Euclidean Geometry: He formalised plane geometry explicitly as a theory of formal logic. Throughout the centuries Euclidean Geometry had been looked upon as the paradigm of an axiomatic theory. But this view only focussed on the role and meaning of Euclid's postulates. The perception of what constitutes a geometrical proof was based on intuitions about valid mathematical reasoning in general and valid geometrical reasoning in particular and was at best marginally connected to an understanding of reasoning in pure logic. Hilbert's formalisation (which with hindsight we can see as one of the first applications of the DN Method) - substituted for this intuitive conception of ehat constitutes valid geometrical reasoning a notion of entailment that was based on a precise logical analysis. It was this that enabled him to show that in last analysis there is nothing that sets geometrical reasoning apart from reasoning about any other domain.

---

16 Well-known examples are the standard proofs of the theorem that the three perpendiculars of a triangle meet each other in a single point, the theorem that the three bisectors meet in one point and the theorem that the three medians meet in one point.

17 David Hilbert (1862-1943), one of the most important and influential mathematicians of all times.

Geometry is only one scientific theory among many. The reason why Hilbert's demonstration that its logic is like that of any other domain made so much of an impact was that throughout the centuries a good deal of thought had been given to the nature of geometrical reasoning; it was in particular the views of those who had argued explicitly and extensively in favour of a mode of proof particular to geometry that Hilbert was sperceived as having refuted.[18] For other scientific domains the thought that they were or might be governed by their own special logics tended to be less specific. But as far as is possible to tell in restrospect, the thought that they too involve special kinds of logic, if perhaps not wholeheartedly embraced, wasn't firmly refected either. And for those domains the message of the DN method was as clear and unequivocal as it was for the domain of geometry: none of these domains is distinguished by a logic of its own.

Obviously it is the axioms of a theory that is formalised within first order logic which determine its properties. But even iif that is so, that doesn't settle the *identity conditions* of such theories - it doesn't settle teh question when a theory given as T and another theory given as T' are to count as one and the same theory. Two points of view are possible here. According to the first the only thing that really matters about a formal theory is which statements can be derived in it as theorems. From this point of view any two axiom sets that generate the same set of theorems are equivalent and there is no reason to distinguish between them. On this conception, then, a first order theory can be identified with the set of its theorems. There may be various ways of axiomatising the theory, but these should be seen as *different axiomatisations of the same theory*.

Sometimes. however, it isn't just the set of theorems that matters, but also the *syntactic form* of the chosen set of axioms which generates

---

18 Perhaps te most celebrated of those who argued for the specifically geometrical character of geometrical demonstrations was the British empiricist George Berkeley (1685-1753), also known as "Bishop Berkeley".

that set. Another axiom set might generate exactly the same theorems but its axioms could nevertheless have different forms, from which less can be inferred about the logical properties of the theory.[19] In such a situation it would be natural to make the choice of axioms part of the identity of the theory.

These considerations suggest two "levels of granularity" for the identity conditions of formal theories: a coarse-grained level at which a theory is identified with the set of its theorems and a fine-grained level at which theorems are identified with particular axiom sets. Here we adopt, following what is the standard practise in mathematical logic, the coarse-grained level.

This coarse-grained notion of a formal theory - or *deductive theory*, as one also says, or simply *theory,* the term we will use here - is given explicitly in Def. 18.b. It is defined in terms of the notion of *deductive closure*, which is given in Def. 18.a.

Def. 18 Let L be a first order language.

1. Let Γ be a set of sentences of L. By the *closure of* Γ *in* L, $Cl_L(\Gamma)$, we understand the set of all L-sentences which are logical consequences of Γ:
   $Cl_L(\Gamma) = \{A: A \text{ is a sentence of L} \ \& \ \Gamma \models A\}$

2. A *theory of* L, or L-*theory*, is any set T of sentences of L that is closed under deduction in L:
   T is a theory iff $Cl_L(T) = T$.

Where it is clear which language L is intended we sometimes omit the subscript L in "Cl_L". We also use $Cl(\Gamma)$ as short for $Cl_{L(\Gamma)}(\Gamma)$, where $L(\Gamma)$, *the language of* Γ, is that language which consists of all non-logical constants that occur in at least one sentence of Γ.

The operator Cl_L has a number of fairly obvious but useful properties which are listed in the following proposition.

---

19 For instance, it could be that the axioms in one set have a form from which we can infer that the set of theorems they generate is decidable - in the sense that a computer programme could be written which decides for each statement within a finite number of steps wether or not it is deducible from the axioms - whereas some other axiom set generating the very same set of theorems would not enable us to draw that conclusion because its axioms aren't of the right form.

Prop. 5 Let L be a first order language, Γ, Δ sets of sentences of L. Then the following hold:

1. $\Gamma \subseteq Cl_L(\Gamma)$.
2. $Cl_L(Cl_L(\Gamma)) = Cl_L(\Gamma)$.
3. $Cl_L(\Gamma)$ is a theory of L.
4. If $\Gamma \subseteq \Delta$, then $Cl_L(\Gamma) \subseteq Cl_L(\Delta)$.
5. Let L' be language such that $L \subseteq L'$.
   Then $Cl_L(\Gamma) = Cl_{L'}(\Gamma) \cap \{A$: A is a sentence of L$\}$

Here are some further important notions connected with theories:

Def. 19 1. Suppose that T is a theory of L and that $T = Cl_L(\Gamma)$. Then we say that Γ *axiomatises* T. T is called *finitely axiomatisable* iff there is a finite set Γ which axiomatises T.

2. A theory T is called *inconsistent* iff $T \models \perp$[20]; otherwise T is called *consistent*.

3. A theory T of L is called *complete* iff for each sentence A of L either $A \in T$ or $\neg A \in T$. (Often the term "complete" is used for "complete and consistent". In general it will be clear from the context whether this is intended.)

4. We define $\perp\!\!\!\perp_L$ to be the set of sentences of L which consistsof all sentences of L. (As stated explicitly in Prop. 6 below, this set is a theory.)

Proposition 6 collects some simple facts about theories.

Prop. 6 1. $\perp\!\!\!\perp_L$ is a theory of L.
2. A theory T of L is in consistent iff $T = \perp\!\!\!\perp_L$.
3. The set $\{A$: A is a sentence of L and $\models A\}$ is a theory of L. We refer to this theory as $\mathbb{T}_L$.
4. When T is a theory of L, then $\mathbb{T}_L \subseteq T \subseteq \perp\!\!\!\perp_L$.

---

20 Recall that $\perp$ is some fixed sentence that is a logical conrtadiction. (Our hoice was and conrinues to be the formula $(\exists v_1)\, v_1 \neq v_1$.)

5. When T and T' are theories of L, then $T \cap T'$ is a theory of L.
6. If T and T' are complete theories of L, then either $T = T'$ or $T \cup T' \vDash \bot$.

More about first order theories can be found in the exercises to this Chapter and in Chapter 2.

There is one basic notion conected with axiomatisation that we have not yet mentioned. Often, when formalising a theory by providing a set of axioms for it, we try to make sure that the axiom set contains no redundancies. Firmally: a set of sentences Γ is called *redundant* iff there is at least one sentence in the set which can be derived from the other sentences in the set; in such a situation we also call a sentence in Γ that can be derived from the other sentences in Γ *redundant in* Γ.

<u>Def. 20</u> Let Γ be a set of sentences from some first order language L.

a. Let A be a member of Γ. Then A is *redundant in* Γ iff $\Gamma \setminus \{A\} \vDash A$.
b. Γ is called *redundant* iff it has at least one redundant member.

When the purpose of choosing a set Γ is simply to provide a set whose theorems are all and only the sentences in some other set that is given in advance, then redundant members of Γ don't do any work that wouldn't be accomplished without them. In such situations it seems a matter of "logical hygiene" to replace redundant sets by smaller non-redundant ones. When the redundant set Γ is finite to start with one can always obtain a redundant subset by dropping redundant axioms one by one until a non-redundant subset of the orginal set remains which still produces the same set of theorems. (When Γ is infinite, this is in general not possible.)

Just as it is often considered a matter of logical hygiene to come up with axiomatisations that are non-redundant in the sense just defined, so it is sometimes also seen as a requirement of proper formalisation that the set of *primitive concepts* of the axiomatisation - i.e. the set of those non-logical constants that occur somewhere within the given axiom set - be "non-redundant". Here "non-redundant" is meanr in the sense that none of the concepts in the set of primitives can be defined within the given theory using the remaining concepts. Exactly what this anounts to won't be obvious and in fact it is something that requires careful explication. To do this here would carry us too far afield. However the matter in Section 2.5, of Ch. 2, which is devoted to the theory of definition.

1.5.2 Model Extension, Model Expansion and Preservation

In this section we introduce the model-theoretic relations of *submodel* and of *reduction*. Both play a part in many important theorems of Model Theory. In this section we only give an application of the submodel relation.

Def. 21 Let M = <U, F> and M' = <U', F´> be models for some language L. We say that M is a *submodel of* M' if the following conditions satisfied:

(i) $U \subseteq U'$
(ii) for each n-place predicate P of L and elements $a_1, ..., a_n$ of U, $F'(P)(<a_1, ..., a_n>) = F(P)(<a_1, ..., a_n>)$
(iii) for each n-place functor f of L and elements $a_1, ..., a_n$ of U, $F'(f)(<a_1, ..., a_n>) = F(f)(<a_1, ..., a_n>)$

When (i)-(iii) are satisfied, we also say that M' is an *extension of* M.

When M = <U,F> is a submodel of the model M' = <U',F'> for L, we sometimes denote M as "$M' \upharpoonright U$".

If the language L does not contain any function constants, then there exists for every model M' = <U', F'> for L and non-empty subset U of U' a (unique) submodel M = <U,F> of M', viz. the model obtained by defining, for each predicate P of L, F(P) as in (ii). However, when L does contain function constants, then in general this is not so. For suppose that f is an n-place function constant of L. Then the subset U of U' need not be closed under F'(f), i.e. it may be that there are $a_1, ..., a_n \varepsilon U$, such that $F'(f)(<a_1, ..., a_n>)$ belongs to $U' \setminus U$. In that case a submodel of M' with universe U cannot be defined.

The reduction relation is one that holds between models for different languages, one of which is included in the other.

<u>Def. 22</u> Let L and L' be first order languages such that $L \subseteq L'$. Let M <U,F> be a model for L and M' = <U',F'> a model for L'. Then we say that M is *the reduction of* M' *to* L, in symbols: $M = M' \upharpoonright L$, iff the following two conditions are satisfied:

(i) $U = U'$
(ii) For every non-logical constant α of L, $F(\alpha) = F'(\alpha)$

When M is the reduction of M' to L, we also say that M' is an *expansion of* M *to* L'

The following proposition is immediate from the definition of the reduction relation.

<u>Prop. 6</u> Suppose that M' is a model for the language L' and that M is the reduction of M' to the sublanguage L of L'. Then for every sentence A of L, $M \models A$ iff $M' \models A$.

Prop. 6 says that a model and its reduction verify exactly the same sentences that are interpretable in both of them. No such relation obtains in general between two models M' and M for some language L when M is a submodel of M'. In general, the only sentences whose truth values are preserved between M and M' in both directions are the quantifier free sentences of L. When we restrict attention to preservation in just one direction, we can do a little better: The truth of purely universal sentences (i.e. sentences consisting of a block of universal quantifiers followed by a quantifier-free part) is preserved from M' to M, and (ii) the truth of purely existential sentences (those sentences which consist of a block of universal quantifiers followed by a quantifier-free part) is preserved from M to M'. (Note that each of these statements can be obtained from the other by contraposition.)

<u>Def. 23</u> Let A be a formula of some language L.

(i) A is said to be *purely universal* if A is of the form $(\forall x_1) \ldots (\forall x_n)$ B, where B is quantifier free and $(\forall x_1) \ldots (\forall x_n)$ is a string of 0 or more universal quantifiers.

(ii) A is said to be *purely existential* iff A is of the form $(\exists x_1) \ldots (\exists x_n)$ B with B quantifier-free.

The following theorem is straightforward and its proof left to the reader.

Thm 7. Let M and M' be models for some language L and let M be a submodel of M'. Then for any assignment **a** in M

(i) If A is a quantifier free formula of L, then $[[A]]^{M,\mathbf{a}} = [[A]]^{M',\mathbf{a}}$.

(ii) If A is a purely universal formula, then if $[[A]]^{M',\mathbf{a}} = 1$, then $[[A]]^{M,\mathbf{a}} = 1$.

(iii) If A is a purely existential formula, then if $[[A]]^{M,\mathbf{a}} = 1$, then $[[A]]^{M',\mathbf{a}} = 1$.

Proof. To prove (1), distinguish between the case where L does not have any function constants and the case where it does. For the case where L is without function constants, it suffices to prove that for arbitrary assignments **a** in M, $[[A]]^{M,\mathbf{a}} = [[A]]^{M',\mathbf{a}}$ by induction on the complexity of A. To show (i) for the more general case where L may have function constants, we must first show (1) by induction on the complexity of t that for arbitrary assignments **a** in M, noting that (1) entails that $[[t]]^{M',\mathbf{a}} \varepsilon U_M$.

$$[[t]]^{M,\mathbf{a}} = [[t]]^{M',\mathbf{a}}, \qquad (1)$$

The proof then proceeds as for the case where L has no function constants.

q.e.d.

Cor. (i) Suppose that L, M and M' are as above and that A is a purely universal sentence of L. Then, if M' ⊨ A, then M ⊨ A.

(ii) Similarly, if L, M and M' are as above and A is a purely existential sentence of L, then, if M ⊨ A, then M' ⊨ A.

In a certain sense the results of Theorem 7 are the best we could hope for: While universal formulas are preserved by submodels, this is not generally true for formulas of a more complex quantifier structure -

Neither ∀∃-formulas (formulas consisting of a block of universal quantifiers followed by a block of existential quantifiers followed by a quantifier-free part), nor ∀∃-formulas (formulas consisting of a block of existential quantifiers followed by a block of universal quantifiers followed by a quantifier-free part) are in general preserved in either direction. Both these results follow from the stronger result that not even purely existential formulas are preserved when we go from a model to a submodel of it. One easy way to see this is to consider the language L whose only non-logical constant is the 1-place predicate P, the model $M' = \langle\{a,b\},F'\rangle$ and its submodel $M = \langle\{a\},F\rangle$, where $F(P)(\langle a\rangle) = F'(P)(\langle a\rangle) = 0$ and $F'(P)(\langle b\rangle) = 1$. Then the purely existential sentence $(\exists x)P(x)$ will be true in M' but not in M. In the same way it can be shown that the truth of purely universal sentences is in general not preserved when we go from a given model to an extension of it.

The preservation properties that Thm. 7 attributes to purely universal and purely existential sentences are obviously not restricted just to formulas of those particular forms. Any sentence that is logically equivalent to a sentence of either of these forms will necessarily share its preservation properties. For instance, if A is a purely universal sentence and B is logically equivalent to A, then B too is preserved by going from models to submodels. For suppose that M is a submodel of M' and that B is true in M'. Then A is also true in M', since it is logically equivalent to B and thus true in the same models. Since A is a purely universal sentence, A will be true in the submodel M. So, again because of the logical equivalence of A and B, B will also be true in M. The same reasoning applies to sentences logically equivalent to purely existential sentences

Interestingly, however, this set - the set of sentences that are logically equivalent to some puerly universal sentence - exhausts the set of sentences preserved by submodels. This is the content of Theorem 8.

<u>Thm 8</u>. Suppose B is a sentence that is preserved by taking submodels. Then there is a purely universal sentence A such that B is logically equivalent to A.

Theorem 8 is one of a number of model-theoretic results to the effect that if a sentence is preserved by certain model-theoretic relations then it will be logically equivalent to a formula of some special syntactic form. Such results are called "preservation theorems". The proofs of such theorems are as a rule non-trivial and in some cases they can be quite complicated. The proof of Theorem 8 is among the simpler ones. We present it as an illustration of the genre as a whole.

Proof of Thm. 8. Suppose B is a sentence of language L for which the assumption of Thm. 8 holds. Let G be the set of all purely universal sentences of L which are logically entailed by B:

$$G = \{A: A \text{ is a purely universal sentence of } L \text{ such that } B \models A\}.$$

We will show

(1) the set $G' = G \cup \{\neg B\}$ is inconsistent.

From (1) the conclusion of the theorem follows easily. For suppose G' is inconsistent. Then there are finitely many sentences $A_1, \ldots, A_n$ from G such that $\models (A_1 \& \ldots \& A_n) \rightarrow B$. It is easily seen that the conjunction $A_1 \& \ldots \& A_n$ of the purely universal sentences $A_1, \ldots, A_n$ is logically equivalent to a single purely universal sentence A. (First rename the bound variables of $A_1, \ldots, A_n$ in such a way that they are all different, i.e. that no two quantifiers in $A_1 \& \ldots \& A_n$ bind the same variable. Then the conjunction can be turned into a prenex formula that will again be purely universal.). So $A \models B$. On the other hand all the A_i belong to G. So we have $B \models A_i$ for $i = 1, \ldots, n$. So $B \models A$. So B is logically equivalent to the purely universal sentence A.

To prove that G' is inconsistent, suppose that G' is consistent. Then by Corr. 2 to the completeness theorem it has a finite or denumerably infinite model M. Let C be a function which maps each element u of U_M to a distinct individual constant c_u not occurring in L. Let L' be the expansion of L with all these new constants and let M' be the corresponding expansion of M. By D(M'), the *diagram of* M, we understand the set of all atomic sentences of L' that are true in M'. Note that the following holds for any model N for L'.

(2) N is an extension of M' iff $N \models D(M')$.

We next show that the set $D(M') \cup \{B\}$ is consistent. Suppose not. Then there are finitely many sentences $D_1, \ldots, D_k$ from D(M') such that $\models (D_1 \& \ldots \& D_k) \rightarrow \neg B$, or, equivalently,

(3) $\models B \rightarrow \neg(D_1 \& \ldots \& D_k)$.

Let $D'_1, ..., D'_k$ be obtained from $D_1, ..., D_k$ by replacing all those constants from the range of C which occur in any of the formulas $D_1, ..., D_k$ by distinct variables $y_1, ..., y_r$ not occurring in $D_1, ..., D_k$ or B. This substitution will preserve the validity of (3). Moreover, since none of the constants that are involved in the substitution occur in B, the substitution leaves B invariant. So we can conclude that the formula $B \rightarrow \neg(D'_1 \& ... \& D'_k)$ is logically valid. But then it is easy to see that $B \rightarrow (\forall y_1)..(\forall y_r)\neg(D'_1 \& ... \& D'_k)$ is also logically valid. So $B \models (\forall y_1)..(\forall y_r)\neg(D'_1 \& ... \& D'_k)$, which means that $(\forall y_1)..(\forall y_r)\neg(D'_1 \& ... \& D'_k)$ is a purely universal sentence of L logically entailed by B. Therefore $(\forall y_1)..(\forall y_r) \neg(D'_1 \& ... \& D'_k)$ is a member of G. So $(\forall y_1)..(\forall y_r) \neg(D'_1 \& ... \& D'_k)$ is true in M. But then $(\forall y_1)..(\forall y_r) \neg(D'_1 \& ... \& D'_k)$ is also true in M', which is impossible, since its instantiation $\neg (D_1 \& ... \& D_k)$ is false in M'. (Recall that M' was a model of D(M), so that $D_1, .., D_k$ are all true in M'.)

So we have shown that $D(M') \cup \{B\}$ is consistent. But this means that there is a model N of D(M') in which B is true. But if N is a model of D(M'), then M is a submodel of N. So because of the original assumption about B, $M \models B$. But this contradicts our earlier assumption that $M \models G'$, from which it follows that $M \models \neg B$. Thus this earlier assumption is refuted and with it our assumption of the consistency of G'.

q.e.d.

It is easy to infer from Theorem 8 that a sentence is purely existential iff it is preserved by model extensions. A more difficult result is the following:

<u>Thm. 9</u>. A sentence is logically equivalent to an $\forall\exists$ sentence iff it is preserved by unions of chains of models.

An $\forall\exists$ sentence is a sentence which consists of a block of universal quantifiers followed by a block of existential quantifiers followed by a quantifier-free part. (Again either block or both may be empty.) The notion of a *chain* of models, to which Thm. 9 also refers, is defined as follows. A *chain of models* for a language L is a sequence of models M_i for L such that for all n, m, if n < m, then M_n is a submodel of M_m. By the *union* of such a chain of models M_i we understand that model M such that $U_M = \cup\{U_{M_i}: i = 1, 2,.. \}$ and for any predicate P the extension in M of any non-logical constant of L is the union of its extensions in the models M_i. (It should be checked that this is a proper definition of a model for L, but the checking is easy.) Lastly we say that a sentence B is *preserved under unions of chains* iff for any chain of models M_1, M_2, ... such that B is true in all M_i, B is true in the union model M.

The proof of Thm. 9 is significantly harder than that of Thm. 8. The proof will not be given here.

1.5.2 More on Formalisation of First Order Theories in Mathematics, Science and the Systematisation of Knowledge.

In the Introduction to Section 1.5.1 we pointed out an important implication of the claim that any serious scientific theory can, no matter what its subject matter, be formalised as a theory of first order logic: the methods of proof and inference in argumentation are the same everywhere; there is only one concept of valid inference, and that is the one which is given by the logical consequence relation $\models$. To show that a sentence A and a set of sentences Γ stand in this relation, one can make use of any proof system that has been proved to be correct, and so long as only first order logic is involved it is possible to use any systems that hasve been shown to be both correct and complete. In a derivation on the basis of T that the sentence A follows from the premise set the axioms of T (and by implication any other sentences that have already been shown to be theorems) can be used as additional premises; and in fact, that is the only way in which what distinguishes T from other theories make its impact on the derivation. In other words, if T_1 and T_2 are two axiomatised theories, what follows in theory T_1 can differ from what follows in theory T_2 only when the axioms of T_1 are non-equivalent to those of T_2. It is in this way, and only in this way, that any differences between T_1 and T_2 can manifest themselves in their consequences, and thus in their content.

We mentioned in the Introduction to 1.5.1 that this conception of the the construction, use and significance of scientific theories is known as the Deductive-Nomological Model of scientific method. In this section we will address a few additional issues that the DN model raises.

The first of these has to do with what has been arguably the paradigm of the axiomatic method for more than two millennia, viz. Euclidean Geometry. In his *Elements* Euclid (ca. 300 B.C.) systematised plane geometry by reducing the facts about this domain that were known at his time to five "postulates" - five geometrical statements which were taken to be self-evidently true[21]: all other true statements about plane geometry should be derivable from these five. (The *Elements* show this to be the case for the already impressive range of geometrical statements that had been established as true in Euclid's own day.) Since then, for a total of more than 23 centuries, Euclidean Geometry has been perceived as something that anyone who wanted to lay claim to a proper education should have been exposed to. In this way it became part of the core of high school curricula in most European countries.[22]

At the same time, however - we mentioned this already in the introduction to Section 1.5 - it was thought that there are aspects to the method of geometrical proof that are unique to geometry. More specifically, the use of diagrams of "ideal", "arbitrary" figures (such as triangles, circles, parallelograms, ellipses, etc.) was held to be indispensible to such proofs and at the same time essentially geometric (i.e. irreducible to principles valid outside geometry). As noted in the introduction to 1.5, this assumption - that the "logic" of geometrical demonstration was specific to the subject of geometry - was finally dismissed by Hilbert in 1900. Hilbert was able to show that geometrical proof was in last analysis no different from proof in other areas of scientific reasoning. And he was able to show this by doing something that had never been tried before (notwithstanding the fact that Euclidean Geometry had been treated since Euclid's day as the paradigm of the axiomatic method): Hilbert spelled out the axioms with a hitherto unknown concern for logical explicitness and detail. This enabled him to bring to light certain aspects of the logic of Euclidean Geometry which had been concealed from view until then, and to show in his proofs from these axioms where those aspects play a decisive part. When one proceeds in this way it becomes clear that the diagrams which had always seemed an essential ingredient of Euclidean proofs are nothing but a visual substitute for the application of certain existence postulates, which license the steps that typically manifest

---

[21] Euclid's fifth postulate, the so-called "parallels postulate", is the one irksome case of a postulate for which self-evidence was considered problematic from the start. (The postulate was considered dubious already by Euclid himself.). In an effort to justify the parallels postulate by reducing it to less problematic assumptions mathematicians kept trying for over 2000 years to derive it from the other Euclidean postulates, which were generally accpted as self-evident. It wasn't until the early 19th century when, partly as a spin-off from the indefatigable attempts to prove the parallels postulate from Euclid's other postulates, both its consistency with and its independence from the other Euclidean postulates were at last demonstrated.

[22] It has been only during the past fifty years or so that geometry has gradually disappeared from the core curriculum. This is a development of which the full intellectual implications cannot yet be properly fathomed. They may well prove more significant than many people currently seem to think.

themselves in the form of drawing of "auxiliary lines" when proofs are given in the traditional mode, in which diagrams play their apparently essential part.[23]

Although Hilbert's formulation stops short of formalising geometry as a theory in the formal sense that it has been given in formal logic (see Def. 18) it shows clearly how such a formalisation should go. The language in which his system of plane geometry is to be formulated as a formal theory in our sense has the 1-place predicates P(oint) and L(ine) and the 2-place predicate (lies) O(n) which stands for the relation between points p and lines l w.hich holds between p and l iff p is "on" l (or, what comes to the ame thing, is one of the points that make up l).[24] One difficulty with the axioms that Hilbert proposed for a formalisation in our sense, however, is that some of his axioms cannot be stated within first order logic. This means that a straightforward

---

23 Among Hilbert´s axioms we find not only statements familiar from Euclid, such as that through any two points there goes exactly one straight line, but also that for each line there is at least one point that does not lie on it, or that for two points A and B on a line l there is at least one point C on l such that B is between A and C. All steps in geometrical proofs that seem to rely on some kind of "geometrical intuition" prove to be instantiations of general principles of this kind. The difficulty we find in deciding which auxiliar lines we should draw in order to obtain a proof for a given theeorem if geometry are just illustrations of the difficulty well known to anyone familiar with deduction within predicate logic: How do we decide which instantations of universally quantified premises will be useful in the subsequent course of a given derivation and should therefore be carried out?

Ever since computers came of age, the possibility has been explored of making them take over various tasks that arise within mathematics. Although Gödel's incompleteness and undecidability results (which antedate the birth of the modern computer by roughly 15 years) had established that mathematics cannot be reduced to mere computation, there are nevertheless certain mathematical tasks at which computers are much better than human beings, simply because they can perform certain elementary operations with such vertiginous speed that it doesn't matter if they perform lots and lots of these without tangable benefit as long as there are just a few that enable them to go ahead. Among the successful applications of computer power within mathematics are programs which make the computer search for proofs in formalised geometry, in which instantiations of universally quantified axioms play a pivotal role. In this way it has actually been possible to discover geometrical theorems, which until then had escaped attention, notwithstanding the huge amount of energy that man has spent on the discovery of new facts about geometry since antiquity. In some cases one could only be amazed that noone had stumbled on the theorem before. [Reference to Boyer & Moore].

24 Hilbert's system is an axiomatiosation of 3-dimensional geometry which contains Euclidean plane geometry as a proper part. Here we focus just on this part.

formalisation would lead to a theory within second order logic, i.e. within a logical formalism which we have not so far considered.

Moreover - and more importantly - second order logic differs from first oder logic in that it does not admit of a correct and complete proof system; there can be no Correctness-and-Completeness Theorem for second order logic. (This is one of the consequences of Gödel's Incompleteness results.) Therefore, from a methodological point of view formalisations within second order logic are less satisfactory than formalisations within first order logic; they do not permit the kind of algorithmisation of inference that correct-and-complete proof systems for first order logic provide for axiomatic first order theories. It is true that there exist certain general methods for approximating second order theories by first order theories, in which one makes use of first order axiomatisations of set theory (see section 1.3 in the present chapter and, for details on Set Theory, Ch. 3). But in general the results of these methods are genuine approximations, which are logically weaker than the theories they approximate not only with regard to their second order but also to their first order consequences. (In other words, there will be statements from the first order language of the approximating theory which are not theorms of that theory although they are logical consequences of the original theory which the first order theory approximates.)

In the particular instance of Euclidean Geometry, however, it is possible to do better. Hilbert's second order axiomatisation can be replaced by a first order theory that covers all of its first order consequences. In fact, this first order theory is complete in the sense of Def. 19: each sentence A belonging to the language of the theory is either itself a theorem or else its negation is. One way in which this complete theory can be obtained is to interpret plane geometry "analytically", i.e. as speaking of "points" that are given by pairs of real numbers (which we can think of as their x- and y-coordinates). In this analytical interpretation lines can also be identified with pairs of real numbers, to be thought of as the coefficients of linear equations. (The line consisting of all points satisfying the equation $y = ax + b$ can be identified with the pair of numbers a and b.) The relation of a point lying on a line then becomes the relation which holds between a number pair (r,s) and a number pair (a,b) iff $s = ar + b$. In this way geometrical statements translate into statements about real number arithmetic. It was proved by Alfred Tarski (1901-1983) that the arithmetic of the real numbers admits of a complete firsst order axiomatisation. This is one of the most striking results of modern mathematical logic. It is especially suprising in the light of Gödel's proof of the impossibility of a complete axiomatisation of arithmetic on the natural numbers. In fact, in combination these two results may seem quite paradoxical. More on this in Ch.2, Section 2.6.

Formal Deduction and Human Reasoning

Tarski´s axiomatisation of real number arithmetic provides us with a theory which contains as theorems not only every statement in the language of real number arithmetic that has a geometrical interpretation (in the sense of analytical geometry indicated in the last subsection) and is true on that interpretation. It also has numerous theorems that have no such geometrical interpretation. (In fact, those are, speaking somewhat loosely, the vast majority.) This is an indication that the theory isn't dealing with geometry directly, but rather with a kind of (numerical) interpretation or analogue of it. This observation brings us to another aspect of formalised geometry. We claimed earlier that the possibility of formalising plane geometry within predicate logic showed that methods of proof and inference in geometry are in last instance reducible to the universally valid deduction principles of general logic. From a purely formal perspective this claim is correct and incontrovertable. But there is also another dimension to this issue, which concerns the way in which we, human beings with the particular kind of cognitive endowment with which evolution has equipped us, reason about spatial information.

The question how we process information can be raised in relation to information of all sorts. But it has a particular importance in connection with information about space. In the lives of the vast majority of us visual information occupies a central and exceptionally important place. By and large it is what we take in through our eyes on which we rely in almost everything we do.[25] It is this kind of information that we use to find our way, to find food, to keep ourselves from stumbling or bumping into things when we move around, to recognise dangerous things and creatures from a distance at which it is still possible for us to avoid or outrun them; and so on. Furthermore, while visually acquired information tends to be very rich and complex there are many situations in which it must be processed very rapidly. Fast processing of visual information is of ubiquitous practical importance and often it is what decides between life and death. Had we not been as good at it as we have become in the course of evolution the human race (or some ancestor of it) would have been wiped out long

---

[25] This is not to deny that losing the use of any of our other senses constitutes a serious handicap too.

before we would have reached our present stage of development, in which we have the capacity to reflect on the properties of our own cognitive system and the way it relates to questions of formal logic.

Such considerations suggest that the ways in which *we* reason in geometry - the ways in which we find, state and understand proofs of geometrical theorems - may well be an outcrop of the ways in which we handle spatial, visually accessed or visualised information generally. That geometry - the science which deals with the structure of the space in which we exist and move and must see that we somehow survive - can be reduced to pure logic in the way indicated above was without any doubt a major scientific discovery. But that discovery tells us nothing about the ways in which humans reason - or how they reason most comfortably and effectively - about thee contents and structure of space.

How human beings process visual information, and how they process spatial information that is not visually acquired (which for all that is known at present need not be the same thing), are questions of the utmost importance to cognitive science. And they are questions about which much is still unknown. But they are not among the questions on the agenda of formal logic and they will play no further role in these notes.

Truth

Directly related to the cognitive issues raised in the last four paragraphs is the third issue to be discussed in this section. This is the question in what sense spatial or geometrical statements can actually be said to be true or false. Fast and accurate processing of spatial information is important because it is information about the world in which we live and struggle to keep alive. If the premises from which we draw spatial conclusions - about how far a predator or a prey is away from us, where a projectile approaching us will hit us if we do not protect ourselves from its impact or step out of its way, etc. - aren't true, then there is no relying on the conclusions we draw no matter how sound the principles we apply in drawing them may be. Sound inferencing is truth-preserving; but when premises aren't true, then there is nothing to preserve.

Fortunately much of the information that we obtain by looking around is quite trustworthy, and so are the general principles about space and motion which our cognitive system makes use of when we draw inferences from information thus obtained. So for the most part the inferences we draw *are* true in their turn and it is on the whole good policy to make use of them in our further deliberations and actions.

All this presupposes that statements about space - the general claims of geometry among them - can be distinguished into those that are true and those that are false. But what does it mean for a statement of this kind to be true or false? Before we address this question first a few words about one that is even more fundamental: What is truth in general - what is it for any statement, whatever its form or subject matter, to be true or false? Questions that are phrased in such very general terms may not admit of useful answers, and it is wise to approach them with care. But of course that is no reason for shying away from them altogether.

In fact, the question of truth has been a central concern of philosophers at least since Socrates and Plato, and it plays an important, and often central part in the thought of many of the leading philosophers from antiquity to the present. Neverthelesss, it wasn't until the 20th century that a method for defining truth was developed which is exact and at the same time very general. This is another major accomplishment of Tarski. Tarski's contributions to the theory of truth are among the most important results in philosophy of the past century and they have become the foundation of essentially all semantics within formal logic. Tarski's work on truth involves two learly distinct stages. In his essay "The Concept of Truth in formalised Languages" from 1935 he showed how truth can de defined for a quite special case - that of the sentences of a language designed for talking about one particular, comparatively simple but well-defined domain, consisting of classes structured by the relation of class inclusion. Tarski showed in a fully explicit way how the truth value of any sentence of this language is determined by on the one hand the subject matter about which it speaks and on the other by its own syntactic form.[26]

This definition is a definition of an *absolute* notion of truth, for one particular language with a fixed and well-defined subject matter. Eventually this absolute notion gave way to the *relative* notion of truth

[26] Another important result of this essay is that it spells out in the clearest possible detail what conditions have to be in place in order for a truth definition of this kind to be possible: The definition must be stated in a metalanguage which is capable of describing on the one hand the "object" language for whose statements truth is to be defined and on the other the relevant properties of the domain that the object language is designed to speak about. (Another, obvious, condition is that both the object language and its subject matter must be understood well enough to begin with in oder that descriptions of them can be exact and yet recognisable correct.)

that is the central concept in what has come to be known as model theory. (This is the notion of truth that was given in the opening section of this chapter - see Def. 7 of Section 1.1.2 - and that has been explicitly or implicitly present in more or less everything that has been discussed in this chapter from that point onwards.) In definitions of this relative notion of truth - i.e. of *truth in* a model - the fixed application domain of Tarski's 1935 essay is replaced by a quantification over arbitrary domains. These domains, we have seen, are specified in the form of models for the given object language - arbitrary structures consisting of a "universe" together with interpretations, relative to this universe, of the language's non-logical constants. In this way the truth definition becomes a complex statement in the meta-language which involves wide scope universal quantification both over expressions of the object language L and over models for L. We can get back from this more general definition of relative truth for a language L to a notion of absolute truth by instantiating the universally quantified variable which ranges over models for L to the particular structure that is L's intended subject matter.

Suppose now that we have a language L which we use for talking about some part of reality - in other words, that this part of reality is the intended subject matter of L. And let us suppose that a division of the sentences of L into those that make true statements about this subject matter and those that make false claims about it is somehow given. Suppose further that we want to come up with a formal theory T that contains the true statements of L as theorems - or, if that turns out to be asking too much, then as many of the true sentences as possible - and none of the false ones. In general the design of T will involve the choice of a particular logical language L' in which T is to be stated, and in that case the relationship between L' and L will have to be made explicit. (Typically this is done by specifying how sentences of L are to be translated into sentences of L'.) However, for the present discussion there is no harm in making the simplifying assumption that L and L' coincide. Under this assumption the requirement on T can be formulated as follows:

> It must be possible to cast the part of reality that L is used to speak about in the form of a model for L (in the sense of 'model' defined in the model theory for first order logic) and moreover this model must be a model of T.

In the optimal case where T captures as theorems all sentences of L that are true in its intended domain, T will be a complete theory and the model in question will be elementarily equivalent to all other models of T. In the suboptimal case, where T captures some but not all true sentences of L, T will have other non-equivalent models besides.

What does it mean for the given part of reality to be able to play the part of a model of T? Since we are focussing on theories formalised within first order logic the answer to this question might seem straightforward: (i) the given part of reality must determine a universe U and (ii) it must determine interpretations relative to U for each of the non-logical constants of L. But how are these components fixed? In particular - and here we return to the example that provoked this discussion - how are they fixed in the case of plane geometry? This is yet another question that may look simple at first sight, but which, when we look more closely, reveals itself as anything but. First, what is the "part of reality" that the language L of plane geometry is used to speak about?[27] Actually, in the case of plane geometry this isn't quite the right way of putting the question, for there isn't just one such part of reality, but - for all we know - indefinitely many: each "flat" plane in the three-dimensional space in which we live is a part of reality in which we expect the full range of truths of plane geometry to be exemplified. Which parts of this space qualify as "flat planes" is a non-trivial question (about which more below). But it is one of the deep-seated commitments that are part of our conception of plane geometry that if there is one part of space that qualifies as a flat plane in the sense of Euclidean Geometry, then there must be an unlimited supply of such parts.

What are examples of flat planes - or, rather, fragments of flat planes since according to the theory a Euclidean plane extends infinitely in all directions and such planes are hard to come by - in the world in which we live? Answers that might come to mind to someone who hasn't thought about the matter too much might be: the surface of a pond or a lake on a day when there is no wind; a sheet of well-made paper (which has no unevennesses); the floor of a properly constructed building; the surfaces of well-constructed tables or desks; an area of land that is without hills or dips or crevasses; and .. and .. .

Let us accept this answer for what it is and ask what would be the points and straight lines (or fragments of straight lines) that are

---

27 We may assume here that L is the first order language indicated above when we said what a formalisation of Hilbert's theory of geometry as a theory of formal logic would look like: Recall what we aaid there: the non-logtical vocbulary of L should consist of: two 1-place predicates for "point" and "line" and a 2-phace predicate for "lies on".

contained in such physically concrete (fragments of) planes. In reflecting on this question it is helpful to concentrate on just one of the cases mentioned in the answer suggested in the last paragraph. We choose the case of a sheet of paper, which for our purpose is a particularly natural choice, since it ithis kind of "flat plane fragment" that people engaged in doing geometry in the familiar traditional way, using diagrams for guidance inspiration and support, often use.[28] The "points" and "straight lines" that somebody doing geometry on paper will be actually working with are dots he makes on the sheet with a pencil or pen, and lines that he draws on it, typically with the use of a ruler. But dots, no matter how fine the pencil or pen that we make them with, have a finite diameter, whereas the points of Euclidean Geometry are assumed to be infinitesimally small. Similarly, the lines we draw will always have finite width, while the width of a straight line in Euclidean Geometry is, like the diameter of a point, supposed to be infinitesimally small too. What does this mean for the question whether the kinds of statements that geometry is promarily concerned with - such as, to pick out just two examples more or less at random, the statement that the three angles of a triangle always sum up to 180° or the statement that the bisectors of a triangle meet each other at a single point - to be true? That is actually a quite difficult problem and at the same time it is one whose importance it would be hard to overestimate. Roughly what one would like to say is that figures composed of the "points" and "lines" realised on a sheet of paper in the manner just described can do no more than confirm the statement "approximately", or "within a certain margin of error", where the margin of error is determined in some way by the finite "thickness" of the given "points" and "lines" of which the figure is made up.

The first difficulty here is that we would need a precise way of assessing *how* the margin is determined by the imperfections of the given "points" and lines (i.e. by their diamteters and the extent of their "thickness"). But even if this problem can be satisfactorily solved, there still is the further problem how confirmation is related to truth. One aspect of this second problem is revealed by a distinction familiar from the philosophy of science: given a certain margin of error that is associated with a concrete figure the figure can in principle provide a conclusive *refutation* for a geometrical statement of the kind exemplified by the two mentioned above. For instance, consider the (plainly false) analogue of the statement that the bisectors of the angles of a triangle meet each other in a single point, viz. the statement that

---

28 It is easy to see, however, that the fundamental difficulties we are about to point out arise equally in relation to any of the other examples of concrete planes just mentioned. We will cometo this presently.

the bisectors of the four angles of an arbitrary quadrangle all go through the same point. If we can draw a quadrangle in which the intersection points of two pairs of bisectors are farther apart from each other than the margin of error associated with the figure permits on the assumption that they should coincide, then that shows conclusivelythat the statement is false. In contrast, concrete figures can confirm geometrical statements of this sort only to the degree that their error margin allows. Consider for instance the statement that there is a common intersection point for the bisectors of a triangle. The best we can expect from a drawin diagram of a triangle with its three bisectors is that it confirms the statement within the given error margin. But that tells us nothing about what we will find when we test the statement at the hand of figures for which the associated error margin is significantly smaller. Thus, no matter how "good" our figures, no matter how small their error margins, agreement with all those we have considered would be partial evidence at best that the statement would also be confirmed by figures with even smaller error margins.

It should be plausible without further discussion that these problems arise not only for the case where the concrete realisations of points, lines and figures are dots and drawings on sheets of paper, but also for other ways in which points and lines can be concretely realised. And it should also be intuitively clear that these difficulties are compounded by the deviations from perfect flatness that afflict the planes or surfaces in which the given realisations of points and lines are embedded.

In short, the relation between the theory of geometry and its physical realisations is full of pitfalls and surprises.[29] And what can be observed

---

[29] One of the ironies in the history of science is that when straight lines are identified with the paths of light rays - and that, it has been agreeed for centuries, is about as good a conrete identification of the geometrical concept of a straight line there is to be had; in fact, the method of triangulation in land surveying and in astronomy is based on it -, then,the geometry of the space in which we live is *not* Euclidean (e.g. the sum of the angles of triangles whose sides are formed by light rays is not equal to 180°). This conclusion follows from Einstein's theory of General Relativity and at the present time it is also supported by substantial emporocal evidence. (e.g.. by certain (very large) triangles whose sides are paths of lihgt rays and whose angles do not add up to 180°).

It is important, by the way, to distinguish between this issue - whether on this or any other physical identification of straight lines physical space is or is not Euclidean - and the question whether Euclidean Geometry is, as Kant had it, built into the way in which we process spatial information. Although this conjunction - a non-Euclidean space determined by physical phenomena combined with a human cognition based on Euclidean geometry - is something that cannot really be accommodated within Kant's general conception of mind and world, one should for this particular relation is in many ways paradigmatic for what we find with theories of other real world phenomena, such as. among others, those of physics, chemistry or astronomy. Even in the best of cases the general statements that play the part of axioms or theorems when the theory is formalised are only confirmed by the relevant phenomena that have been considered within the error margins associated with these. A more detailed analysis of such theeories reveals that in each individual case - consisting of the set of phenomena to be accounted for and the theory that is porposed to account for them - the relationship between confirmation and truth comes with its own special difficulties. But there are nevertheless also a range of problems that all such cases have in common. An entire discipline, known as "Scientific Methodology" or as the "theory of Scientific Method", has grown up around the investigation of these general problems. Among other things it currently includes substantial parts of statistics and the theory of probability.

Scientific methodology is not among the topics of these notes. Nor does it have to be. For our actual concern here, viz. the formalisation of scientific theories, a detailed analysis of statement conformation isn't really needed. In this context it is enough to assume that such an analysis is in place and that it will provide us in each relevant case - in each case where the question arises how a theory of some part or aspect of reality might be formalised as a theory in the sense of logic - with (i) a set Tr of sentences from the given language L within which the theory is to be formalised that count as true, (ii) a set Fa of sentences of L that count as false and (usually) (iii) a remaining set Un of sentences of L which neither count as sufficiently confirmed to be included in Tr nor as sufficiently disconfirmed to be included in Fa. Any formalised version T of the theory will have to be consistent with this tripartite division of the sentences of its language in that (i) none of the sentences in Fa are among the theorems of T and (ii) the theorems of T include as many sentences from Tr as possible.[30] In cases where Un is non-empty - and it may be assumed that in practice that will allways be so - T will have new predictive power vis-a-vis the data set (Tr,Fa,Un> if and when it contains theorems that belong to the set Un. For if S is a sentence of L such that $S \varepsilon T \cap Un$, this means that according to T S should really belong to Tr rather than to Un. Further empirical investigations will then be needed to see if this prediction is correct.

---

nevertheless credit Kant with having discovered a way of looking at the question of spatial structure from an essentially cognitive perspective.

[30] Note that this representation of the general situation is a refinement of the one given on p. 98.

There is one further aspect of what we have said about the verification of geometrical statements in concrete settings that deserves to be mentioned. This is the question what should be considered the "true" subject matter of Euclidean Plane Geometry. As was already implied earlier one natural way of seeing Euclidean Geometry is as a theory that talks such ideal entities as dimension-less points and lines, and only indirectly about their concrete but imperfect realisations. And indded, it is structures made up of such ideal entities, and not parts of physical reality, that we find among the models of Eisclidean geometry when it is formalised as a logical theory. Or, put in almost equivalent terms but from a slightly different perspective: to the extent that the models of this theory can be thought of as "gemoetrical structures at all, they should be thought of as made up of ideal, dimension-less points and lines rather than of entities with non-infinitesinal size or width. Seen from this angle the accomplishment of Euclidean Geometry, when we look upon it as a theory of the "points" and "lines" that we encounter in real life, isn't just that it offers a certain set of postulates towards the description of these entities with their spatial properties and relations, but also that it presents us with a certain idealised conception, which manifests itself formally in the model (or models) of these postulates.

In the case of Euclidean geometry this way of seeing the theory's true accomplisment is particularly compelling. For as Hilbert was able to show, the axioms that he had come up with define (up to isomorphism) a single model, viz. the structure $\mathbb{R} \times \mathbb{R}$, the cartesian square of the structure $\mathbb{R}$ of the real numbers with the usual arithmetical operations of addition and multiplication.[31] The models of theories for other empirical phemomena are not always reconstructable from their axioms in this unique and explicit way. But nevertheless many of those theories can also be seen as providing not simply some set of postulates, but rather a combination of postulates and an abstract

31 To obtain this result one has to make use of certain axioms that are essentially second order. (If all axioms were first order, then Hilbert's unique model result could not hold, as we have seen in connection with the Skolem-Löwenheim Theorems.) Indeed, as we noted earlier, Hilbert's axiom system does include such axioms, the Archemedean Axiom and what he called the Completeness Axiom. (We must refer the reader to Hilbert's Foundations of Geometry or some other foundational text on geometry for an explanation of what these axioms say.) Tarski's complete axiomatisation of the first order fragment of Hilbert's theory comes (almost) as close to the ideal of unique characterisation as a first order theory ever can, in that it is not just complete, but categorical in the cardinality of the target structure $\mathbb{R} \times \mathbb{R}$. (It then follows from Morley's Theorem that the axiomatisation is categorical in all uncountable categories. However, the axiomatisation is not categorical for countable models. For more on Tarski's axiomatisation see Section 2.6.2.)

conceptualisation that manifests itself as one or more of the postulates' models and to which the phenomena themselves are related by approximation. Examples o such theories abound. Newtonian celestial mechanics, which should be thought of as speaking of structures consisting of entities that are dimension-less points with finite mass, Galilei's theory of free fall, which directly speaks of objects that are propelled by gravity but are not affected by friction, are among the cases that most of us have heard of; but there are countless others.

Theories which do not describe the phenomena they aim to account for strictly and directly, but are most naturally viewed as descriptions of idealised structures, to which the phenomena themselves stand in complex approximation relations, throw an interesting light on the meaning of the term 'model'. On the face of it, the use that is made of this term in formal logic does not seem to correspond to what most people - scientists as well as persons without a specific scientific background - understand by it when they talk about 'modelling' certain phenomena or aspects of the world. In their use of the term there is no clear distinction between model and theory. The theory itself is said to "model the phenomena". On this use of "model", theory and model are one. This is clearly a quite different way of understanding the relation between theories and models from the one that is favoured in formal logic. According to the model-theoretic conception adopted there, model and theory are, as we have seen, to be distinguished sharply: theories are syntactic objects (sets of sentences) and models semantic structures, about which the sentences from the language of the theory make true or false assertions.

Formal theories which treat the phenomena that they are meant to account for as approximations to some ideal structure can be seen as providing a link between these two conceptions of 'model'. The structures that are models of such a theory in the sense of model theory - those in which the axioms of the theory are strictly and literally true - can be seen at the same time as abstract structures which model the phenomena in the sense in which the term is used by most other people. Inasmuch as the abstract structures can be considered part of the package that theories offer towards description and explanation of the phenomena, the theories can be seen as providing us with models of the phenomena (or, to use the saame phrase once again, as modelling them) in the non-logicians' sense. But when we look inside the packages, what we see are theories and models as sharply distinct as the logicians want them to be, with the theories as syntactic objects identified by their axioms or theorems and the models as the non-syntactic structures in which the axioms and theorems are true.

As argued in the last two paragraphs, theories that come with abstract structures of which they can be seen as the direct and literal descriptions, but which at the same time function as idealisations of some empirical domain, play a kind of double role. On the one hand they can be regarded as theories of the empirical domain in question and thus as empirical theories. On the other they can be seen as formal descriptions of the given abstract structure or structures to which their theorems are directly applicable. In a case like that of plane geometry, where the abstract structure is one that can be defined in purely mathematical terms (viz. as $\mathbb{R} \times \mathbb{R}$), it is therefore possible to look upon the formal theory itself either as a theory of *applied mathematics*, which tells us something about the structure of physical space, or alternatively as an account of a purely mathematical structure and thus as a theory of *pure mathematics*. Both views are legitimate, and at least in this particular case the question which way the theory should be classified is not something that can be settled once and for all. What anyone will want to say will depend on the particular context in which the theory is viewed by him or used.

The distinction between pure and applied mathematics is fraught with difficulties and the difficulties vary with the particular branch of mathematics that we consider. But the ambivalence we have just noted for the case of geometry arises for many other branches of mathematics as well.

This is all that will be said in these notes about the meaning and use of formal theories within a wider scientific context. It should have been clear that what we have said is no more than the tip of a very large iceberg. But it is enough to enable us to raise the last question that is to be considered in this section: How *useful* can formalisations be?

How useful is Formalisation?

When you ask an empirical scientist - e.g. a physicist or a chemist - what he thinks about the usefulness of dformalising the theories he is concerned with within formal logic, his reaction is likely to be one of scepticism, perhaps even of derision. And much the same reaction can be expected from most mathematicians. The reason for this is simple. What is perhaps the most important conceptual advance connected with logical formalisation - the implication that any form of valid inference can be reduced to principles of general logic - turns out in practice to be more of a nuisance than an advantage. When proofs in pure or applied mathematics are cast in the form of logical derivations, in which every step is an application of such principles, they tend to become inordinately long, unsurveyable and well-neigh impenetrable to human understanding. Moreover, it is only rarely that such logical proof expansions reveal anything new or important. Actual formalisations of mathematical or scientific theories, in which proofs take the form of such derivations, are thus the source of unnecessary complications, and that almost always without compensating benefits.

It is important however to distinguish between (i) actual formalisation of theories and their use in mathematical or scientific research and (ii) the *possibility* of formalisation: When can a theory be formalised, and what does its formalisation look like, and what can that tell us about the theory's intrinsic structure (the structure it possesses whether we formalise it or not)? We have already encountered a number of non-trivial questions connected with formalisability and seen glimpses of the light that formal logic can throw on them. The results we mentioned about the formalisation of geometry are a telling example: Hilbert's axiomatisation determines a unique model, a structure that can be defined independently, by using methods and principles of arithmetic rather than geometry (successive applications of certain number-theoretic closure operations, leading from the natural all the way to the real numbers); this axiomatisation is therefore essentially second order, but a complete first order axiomatisation of the first order fragment of his theory is possible as well. These are deep results, that have been obtained - and could only have been obtained - by the methods of logic; and yet their importance is not restricted to logic as such, but extends to the theory's intended subject, the structure of space. In this regard they are representative of formal results about logical theories, which give us insight into the possible forms that formalised theories can take and into the logical properties associated with different forms of formalisation.

What was presented in Section 1.5.1 are the very first steps of the logical investigation of theories formalised within first order predicate logic. In Chapter 2 we will look at a number of such theories, each of which will reveal new aspects of this investigation. Not all of these aspects are directly relevant to the importance of formalisation (as a possibility, rather than an actual practice) for mathematics and science. But many of them are, and between them they yield an understanding of the logical stucture of theories (whether they be stated in the form of logical theories or not) that we could not have reached in any other way.

Exercises to Ch. 1.

1. (Comparative cardinalities of some infinite sets.)

(i) Show that the following sets are equipollent with the set $\mathbb{N}$ of natural numbers.

a. the set of all positive natural numbers
b. the set of all odd natural numbers
c. the set of all multiples of 51
d. the set of all natural numbers that are squares
e. the set of all prime numbers
f. the set $\mathbb{Z}$ of the integers
g. the set $\mathbb{Q}$ of the rational numbers
h. the set of all complex rational numbers (= the set of all numbers r + i.s, where r,s ε $\mathbb{Q}$ and $i = \sqrt{-1}$)
i. the set of all pairs <n,m> of ingtegers n and m
j. the set of all finite sequences of natural numbers

(ii) Show that the following sets are equipollent with the set $\mathbb{R}$ of real numbers.

a. the set of all real numbers $\neq 0$
b. the positive real numbers
c. the closed real number interval [0,1]
d. the open rela number interval (0,1)
e. the set $\mathbb{C}$ of complex numbers, i.e. the numbers
f. the set $\mathbb{R} \setminus \mathbb{Q}$, of the irrational real numbers

2. (Finite and Infinite)
Suppose that X, Y and Z are sets and that X $\sim$ Y. Prove:

(i) $X \preccurlyeq Z$ iff $Y \preccurlyeq Z$;
(ii) $Z \preccurlyeq X$ iff $Z \preccurlyeq Y$;
(iii) $X \prec Z$ iff $Y \prec Z$;
(iv) $Z \prec X$ iff $Z \prec Y$.

3. Suppose that Y is a finite set. Show:

(i) If $X \subseteq Y$, then X is fintie
(ii) If $X \preccurlyeq Y$, then X is finite.

4. Suppose that X, Y and Z are sets, that $Y \preccurlyeq X$ and that $X \cap Z = \varnothing$.
(i) Show: $Y \cup Z \preccurlyeq X \cup Z$.
(ii) Show that the condition that $X \cap Z = \varnothing$ cannot be dropped.

5. (i) Suppose that X is finite and Y infinite. Show that $\neg(X \preccurlyeq Y)$.

(N.B. Intuitively one would want a stronger result, viz. that $X \prec Y$. This would follow from the general principle that for any two sets A and B $X \preccurlyeq Z$ or $X \preccurlyeq Z$. We will establsih this result only in Ch. 3. One might have thought that under the special conditions that X is finnite and Y infinite this result could be obtained with elementary means. But as far as we know this is not so.)

(ii) Suppose that X and Y are finite sets. Show that $X \cup Y$ is finite.

6. Prove Propositions 5 and 6. (See pp, 77,79)

7. a. Let M be a model for some language L, and let Th(M) be the set of all sentences of L which are true in M. Show: Th(M) is a complete consistent theory of L.

b. Let M be a non-empty class of models for the language L. Let Th(M) be the set of all sentences of L which are true in each model M from M. Show: Th(M) is a consistent theory of L.

8. Show: Every infinite model is elementarily equivalent to a denumerably infinite model.

9.· Let L be some first order language, let X be some denumerably infinite set and let *K* be the set of all finite models M for L with $U_M \subseteq X$. Let T be the theory Th(*K*). Prove that T has infinite models.

10. Let L be the language {<}, with < a 2-place predicate. For each positive integer n, let M_n be the model U_n, $<_n$ >, where U_n is the set of the numbers {1, 2, ..., n} and $<_n$ is the standard 'less than' relation between the numbers in U_{M_n}. Let T be the set of sentences of L which are true in every model M_n (i.e. in all models M_n for n = 1, 2, ..).

    (i) Show that T has infinite models and that these are all liner orderings. (That is, if M = <U,<> is such a model then, < is a linear ordering of U_M.)

    (ii) Show that rthere are infinite linear orderings that are not models of T.

11. Let *M* be a finite set of finite models for some given finite language L. Show that there is a sentence $A_{\mathcal{M}}$ such that for every model M' for L:

    M' ⊨ $A_{\mathcal{M}}$ iff M' is isomorphic to one of the models in *M*.

12. A theory T of some first order language L is said to be axiomatised by the set *A* of sentences of L iff T =Cl(*A*). T is said to be *finitely axiomatisable* iff there exists some finite set *A* which axiomatises T.

    a. Show that T is finitely axiomatisable iff there is a single sentence of L which axiomatises T.

    b. Show that T is not finitely axiomatisable iff there is an infinite set *A* of sentences {A_1, A_2, A_3, ...}, which axiomatises T and which has the property that for n = 1,2,.. A_n is *properly entailed* by A_{n+1}:

$A_{n+1} \models A_n$, but not $A_n \models A_{n+1}$.

13. Let T be a theory of some 1-st order language L which only has finite models. Then there is some natural number n such that every model of T has cardinality < n.

14. Let T and T' be theories of L such that both $T \cap T'$ and $Cl(T \cup T')$ are finitely axiomatisable. Then T and T' are themselves finitely axiomatisable.

15. Let L be a 1-st order language with a finite set of non-logical constants and let T_1, T_2, .. be an infinite sequence of theories of L such that for i = 1,2,... T_{i+1} is a proper extension of T_i (i.e. $T_i \subseteq T_{i+1}$ but not $T_{i+1} \subseteq T_i$). Show that every T_i has infinite models.

16. Let L be a language of first order predicate logic which does not contain function constants of arity > 0 (i.e. of more than 0 places), let P be a predicate not occurring in L and let $L' = L \cup \{P\}$. Let the translation * of arbitrary formulas A of L into formulas A* of L' be defined as follows:

(i) $A^* = A$, in case A is atomic;

(ii) $(\neg A)^* = \neg(A)^*$, $(A \,\&\, B)^* = A^* \,\&\, B^*$, $(A \vee B)^* = A^* \vee B^*$,
$(A \rightarrow B)^* = A^* \rightarrow B^*$, $(A \leftrightarrow B)^* = A^* \leftrightarrow B^*$;

(iii) $((\exists x)A)^* = (\exists x)(P(x) \,\&\, A^*)$, $((\forall x)A)^* = (\forall x)(P(x) \rightarrow A^*)$.

Let B be the set of all sentences A* of L' that are translations of sentences A which are tautologies of L:

$B = \{A^*\colon$ A is a sentence of L and $\models A\}$.

a. Show that $B \models (\exists x)P(x)$.

b. Show that for all sentences $B \in B$, $(\exists x)P(x) \models B$.

c. Show that B is not a theory of L'.

17. Let A be a sentence from the 'pure language of identity'. i.e. from that language {} of predicate logic which doesn't contain any non-logical constants. (So the only atomic fomulas of this language are of

the form '$v_i = v_j$', where v_i and v_j are variables.) Assume that the only variables occurring in A are among $v_1, \ldots, v_n$.

Show:

(*) If A is consistent, then A has a model of at most n elements.

<u>Hint</u>: Let M und N be models für the language {}. For assignments f in M and g in N we define

$$f \sim g \quad \text{iff} \quad (\forall v_i)(\forall v_j)(\, v_i, v_j \;\varepsilon\; \{v_1, \ldots, v_n\} \rightarrow (f(v_i) = f(v_j) \leftrightarrow g(v_i) = g(v_j)))$$

By induction on the complexity of the formulas of {} we can prove for the subformulas B of A (including A itself):

(**) Iff f and g are assignments in M and N such that $f \sim g$, then

$$[[B]]^{M,f} = [[B]]^{N,g}$$

Show (**) and then prove (*) with the help of (**).

18. Let T be a theory of the language L.

(i) Let S be an infitnite set of sentences of L and let $T = Cl_L(S)$ be the theory 'axiomatised by S.

Show: T is finitely axiomatisable iff there is a finite subset S' of S such that $T = Cl(S')$.

(ii) Let $L_o = \{\}$ be the language of first order logic which contains no non-logical constants whatever. (So the only atomic formulas are those of the form "x = y", where x and y are variables.)
Let S_o the set consisting of the sentences $A_1, A_2, \ldots$ of L_o, which are defined as follows:

$$A_1 = (\exists v_1)(\exists v_2)\,(\, v_1 \neq v_2)$$
$$A_2 = (\exists v_1)(\exists v_2)(\exists v_3)\,(\, v_1 \neq v_2 \;\&\; v_1 \neq v_3 \;\&\; v_2 \neq v_3)$$
.
.
$$A_n = (\exists v_1)..(\exists v_{n+1})\,(\textstyle\bigwedge_{i\neq j} v_i \neq v_j)$$

(It is easy to see that A_n says that there are at least n+1 individuals.)

Let $T_o = Cl_L(S_o)$ the theory axiomatised by S_o.
Show that T_o is not finitely axiomatisable.

(iv) Let L be a finite language (i.e. one with finitely many non-
(iii) Let L be a finite language (i.e. ne with finitely many non-logical constants), let T be an arbritrary theory of L and let T_o be the theory defined under (ii)

Show : When $T \cup T_o$ inconsistent, then T is finitely axiomatisable.

19. Let L be a first order language and T a theory of L. For arbitrary sentences A, B of L we define:

$A \equiv_T B$ iff $T \models A \leftrightarrow B$

(i) Show that $\equiv_T$ is an in equivalence relation.

(ii) Let U be the set of all equivalence classes determined by $\equiv_T$. For sentences A of L ist we write "[A]" for the equivalence class A generaetd by A: $[A] = \{B: A \equiv_T B\}$.

On U we define the following 2-, 1- and 0-place functions:

D$\underline{\cap}$	$[A] \,\underline{\cap}\, [B]$	$= [A \,\&\, B]$
D$\underline{\cup}$	$[A] \,\underline{\cup}\, [B]$	$= [A \vee B]$
D$\underline{-1}$	$[A]^{\underline{-1}}$	$= [\neg A]$
D$\underline{0}$	$\underline{0}$	$= [A \,\&\, \neg A]$
D$\underline{1}$	$\underline{1}$	$= [A \vee \neg A]$

Show that the structure <U, dass U, $\underline{\cap}$, $\underline{\cup}$, $\underline{\neg}$, $\underline{0}$, $\underline{1}$> is a boolean algebra. This algebra is known as the *Lindenbaum algebra of* T *in* L, 'LB(T,L)' for short.

(iii). Show the following:

(a) [A] is an atom of LB(T,L) iff $Cl(T \cup \{\neg A\})$ is a complete consistent theory.

(b) LB(T,L) consists of exactly two elements iff T is a complete and consistent theory of L

(c) Let $L_o = \{\}$ be the language of first order logic which contains no non-logical constants whatever. Let V be any logically valid sentence of L_o.
Then the atoms of LB(V,L_o) are the equivalence classes $[\neg B_n]$ of the sentences B_n, which assert that there are exactly n individuals.

(iv). Give an example of a language L and theory T such that LB(T,L) is finite but consists of more than two elements.

20. T_1 and T_2 are theories of some first order language L.

Show: (i) $T_1 \cap T_2$ is a theory of L.
(ii) $T_1 \cup T_2$ is a theory of L iff either $T_1 \subseteq T_2$ or $T_2 \subseteq T_1$.

21. L is a language of first order predicate logic. recall that by $\mathbb{T}_L$ we understand that theory of L which consists of all and only the tautologies of L. Let T be an arbitrary theory of L. We define:

$T^{\perp} = \bigcap\{ T' : T'$ is a theory of L and $T \cup T'$ is inconsistent$\}$

$T_{\perp} = \bigcup\{ T' : T'$ is a theory of L and $T \cap T' = \mathbb{T}_L\}$.

Show: (i) $T^{\perp}$ and $T_{\perp}$ are both theories of L.

(ii) $T_{\perp} \subseteq T^{\perp}$.

(iii) For any theory T of L there are the following two possibilities:

(a) T is finitely axiomatisable. Then there is a sentence A such that A axiomatises T,
$T_{\perp} = T^{\perp} = Cl(\neg A)$ and $T \cup T_{\perp} \models \perp$

(b) T is not finitely axiomatisable.
Then $T_{\perp} = T^{\perp} = \mathbb{T}_L$ but not $T \cup T_{\perp} \models \perp$.

22. Let M be a model for a language L and let N be the following class of models for L: $N = \{M' : (\exists A)(M \models A \ \& \ M' \models \neg A\}$. Let $\varnothing_L$ be the set of all tautologies of L.
Show: $Th(N) = \varnothing_L$ iff Th(M) is not finitely axiomatisable.

23.· Let L be some language for predicate logic let X be some denumerably infinite set and let K be the set of all finite models M for L with $U_M \subseteq X$. Let T be the theory Th(K). Prove that T has infinite models.

24. Let L_1 be the language $\{0, S, <, c_1\}$ of first order predicate logic, in which 0 and c_1 are individual constants, S is a 1-place predicate constant and < is a 2-place predicate; and let L_2 be the language $L_1 \cup \{c_2\}$, where c_2 is some individual constant not in L_1.

Let T_1 be the theory of L_1 which is axiomatised by A1-A6 and let T_2 be the theory of L_2 which is axiomatised by A1-A7.

A1. $(\forall x)\ (x \neq 0 \leftrightarrow\ (\exists y)\ x = Sy)$
A2. $(\forall x)(\forall y)\ (Sx = Sy \rightarrow\ x = y)$
A3. $(\forall x)(\forall y)\ (x < y \ \rightarrow\ \neg\, y < x)$
A4. $(\forall x)(\forall y)\ (\forall z)((x < y \ \&\ y < z \ \&\ x < z) \ \rightarrow\ x < z)$
A5. $(\forall x)(\ x \ < \ Sx)$

A6. $\mathbf{n}_0 < c_1$, for n = 1,2,3, ... ,
where for any natural number n, $\mathbf{n}_0$ is the term "SS....S0", consisting of a "0" followed by n occurrences of "S".
(Thus A6 is an axiom schema which consists of an infinite number of individual axioms, one for each n.)

A7 $\mathbf{n}_{c1} < c_2$, for n = 1,2,3, ... ,
where for natural numbers n, the term $\mathbf{n}_{c1}$ is defined just as $\mathbf{n}_0$ except that its first symbol isn't "0" but "c_1".
(So A7 also consists of an infinity of axioms.)

Show that

(i) T_1 and T_2 are both consistent.
(ii) T_1 and T_2 only have infinite models.
(iii) There exists a model M for the language L_1 such that

(a) M verifies all the sentences of T_1.
(b) There is no expansion M' of M to the language L_2 which verifies all the sentences of T2.

25. Let L be the language {f} of first order predicate logic, with f a 2-place function constant. Let Γ be the set consisting of the following five sentences B1-B5.

B1. $\forall x \forall y\ (f(x,y) = f(y,x))$

B2. $\forall x \forall y\ (f(x,y) = x \ \vee \ f(x,y) = y)$

B3. $\forall x \forall y \forall z\ ((f(x,y) = x \ \& \ f(y,z) = y) \rightarrow f(x,z) = x)$

B4. $\forall x \forall y\ (f(x,y) \neq y \rightarrow \exists z\ (f(x,z) \neq z \ \& \ f(z,y) \neq y))$

B5. $\exists x \exists y\ x \neq y$

Show that Γ has an infinite model but no finite models.

(Hint: A function which satisfies the axioms B1-B4 defines a weak linear order $\leq$: $x \leq y$ iff_{def} $f(x,y) = x$.)

Solutions to some of the exercises to Ch. 1.

2. (Finite and Infinite)
Suppose that X, Y and Z are sets and that $X \sim Y$. Prove:

(i) $X \preccurlyeq Z$ iff $Y \preccurlyeq Z$;
(ii) $Z \preccurlyeq X$ iff $Z \preccurlyeq Y$;
(iii) $X \prec Z$ iff $Y \prec Z$;
(iv) $Z \prec X$ iff $Z \prec Y$.

4. Suppose that X, Y and Z are sets, that $Y \preccurlyeq X$ and that $X \cap Z = \varnothing$.
(i) Show: $Y \cup Z \preccurlyeq X \cup Z$.
(ii) Show that the condition that $X \cap Z = \varnothing$ cannot be dropped.

5. Suppose that X and Y are finite sets. Show that $X \cup Y$ is finite.

2. (Finite and Infinite)
Suppose that X, Y and Z are sets and that $X \sim Y$. Prove:

(i) $X \preccurlyeq Z$ iff $Y \preccurlyeq Z$;
(ii) $Z \preccurlyeq X$ iff $Z \preccurlyeq Y$;
(iii) $X \prec Z$ iff $Y \prec Z$;
(iv) $Z \prec X$ iff $Z \prec Y$.

Solution to (2.iii). Suppose that $X \sim Y$ and $X \prec Z$. Let h be a bijection from Y to X. We first show that $Y \preccurlyeq Z$. Let f be an injection from X into Z. Then $h \circ f$ is an injection of Y into Z. Secondly, suppose that $Z \preccurlyeq Y$. Then there is an injection g from Z into Y. But then $g \circ h$ is an injection of Z into X, which contradicts the assumption that $X \prec Z$. So there can't be an injection of Z into Y. So $\neg(Z \preccurlyeq Y)$. Putting the two conclusions together we get: $Y \prec Z$.

3. Suppose that Y is a finite set. Show:

(i) If $X \subseteq Y$, then X is fintie
(ii) If $X \preccurlyeq Y$, then X is finite.

Solution to (3.i). Suppose X were infinite. Then there would be a bijection f from some proper subset Z of X to X. Let g be the union of f and the identity function on Y \X. Then g is a bijection from Z $\cup$ (Y \ X) to Y. Since Z is a proper subset of X, Z $\cup$ (Y \ X) is a proper subset of Y. So Y would infinite, contrary to assumption.

Solution to (3.ii). Let f be an injection of X into Y. Suppose that X were infinite. Then there would be a bijection g from X to some proper subset Z of X. Then $f^{-1} \circ g \circ f$ to Z is a bijection from f[X] to the set $(f^{-1} \circ g \circ f)[f[X]]$. Since $(f^{-1} \circ g)[Y] = Z$ is a proper subset of X, $(f^{-1} \circ g \circ f)[f[X]]$ is a proper subset f[X]. So Y would have an infinite subset, contradicting (3.i).

4. Suppose that X, Y and Z are sets, that $Y \preccurlyeq X$ and that $X \cap Z = \varnothing$.

(i) Show: $Y \cup Z \preccurlyeq X \cup Z$.

(ii) Show that the condition that $X \cap Z = \varnothing$ cannot be dropped.

Solution to (4.i). Let f be an injection of Y into X. Let g be the union of f and the identity function on Z \ Y. Then, since $X \cap Z = \varnothing$, g is 1-1. Furthermore DOM(g) = $Y \cup Z$ and RAN(g) $\subseteq X \cup Z$.

5. Suppose that X and Y are finite sets. Show that $X \cup Y$ is finite.

Solution to (5.ii). Assume that both X and Y are finite. Suppose that
$X \cup Y$ is infinite. Then there is a proper subset Z of $X \cup Y$ and a bijection f of $X \cup Y$ to Z. Since Z is a proper subset of $X \cup Y$, there is a $u \in X \cup Y$ which does not belong to Z. Since $u \in X \cup Y$, $u \in X$ or $u \in Y$. Suppose that $u \in X$. Define for n = 1,2,.. $f^n(u)$ as follows:

(i) $f^0(u) = u$
(ii) $f^{n+1}(u) = f(f^n(u))$

Consider the set $\{f^n(u): n \in \mathbb{N}\}$. We distinguish two possibilities:

(a) for infintely many n $f^n(u) \in X$;
(b) there is an n such that for all m > n $f^m(u) \in Y$.

First consider case (b). Let n be a number instantiating ther existential statement (b) and let Y' = $\{f^m(u); m > n\}$. Then it is easily verified that f is a bijection from Y' to its proper subset Y' \ $\{f^{n+1}(u)\}$. This contradicts the assumption that Y is finite.

Next we consider case (a). Let X' = $\{f^n(u): n \in \mathbb{N}\} \cap X$. Define the function g on X' by the condition that if x ε X', then g(x) is that element x' such that (i) x' ε X, (ii) x' = $f^n(x)$ for some n, and (iii) there is no positive m < n such that $f^m(x)$ ε X. Then g is a bijection from X' to the proper subset X' \ {u} of X'. This contradicts the assumption that X is finite.

(The following 'solution' to (5) is not correct. What is the mistake?

<u>'Solution' to (5).</u> Assume that both X and Y are finite. Suppose that X ∪ Y is infinite. Then there is a proper subset Z of X ∪ Y and a bijection f of X ∪ Y to Z. Since Z is a proper subset of X ∪ Y, there is a u ε X ∪ Y which does not belong to Z. Since u ε X ∪ Y, u ε X or u ε Y. Suppose that u ε X.

First assume f(u) ε X. Note that f[X] = (f[X] ∩ X) ∪ (f[X] ∩ (Y \ X)). So X = f^{-1}[(f[X] ∩ X)] ∪ f^{-1}[(f[X] ∩ (Y \ X))]. Put X_1 = f^{-1}[(f[X] ∩ X)] and X_2 = f^{-1}[(f[X] ∩ (Y \ X))]. Clearly, $X_1 \cap X_2 = \varnothing$ and $X_1 \cup X_2 = X$. define the function g on X as follows: (i) for x ε X_1, g(x) = f(x); (ii) for x ε X_2, g(x) = x. Then DOM(g) = X, RAN(g) ⊆ X and g is 1-1. Moreover, u ε (X \ g[X]); thast is, g maps X 1-1 onto a proper subset of X. But this contradicts the assumption that X is finite.

Now suppose that f(u) does not belong to X. So f(u) ε Y. If f[Y] ⊆ Y, then we are done. For then f[Y] is a proper subset of Y, since f(u) ε Y \ f[Y], and thus f restricted to Y is a bijection from Y to a proper subset of Y. So we may assume that it is not the case that f[Y] ⊆ Y. So there is a y ε Y such that f(y) ε X. Let f' be the function which is like f except that it switches the values of u and y. (That is: f'(u) = f(y), f'(y) = f(u) and for all v ε X ∪ Y such that v ≠ u and u ≠ y, f'(v) = f(v).) Then we have that f'[X ∪ Y] = f[X ∪ Y] = Z, u does not belong to f'[X ∪ Y] and f'(u) ε X. This reduces the second case to the first. The cxase where u ε Y is com,pletely parallel to that where u ε X.)

Appendix.

Soundness and Completeness for the Method of Proof by Semantic Tableaus.

The proofs of soundness and completeness that were given earlier in this Chapter concern the axiomatic deduction system presented in Section 1.1.3. The completeness proof is fairly involved and this is so for one thing because it requires showing for a substantial number of logical theorems that they can be derived from the given axioms. To make this task somewhat easier and less tedious a proof was given early on of the Deduction Theorem. But that proof involves complications of its own. Most of these various complications leave one with a feeling that they are peripheral to the central ideas of the completeness proof as it is given in 1.1.3 and nourish the wish for a proof that circumvents them.

This Appendix offers, as an alternative to the proofs of 1.1.3, proofs of soundness and completeness for the method of demonstration by semantic tableau construction. In some ways these proofs are easier, since the Tableau Method is, by conception and general architecture, much closer than the axiomatic method to the semantic conception of logical consequence with which it has to be shown equivalent. For after all, proving validity for an argument by the Tableau Method is nothing other than showing that an attempt to find a counterexample for it necessarily fails. (Furthermore, proving soundness and completeness for the Tableau Method is natural for most of those for whose benefit these notes have been produced, since the tableau method is the principal deduction method with which they were familiarised in the logic course that standardly serves as prerequisite for the present one.)

Unfortunately, proving soundness and completeness for the Tableau Method isn't quite as straightforward as one might have hoped, in spite of the fundamentally semantic conception on which the method is based. This is because as soon as one sits down to define them with mathematical rigour semantic tableaux prove to be fairly complex data structures - much more so than the remarkably simple formal objects that are axiomatic derivations. (Recall that these are strings of formulas which satisfy a small number of simple and easily verifiable conditions.) So some of the benefit that one gains from the close connections between the Tableau Method and the notions of truth in a model and logical consequence is lost because of by the need to manipulate these more complex structures. Still, it would seem to me that on balance the completeness proof below is simpler and more natural than the one given in Section 1.1 of this Chapter.

In what follows familiarity with the use of semantic tableaux will be assumed. Nevertheless, as a preliminary to the formal treatment of the Tableau Method, we begin with an informal summary of the important features of this method.

Semantic tableaux are structures that are built from sentences of some particular language L of First Order Predicate Logic. The sentences occur in either one of two columns, the 'TRUE' column and the 'FALSE' column. To prove the validity of an argument with premises $A_1,..., A_n$ and conclusion B one starts with a tableau in which $A_1,..., A_n$ are entered under 'TRUE' and B is entered under FALSE. Rules are then applied to these sentences and to the ones which result from earlier rule applications until, roughly speaking, only atomic sentences are left. In the course of these rule applications the tableau may split into different 'branches', each with its own pair of sets of 'TRUE' and 'FALSE' formulas. A branch is *closed* if it contains the same sentence under both TRUE and FALSE; and the semantic tableau as a whole is *closed* if each of its branches is closed.

The purpose of constructing a semantic tableau for an argument $<A_1,..., A_n \mid B>$, with premises $A_1,..., A_n$ and putative conclusion B is to try and construct a countermodel for it, i.e. a model M in which $A_1,...,$ A_n are true and B is false. This succeeds iff the construction produces a tableau branch in which all reduction operations have been carried out and in which there are no explicit conflicts, of the kind that arises when the same sentence occurs both under TRUE and under FALSE. A conflict-free tableau branch in which no further reductions can be carried out will provide a counter-model for the argument, and thereby establish its non-validity.

From the present point of view a tableau all of whose branches are closed is to be considered a failure: it doesn't provide the counter-model which was the aim of its construction. However, there is also another point of view from which it is precisely tableau closure that should be seen as a success. Failure to find a counter-model this way, which manifests itself as closure of all branches of the tableau, has the status of a proof that no counter-model exists, and thus that the argument is valid. This is so because tableau construction is a fully systematic search for counterexamples - one in which 'no stone is left unturned', so to speak. That the Tableau Method is exhaustive in this strict sense, however, is not immediately obvious and is itself in need of a formal demonstration. So this is one of things we will have to prove in this Appendix. (In the syllabus for LFG II the result followed from the conversion of the tableau method into the method of proof by deduction in the Sequent Calculus.)

This description of the Tableau Method might give the impression that more or less all the work that is needed to establish soundness and completeness of the predicate calculus has already been done: Either the semantic tableau for $\langle A_1,..., A_n \mid B\rangle$ is closed (i.e. all its branches are closed) and then the argument is valid. or else the tableau has at least one branch which is not closed and then there is a counter-model; *tertium non datur*. We can rephrase this in the words of principle (P1):

(P1) An argument is valid iff a semantic tableau constructed for it is closed.

(P1) combines (a) the soundness and (b) the completeness of the Tableau Method: For an argument $\langle A_1,..., A_n \mid B\rangle$ to be valid it is (a) sufficient and (b) necessary that its semantic tableau is closed.

What has just been said constitutes the gist of the proof of soundness and completeness of the Tableau Method. But turning these intuitive ideas into a proper mathematical argument requires some real work.

To begin, let us list the three propositions for which explicit proofs are needed:

PR1. If the tableau for the argument $\langle A_1,..., A_n \mid B\rangle$ closes, then $\langle A_1,..., A_n \mid B\rangle$ has no counter-model (and thus is semantically valid).

PR2. When the tableau for $\langle A_1,..., A_n \mid B\rangle$ has an open branch, then $\langle A_1,..., A_n \mid B\rangle$ has a counter-model (and thus is invalid).

PR3. Every complete tableau (i.e. one in which all possible reductions have been carried out) is either closed or it has at least one open branch.

At first blush PR3 may seem a tautology. It isn't quite that, however, since complete tableaus can be infinite. In fact, infinite. non-closing tableaux are far more common than finite ones. It is for infinite tableaux that PR3 is not altogether self-evident. Its demonstration rests on some (modest) combinatorial properties of set theory.

Tableau construction involves the application of rules to 'reducible' sentences occurring in the tableau. The reduction rules are fully determined by three factors:

(i) the form of the sentence to which the rule is applied. What rule is applied is determined by the operator (connective or quantifier) which has widest scope in the sentence;

(ii) the question whether the sentence occurs under 'TRUE' or 'FALSE';

(iii) (for the quantifier rules) which parameter is to be used in reducing the outer quantifier of the sentence.

(iii) points to one important feature of tableau construction for arguments of predicate logic, viz the substitution of 'parameters' for variables bound by outer quantifiers. In some cases the parameters used belong to the tableau already, but in others they are (and must be) introduced by the reduction operation in question. It is in this way that the universes are constructed for the counter-models that are determined by open tableau branches.

Here are schematic presentations of all the tableau rules for First Order Predicate Logic with Identity:

(8)

(¬,T) TRUE FALSE

<u>¬ C</u> || <u>   </u>
|| C

TRUE FALSE (¬,F)

<u>   </u> || <u>¬ C</u>
C ||

(v,T) TRUE FALSE

<u>C v D</u> || <u>   </u>
C | D

TRUE FALSE (v,F)

<u>   </u> || <u>C v D</u>
C
D

(→,T) TRUE FALSE TRUE FALSE
(→,F)

<u>C → D</u> || <u>   </u>
| D C |

<u>   </u> || <u>C → D</u>
C D

($\leftrightarrow$,T) TRUE FALSE

$$\begin{array}{c|c} \underline{C \leftrightarrow D} & \underline{\qquad\qquad} \\ C \mid & \mid C \\ D & D \end{array}$$

($\leftrightarrow$,F) TRUE FALSE

$$\begin{array}{c|c} \underline{\qquad\qquad} & \underline{C \leftrightarrow D} \\ C \mid D & D \mid C \end{array}$$

($\forall$,T) TRUE FALSE

$$\begin{array}{c|c} (\forall v_i)\ A & \\ \overline{A(t/v_i)} & \overline{\qquad} \end{array}$$

(t an arbitraryclosed term)

TRUE FALSE ($\forall$,F)

$$\begin{array}{c|c} & (\forall v_i)\ A \\ \overline{\qquad\qquad} & \overline{A(c/v_i)} \end{array}$$

(c a new parameter)

($\exists$,T) TRUE FALSE

$$\begin{array}{c|c} (\exists v_i)\ A & \\ \overline{A(c/v_i)} & \overline{\qquad} \end{array}$$

(c a new parameter)

($\exists$,F) TRUE FALSE

$$\begin{array}{c|c} & (\exists v_i)\ A \\ \overline{\qquad\qquad} & \overline{A(t/v_i)} \end{array}$$

(t an arbitraryclosed term)

(=,Sub) TRUE FALSE

$$\begin{array}{c|c} s = t & \\ A & \\ \overline{A'} & \overline{\qquad} \end{array}$$

TRUE FALSE

$$\begin{array}{c|c} s = t & \\ & A \\ \overline{\qquad\qquad} & \overline{A'} \end{array}$$

(s, t arbitrary closed terms; A is an atomic formula and A' is the result of properly substituting t for one occurrence of s in A).

(=, Ref) TRUE FALSE

$$\begin{array}{c|c} \overline{t = t} & \overline{\qquad} \end{array}$$

(t an arbitrary closed term)

Although familiarity with the Tableau Method is assumed, it may be helpful to present a couple of tableau constructions as examples. This will also help us to focus more sharply on the tasks that lie ahead. The tableau constructions we will consider are those for the two arguments that we get by taking as premise and conclusion the standard formalisations in First Order Logic of the two possible scope readings of a sentence like (2)

(2) Some book about semantics has been read by every student.

Abbreviating 'student' as P, 'book' as Q and 'y has been read by x' as R(x,y), we get as formalisations for the two readings:

(3) i. $(\forall x)(P(x) \rightarrow (\exists y)(Q(y) \ \& \ R(x,y)))$
ii. $(\exists y)(Q(y) \ \& \ (\forall x)(P(x) \rightarrow R(x,y)))$

Thus the two arguments are:

(4) i. $< (\exists y)(Q(y) \ \& \ (\forall x)(P(x) \rightarrow R(x,y))) \mid$
$(\forall x)(P(x) \rightarrow (\exists y)(Q(y) \ \& \ R(x,y))) >$
ii. $< (\forall x)(P(x) \rightarrow (\exists y)(Q(y) \ \& \ R(x,y))) \mid$
$(\exists y)(Q(y) \ \& \ (\forall x)(P(x) \rightarrow R(x,y))) >$

Of these (4.i) is valid and (4.ii) is not. The following two tableaus show this.

(5) (Tableau for (4.i))

```
                    TRUE                  ||                   FALSE

 (∃y)(Q(y) & (∀x)(P(x) → R(x,y)))        (∀x)(P(x) → (∃y)(Q(y) & R(x,y)))

   Q(a) & (∀x)(P(x) → R(x,a))
    Q(a)
  (∀x)(P(x) → R(x,a))                    P(b) → (∃y)(Q(y) & R(b, y))
    P(b)                                 (∃y)(Q(y) & R(b,y))

                                         Q(a) & R(b,a)

                  |                                            |
                  |                                            |
                  |                      Q(a)                  |        R(b,a)
=========== |                     ||================|
                   P(b) → R(b,a))       ||
                                        ||
                 |            | R(b,a)                          |  P(b)   |
                 |====== |=========||                          |=======|=====
```

Since this tableau closes, we conclude that (4,i) is valid.

(6) (Tableau for (4.ii))

```
                    TRUE                                       FALSE

  (∀x)(P(x) → (∃y)(Q(y) & R(x,y)))      (∃y)(Q(y) & (∀x)(P(x) → R(x,y)))

    P(b) → (∃y)(Q(y) & R(b,y))
                        | (∃y)(Q(y) &  ||          P(b)                |
                        |       R(b,y)) ||                              |
                        |               ||  Q(b) & (∀x)(P(x) →          |
________________________|_______________||_________________R(x,b))______|___
             |          |               ||  Q(b)    | (∀x)(P(x) →       |
             |          |               ||          |      R(x,b))      |
```

The leftmost branch of this tableau is open. It determines the extremely simple counter-model defined in (7) and thereby shows that the argument is invalid.

(7) (Countermodel to (4.ii))

$U_M = \{b\}$
$P_M = \varnothing$
$Q_M = \varnothing$
$R_M = \varnothing$

(6) is an example of a tableau with a finite open branch in which no further reductions are possible. Besides such tableaux and tableaux in which all branches close there are also those in which there are open branches, but which have no finite open branches without further reduction possibilities. It is these tableaux that are responsible for the fact that the semantic tableau method is not a decision method for validity. (Which is as it should be, since we know that there cannot be such a decision method).

Tableaux with branches that do not close but which offer reduction option at all finite stages of their construction are very common. Perhaps the simplest example of such a tableau is that for the argument $<(\forall x)(\exists y)R(x,y),\varnothing>$. This tableau has no splittings, and its one branch never closes lalhough its construction can be continued indefinitely. The first stages of its construction are given in (8).

(8)

TRUE	FALSE
$(\forall x)(\exists y)R(x,y)$	
$(\exists y)R(a,y)$	
$R(a,b)$	
$(\exists y)R(b,y)$	
$R(b,c)$	
$(\exists y)R(c,y)$	
.	
.	

It is plain how this tableau construction will continue and equally plain that a closure is not in the making. But in general things are not so straightforward. Indeed, it follows from the fact that there is no decision procedure for validity in first order logic that there can't be an algorithm that will tell us when we may stop with the construction of a tableau branch on the grounds that if closure hasn't yet been reached so far, it won't be achieved at any later stage either.

Exercise.

a. For the formula $(\forall x)(\exists y)R(x,y)$ we can find finite models (and thus there are finite countermodels to the argument $< (\forall x)(\exists y)R(x,y) \mid \emptyset >$).

Task: Define a "minimal" model of $(\forall x)(\exists y)R(x,y)$, i.e. one in which the universe has as few elements as possible.

b. However there are also formulas that have models but *only* infiinite ones.

Task: Give one such formula and define a model (necessarily with infinite universe) in which the formula is true.

In order to be able to provide exact proofs of soundness and completeness we need a more rigorous definition of semantic tableaux and their construction than are provided by the semi-formal descriptions of the Tableau Method which suffice for most purposes (such as the description given in the syllabus for LFG II). In the formal definition of semantic tableaux that we will give below it will be convenient to mark the distinction between formulas occurring under TRUE and formulas occurring under FALSE directly on the formulas themselves. That is, we will define semantic tableaux in such a way that each tableau branch will be a set of pairs $<A,T>$ and $<A,F>$, where the A's are sentences and T and F are used to indicate whether A occurs in the TRUE or the FALSE column of the given branch. This means in particular that a branch is to be considered closed if for some sentence A both $<A,T>$ and $<A,F>$ belong to it. We will refer to pairs $<A,T>$ and $<A,F>$ as *positively* and *negatively signed* formulas, respectively, or simply as *signed* formulas.

We also need a formal characterisation of the branching structure of semantic tableaux. To this end we represent semantic tableaux as trees (in the mathematical sense of the term), i.e. as sets of nodes that are connected by a partial order which has the following additional properties (which are distinctive of tree orderings). That is, a *tree* is a strict partial order $<$ such that

(T.i) $(\exists x)(\forall y)(y \neq x \rightarrow x < y)$, and

(T.ii) $(\forall x)(\forall y)(\forall z)((y < x \;\&\; z < x) \rightarrow z < y \vee z = y \vee y < z)$.

In connection with property (T.i), note that it follows from the fact that a tree is a partially ordered set that there is at most one object in the universe which satisfies the free variable formula $(\forall y)(y \neq x \rightarrow x < y)$. This means that when $(\exists x)(\forall y)(y \neq x \rightarrow x < y)$ is true, then there is exactly one such object. This object is called the *root* of the tree.

The nodes of the tree which get created in the course of tableau construction are to be thought of as representing the stages of tree branches which are reached each time a reduction operation is applied to one of the formulas belonging to the given branch.

The trees that arise in the course of tableau construction are thus special in that any given node has either:

(a) two successors; this happens when the reduction rule that is applied to a formula from the set associated with the node leads to a pair of reduction products; this is the case whenever the reduction rule applied is one of (&,R), (v,L), ($\rightarrow$,L), ($\leftrightarrow$,L) or ($\leftrightarrow$,R); or

(b) one successor; this happens when the reduction rule that is applied to a formula belonging to the node leads to a single reduction product, i.e. through an application of one of the remaining rules ((&,l), (v,R), ($\rightarrow$,R), ($\neg$,L), ($\neg$,R), ($\forall$,L), ($\forall$,R). ($\exists$,L) or ($\exists$,R)); or

(c) no successor; this situation arises when either (i) all possible formula reductions in the branch to which the node belongs have been carried out, or else (ii) because the node represents that stage of its branch Z at which closure of Z is achieved.

It will be useful to adopt a special mode of representation for the kinds of trees we will be needing. This mode doesn't cover all tree-like orderings defined above, but it will cover all those we want, and it has the advantage that the partial order is exhaustively characterised by the internal structure of the nodes. The nodes of the trees in question are finite sequences of 0's and 1's. and the ordering relation holds between two such nodes s and s' if and only if s is a proper initial segment of s'. We include the empty sequence <> among the possible tree nodes. Since trees will be defined as non-empty node sets closed under initial segments, this means that <> will be member of every tree. and it will always be its root.

In any tree T of the kind described each node s will have either 0, 1 or 2 immediate successors. s will have two successors in T if both $s^\frown 0$ and $s^\frown 1$ belong to T and it has no successor in T if neither of these belong to T.[32] In the third case, where s has one successor, it could be that this successor is either $s^\frown 0$ or $s^\frown 1$, but to make things as tight as possible we want to exclude the second of these cases. In other words, the successors of s in T will always be one if the following three sets: $\varnothing$, $\{s^\frown 0\}$, $\{s^\frown 0, s^\frown 1\}$. We summarise these stipulations in the following definition.

<u>Def. DA1.</u> (Trees)

A tree *T* is a pair $\langle T, \leq \rangle$, where

(a) T is a non-empty set of sequences of 0's and 1' satisfying the following two conditions:

(i) if $s^\frown 1 \in T$, then $s^\frown 0 \in T$,
(ii) if $s^\frown 0 \in T$, then $s \in T$;

(b) for any nodes s, s' $\in$ T, s $\leq$ s' iff s is a proper initial segment of s'.[33]

N.B. since the ordering relation of a tree *T* = $\langle T, \leq \rangle$ is fully determined by the internal structure of its nodes, we will henceforth identify *T* with its node set T.

The *branches* of a tree T are its maximal linearly ordered subsets. For trees of the kind we are using here this means that if Z is a branch of T and s and s' are nodes in Z then either s = s' or s is a proper initial segment of s' or s' is a proper initial segment of s.

---

32 By $s^\frown n$ we understand the concatenation of s and n, i.e. the result of adding n on to the end of s; so if s is $\langle s_1, ..., s_i \rangle$, then $s^\frown n$ is the sequence $\langle s_1, ..., s_i, n \rangle$

33 Here it is assumed that every sequence counts as an initial segment of itself. Thus $\leq$ is reflexive, and thus as weak partial order, as the symbol '$\leq$' suggests.

A semantic tableau for an argument $<A_1,\ldots,A_n \mid B>$, where the premises $A_1,\ldots,A_n$ and the conclusion B are formulas of some first order language L, is to be thought of as a tree whose nodes are 'decorated' with the information that makes each node into a stage of a tableau construction for this argument. We represent this information by means of a *decoration function*. This is a function which is defined on the nodes of the tree and maps each node onto the information that is to be associated with it.[34] In particular, our semantic tableaux will be defined as decorated trees of certain special sort. More precisely, we will define a semantic tableau as a decorated tree <T,D> in which the decorating function D associates with each node s of T information about which sentences have been included under 'TRUE' at the tableau construction stage identified by s and which have been included under 'FALSE'.

There is an additional feature of semantic tableaux which a mere association of sets of 'true' and 'false' sentences with nodes of the tree may seem to overlook. This is the set of *parameters* which have been introduced into a tableau branch at any one stage of its construction. We recall that parameters are individual constants and that the origin of an individual constant c in a tableau for an argument $<A_1,\ldots,A_n \mid B>$ can be of two kinds: either c occurs somewhere in $A_1,\ldots,A_n$ or B or else c has been introduced (as a 'parameter') in the course of the construction of the tableau through the application of reduction rules applying to quantified formulas. In general these new parameters cannot be assumed to belong to the language L of the argument $<A_1,\ldots,A_n \mid B>$, so their introduction into the tableau means that the tableau, conceived as a structure involving formulas of some first order language L, is strictly speaking no longer a tableau for the language L but rather for some extension L' of L, which is obtained by adding new individual constants to L. In keeping with this observation we assume that before the construction of the tableau for an argument $<A_1,\ldots,A_n \mid B>$, with premises and conclusion belonging to L, is started, L is extended with an infinite sequence $c_1, c_2, ..$ of new constants. From this set the parameters that are needed in course of the tableau construction will then be drawn.

---

[34] Combinations consisting of some abstract mathematical support structure *S* and a function which assigns certain items to each of the elements of S are often referred to as *decorated structures*. *Decorated trees* are a special case of decorated structures in general, but it seems that they are the kind that is used most often.

Strictly speaking the set of constants that have been introduced at the point of tableau construction identified by a tree node s can be recovered from the formulas associated with s by the decoration function. For these constants are just the ones which have occurrences in those formulas. However, it will be convenient to define the tableau construction process in such a way that the set of constants that have been introduced at any stage in any branch is explicitly available and directly accessible.

We need access to the information what constants have already been introduced before a certain stage s of the tableau construction whenever the sentence that is up for reduction at s is either of the form $(\forall x)E(x)$ and occurring under TRUE or of the form $(\exists x)E(x)$ and occurring under FALSE. Reduction of such a formula is required iff there is a constant c that has been introduced at some stage before s with which the formula has not been instantiated before. (That is, E(c) has not yet been added to the TRUE c.q. FALSE column.) Having the sequence of previously introduced constants as a separate item in the decoration of s makes it easier to state whether and how reduction of such a formula is to be executed at s.

There is also another piece of information that we need in order to make the right decisions with regard to such formulas. It could be the case that the formula has in fact been previously instantiated with a given constant c, but that the formula E(c) to which this instantiation led is no longer available at s as a witness to this fact. For E(c) might itself have been a complex formula and might have been reduced in its turn at some stage before s. Therefore it is desirable to keep an explicit record in some other form of what instantiations have already been carried out. The simplest way to do this is to attach to formulas of the kind at issue besides a feature that tells us under which of the two columns they occur also the set of constants with which they have already been instantiated.

This additional piece of information sets the formulas in question apart from all other cases. In the other cases the column in which the formula occurs is all the information about their status in the given tableau branch that we need; for the cases under discussion the set of instantiated constants is needed as well. This distinction is built into the following definition of the notion of a *signed formula*. (The signed formulas will be the items that go into the decorations of the tree nodes.)

<u>Def. DA2.</u> (of *signed formula*)

A *signed formula* of L is either:

(i) a pair <A,π>, where A is a sentence of L, $\pi \varepsilon \{T,F\}$, and neither of the following two conditions (a), (b) holds:

(a) π = T and A is of the form $(\forall x)E(x)$
(b) π = F and A is of the form $(\exists x)E(x)$

or

(ii) a triple <A,π,S>, where A, π are as under (i), one of the conditions (a), (b) obtains and S is a (possibly empty) set of individual constants.

One last point before we come to our formal definition of semantic tableaux. We want the construction of semantic tableaux to be fully deterministic: at every stage the form of the tableau at that stage should fix unequivocally which reduction, if any, is to be performed next and how it is to be carried out. This requires that the (signed) formulas that are part of the decoration of any stage s are given in some particular order. We will assume, moreover, that this is also the case for the constants that have already been introduced into the tableau (although here an ordered presentation isn't absolutely necessary). In other words, the decoration D(s) of a tree node (= tableau construction stage) s will consist of a pair of two finite sequences, the first consisting of signed formulas and the second of individual constants.

For languages with function constants of one or more argument places tableau construction is complicated by the fact that instantiation of formulas of the form $(\forall x)E(x)$ under TRUE and $(\exists x)E(x)$ under FALSE may be needed not only for individual constants, but also for the complex terms that can be built from these constants with the help of function constants of L of one or more argument places. (For instance, if c is an individual constant and f a 1-place function constant, instantiation will in general be required not just with c but also with the terms f(c), f(f(c)), .. and so on.) To carry through the formalisation of tableaux and their construction and the proofs of soundness and completeness based upon that formalisation for languages with function constants doesn't encounter any fundamental obstacles, but it presents extra complications which detract from the central points of the proof. We will therefore restrict attention to languages L without function constants of one or more argument places. The general case, in which L may contain such constants, can be reduced to the one we will consider by translating formulas with such function constants into formulas with corresponding predicate constants; see Exercise EA2 below.

In fact, we will initially restrict the language L even further, by also excluding 0-place function constants (i.e. individual constants). That is, L won't have any individual constants of its own, and so the only constants occurring in a semantic tableau for an argument whose premises and conclusion belong to L will be those introduced in the course of its construction. Finally, as our third initial restriction, we will assume that = occurs neither in the premises nor the conclusion of the arguments we will consider. Note that this entails that = won't occur anywhere in the tableaux for these arguments.

We are now ready for a formal definition of the notion of a s*emantic tableau for* an argument $<A_1,..., A_n \mid B>$. Note well that what will be defined is the notion of a *completed* tableau, i.e. a tableau in which all possible reductions have been carried out. As noted, such tableaux are very often infinite (i.e. they involve an infinite node set T).

Def. DA3 (Formal characterisation of the notion 'Semantic Tableau for an argument $<A_1,..., A_n \mid B>$ in a first order language L)

Let L be a language of First Order Predicate Logic without function constants, $c_1,c_2,..$ an infinite sequence of individual constants not belonging to L, and let $A_1,..., A_n$, B be sentences of L in which = does not occur.
A (*completed*) *semantic tableau for* the argument $<A_1,..., A_n \mid B>$ *given* the sequence $c_1,c_2,..$ is a pair <T,D>, where

(i) T is a tree as defined in Def. DA1 and
(ii) D is a function defined on T which assigns to each node $s \in T$ a pair D(s) consisting of

(a) a finite sequence of signed formulas (see Def. DA2), and
(b) a finite sequence of constants from the sequence $c_1, c_2,..$

(iii) T and D satisfy the conditions specified below.

Before we set about describing these conditions, first a notational convention. For any node s of T we refer to the first component of the pair D(s) (the sequence of signed formulas) as '$D(s)_F$' and to the second component (the sequence of individual constants) as '$D(s)_C$'.[35]

The conditions alluded to under (iii) recapitulate, in strictly formal and strictly deterministic terms, the construction of the tableau from its starting point, when the column TRUE consists just of the premises only the premises $A_1,..., A_n$ and the column FALSE just of the conclusion B.

Our first condition concerns this starting point; it specifies the decoration of the root <>. But before we can state it in the form in which it will be most useful later on, there is one further aspect of tableau construction that we must make explicit. As our examples illustrate, there are in essence two reduction rules for quantified formulas. Reduction of existential formulas under TRUE and universal formulas under FALSE requires replacement of the variable that is bound by the quantifier by a new constant, which does not yet occur in the tableau that is being constructed; and such reductions have to be performed only once. Reductions of existential formulas under FALSE and universal formulas under TRUE, on the other hand, involve constants that have been introduced already. These are reductions that have to be repeated again and again to the same formula, in order to make sure that all constants occurring in the tableau branch to which a given quanified formula belongs are substituted for the bound variable of its outer quantifier eventually.

But there is one exception to the principle that quantified formulas of the second category are only instantiated with constants that have been previously introduced. This is when tableau construction has to be got under way somehow and the only reductions that are possible involve formulas of just this kind. Of the three tableaux that were shown above the second and third are both examples of this. In such cases there is nothing for it but to instantiate one of the quantified formulas with some constant or other, which makes its entry into the tableau in this way. In each tableau such a step needs to be performed at most once. For once one such rule application has occurred and the constant involved in the application has been thereby introduced into the given tableau branch, then from then on tableau construction can proceed in

---

[35] It should be noted that both sequences may in principle be empty. In fact, given the restrictions on L we have adopted here, $D(<>)_C$ will always be empty; $D(<>)_F$ would be empty only when the argument had neither premises nor conclusion. (This, however, is a purely theoretical possibility without any intuitive interest.) It is standard to think of an argument as involving at least a conclusion, even if the premise set of an argument may sometimes be empty.

accordance with the principle that quantifiers of the first kind are instantiated (once) with new constants and quantifiers of the second kind with all and only the previously introduced ones.[36]

It would be possible but awkward if we had to make special provisions for the possibility that tableau constructions may have to start in this particular way. But it is easy to set things up in such a way tzhat no special provisions are needed. It suffices to add one constant to the tableau at the very start of its construction, irrespective of whte form of hte argument for which the tableau is being constructed. Doing this is yet another way of saying that no matter what the (counter) model we are trying to find by constructing the tableau will be like, it will have at least one element (viz. the denotation of this constant) in its universe. As regards the constant we choose for this special role, the most natural choice would seem to be the first constant c_1 from our list; so that is the one we choose.

For the decoration of the root of the tableau this means that the sequence of already introduced constants is not the empty sequence, but the one element sequence $<c_1>$.

With this last bit of informal explanation out of the way we are ready for the exact specification of the decoration of the root .

C_{root} $D(<>) = <<\alpha_1,..., \alpha_n, \beta>, <c_1>>$,
where $\alpha_1,..., \alpha_n$ are signed formulas which establish the premises $A_1,..., A_n$ as occurring in the TRUE column and β is a signed formula establishing B as occurring in the FALSE column.

N.B. that $\alpha_1,..., \alpha_n$ are signed formulas which establish the

---

[36] We recall that the justification for this way of starting tableau constructions is the assumption of classical logic that the universe of discourse is never empty (and thus that models never have empty universes). This means that for instance a universally quantified statement will never be true vacuously, that is, simply because there is nothing at all in the model in which it gets interpreted. Since this possibility of vacuoous truth is excluded in the model theory for classical first order logic, it is always legitimate to instantiate the quantifier of such a statement to a new constant, with which no information about its referent is as yet connected. Instantiating the quantifier in this way is nothing more than making explicit that if the statement is true at all, then there will be at least one thing of which its scope (i.e. the formula to which the quantifier is attached) will be true. The non-empty universe assumption netails that this procedure is sound..

premises $A_1,\ldots, A_n$ as occurring in the TRUE column is to be understood as follows: If A_i begins with a universal quantifier, then α_i has the form $\langle A_i,T,\varnothing\rangle$; otherwise α_i has the form $\langle A_i,T\rangle$. Analogously, β is a signed formula establishing B as occurring under FALSE. That is, β has the form $\langle B,F,\varnothing\rangle$, if B begins with an existential quantifier and otherwise is equal to $\langle B,F\rangle$.

The next two conditions concern the end nodes (or 'leaves') of T. These are the stages s at which either (i) no further formula reductions are needed or (ii) all possible reductions have already been carried out. Case (i) arises when a contradiction (= closure) has been reached in the transition to s. That is, the same formula A occurs in $D(s)_F$ both with the sign T and with the sign F. Given the particular way in which we formalise semantic tableaux here, case (ii), where all possible reductions have been carried out already, manifests itself as follows. As will be described in detail below, all reducible formulas are removed from the decoration when they are reduced except for universally quantified formulas occurring under TRUE and existentially quantified formulas under FALSE. Whether a signed formula of this kind is a candidate for reduction at stage s depends on whether the sequence $D(s)_C$ contains constants that do not occur in the set S that the signed formula contains as its third component. Formally the condition about end nodes can be stated as follows:

C_{leaf} s ε T is an end node of T (in other words, $s^\frown 0$ is not a member of T) iff one the following two conditions (a), (b) is satisfied:

(a) (closure at s)
$D(s)_F$ contains signed formulas α_i and α_j which involve the same formula A but the opposite signs T and F, respectively.

(b) (no further reductions possible)
The only signed formulas in $D(s)_F$ which involve non-atomic formulas are either of the form $\langle(\forall x)E(x),T,S\rangle$ or of the form $\langle(\exists x)E(x),F,S\rangle$, where in each case S contains all the constants occurring in $D(s)_C$.

The remaining conditions concern the relations between the decorations of mother nodes and their daughters. In these cases s is not closed and $D(s)_F$ contains at least one signed formula that is a candidate for reduction. The reduction that is performed will then concern the first such signed formula in $D(s)_F$. The nature of the reduction depends on what kind of signed formula this is, and the precise description of the way in which it is reduced depends on the form of D(s) and relates D(s) to the decorations of the one or two daughters of s. There are as many cases to be distinguished here as there are tableau construction rules (see pp. 93, 94). Strictly speaking it would be necessary to go through each one of those cases separately. We will proceed selectively, however, and leave the majority of the cases as exercises.

We first consider those reductions which lead to a split of the given tableau branch. That is, in these cases s has two daughters, $s^\frown 0$ and $s^\frown 1$. Reductions of this kind arise when the signed formula that is to be reduced has one of the following forms: $<CvD,T>$, $<C\&D,F>$, $<C\rightarrow D,T>$, $<C\leftrightarrow D,T>$ or $<C\leftrightarrow D,F>$. We consider only the first of these possibilities, $<CvD,T>$. In this case the decorations of the successor nodes $s^\frown 0$ and $s^\frown 1$ are obtained by eliminating $<CvD,T>$ from $D(s)_F$ and adding at the end of that sequence a signed formula γ containing C in the case of $s^\frown 0$ and a signed formula δ containing D in the case of $s^\frown 1$. γ is defined as follows: $\gamma = <C,T>$ in case C does not begin with a universal quantifier, and $= <C,T,\varnothing>$ if C does. Likewise for δ. Thus the decorations $D(s^\frown 0)$ and $D(s^\frown 1)$ can be defined as follows:

$C_{(v,T)}$ Suppose the member α_i of $D(s)_F$ that is up for reduction has the form $<CvD,T>$. Then s has the successors $s^\frown 0$ and s $s^\frown 1$ in T, whose decorations are determined as follows:

$D(s^\frown 0) = <<\alpha_1, .., \alpha_{i-1}, \alpha_{i+1}, .., \alpha_n, \gamma>, D(s)_C>$,
$D(s^\frown 1) = <<\alpha_1, .., \alpha_{i-1}, \alpha_{i+1}, .., \alpha_n, \delta>, D(s)_C>$,
where γ, δ are as defined above,.

We now turn to the reductions which do not produce a split. Here we distinguish between three major cases:

(i) the main operator of the reduced formula is a sentential connective:

(ii) the reduced formula either begins with an existential quantifier and occurs under TRUE or begins with a universal quantifier and occurs under FALSE;

(iii) the reduced formula either begins with an existential quantifier and occurs under FALSE or begins with a universal quantifier and occurs under TRUE.

Case (i). In this case the signed formula that is up for reduction is of one of the following forms: $<\neg C,T>$, $<\neg C,F>$, $<C\&D,T>$, $<C \vee D,F>$, $<C \rightarrow D,F>$. This time we only consider the conjunction case. The only difference with condition $C_{(v,T)}$ is that now we have just one successor and both constituents of the reduced formula are added on to the end of the formula decoration of that successor.

$C_{(\&,T)}$ Suppose the member α_i of $D(s)_F$ that is up for reduction has the form $<C\&D,T>$. Then s has one successor, $s^\frown 0$, in T, whose decoration is determined as follows:

$$D(s^\frown 0) = <<\alpha_1, .., \alpha_{i-1}, \alpha_{i+1}, .., \alpha_n, \gamma, \delta>, D(s)_C>,$$

where γ and δ are as defined as in the case of $C_{(v,T)}$.

Case (ii). In cases of this kind reduction involves the introduction of a new parameter c into the given tableau branch. We choose for this parameter the first constant in our fixed sequence $c_1, c_2,..$ that does not occur in $D(s)_C$. We only consider the subcase where the member of $D(s)_F$ that is up for reduction has the form $<(\exists x)E(x),T>$.

$C_{(\exists,T)}$ Suppose the member α_i of $D(s)_F$ that is up for reduction has the form $<(\exists x)E(x),T>$. Then s has one successor, $s^\frown 0$, in T. The decoration of $s^\frown 0$ is given by

$$D(s^\frown 0) = <<\alpha_1, .., \alpha_{i-1}, \alpha_{i+1}, .., \alpha_n, \varepsilon>, D(s)_C {}^\frown c>;$$

here $\varepsilon = <E(c),T,\emptyset>$ if E begins with a universal quantifier and $\varepsilon = <E(c),T>$ otherwise.

(Note that this is the one rule application in which $D(s)_C$ gets extended.)

Case (iii). This case differs from all others in that the reduced formula is not eliminated from $D(s)_F$ but 'recycled' by being added to the end of $D(s)_F$. Also a special check is needed in this case to see whether the formula should be reduced at stage s, and which parameter should be involved in its instantiation. Since we have discussed this issue in considerable detail above, we proceed with the formal specification of the relevant condition right away. We only consider the case where the signed formula that is up for reduction is of the form <(∀x)E(x),T,S>.

C(∀,T) Suppose the member α_i of $D(s)_F$ that is up for reduction has the form <(∀x)E(x),T,S>, that $D(s)_C$ contains at east one member that does not belong to S and that c is the first constant in $D(s)_C$ with this property. Then s has one successor, $s^\frown 0$, in T and the decoration of $s^\frown 0$ is given by

$$D(s^\frown 0) = \langle\langle \alpha_1,.., \alpha_{i-1}, \alpha_{i+1},.., \alpha_n, \varepsilon, \langle(\forall x)E(x),T,S\cup\{c\}\rangle\rangle, D(s)_C\rangle;$$

again ε is equal to <E(c),T,∅> if E begins with a universal quantifier and equal to <E(c),T> otherwise.

This completes the list of conditions that any semantic tableaus must meet and therewith Def. DA3.[37]

It is useful to see at least for one example what a tableau construction according to the specifications of Def. DA3 looks like. Hence the following exercise:

Exercise EA1. Construct a tableau in accordance with the specifications of Def. DA3 for the argument

$$\langle(\exists y)(Q(y)\ \&\ (\forall x)(P(x) \rightarrow R(x,y)))\ |\ (\forall x)(P(x) \rightarrow (\exists y)(Q(y)\ \&\ R(x,y)))\rangle$$

Having given a precise formal reconstruction of semantic tableaux and their construction, we can now proceed to prove, on the basis of our formalisation, the properties of semantic tableaux which jointly establish soundness and completeness of the Tableau Method. As a preliminary we prove a lemma about infinite trees of the kind we are using.

---

[37] As described, the procedure for construc ting semantic tableaux is still not fully deterministic. Usually a tableay leads to splittings, and as soon as the tableau that is being constructed has more gthan one branch, there is the question in which bramch the next reduction is to be performed. This is a question that the tableaqu construction algorithm we have outlined doesn't address. (It is detreministic onlky with regard to the order of reductions within a ny given branch.) It is straightforward to turn the given algorithm into oe which also decides in a fully deterministic way which is to be the next branch in which a reduction step is to be carrie out. But to do so explicitly is yet another burden on notation, so have decided to let this matter rest. The reader can modify the given algorithm so that it is deterministic also in this respect if he or she feels the urge.

Lemma LA1. Every infinite tableau has at least one infinite branch.

Proof. Let T be a tree in the sense of Def. DA1. It is easy to see that the nodes of T can be distinguished into three categories: (i) nodes s such that there are only finitely many successors of s in T; (ii) nodes s whose successor $s^\frown 0$ has infinitely many successors in T; and (iii) nodes s such that $s^\frown 0$ has only finitely many successors in T but $s^\frown 1$ has infinitely many successors in T.

We make use of this tripartite division in defining the following function f on T: For s ε T, f(s) is specified as follows:

	$\varnothing$	in case (i) (s has finitely many successors in T)
f(s) =	$s^\frown 0$	in case (ii) ($s^\frown 0$ has infinitely many successors in T)
	$s^\frown 1$	in case (iii) ($s^\frown 1$ has infinitely many successors in T while $^\frown 0$ has finitely many successors in T)

Since T has infinitely many nodes, its root <> will have infinitely many successors. Moreover, if s is a node which has infinitely many successors, then f(s) will have infinitely many successors as well. This means that if we define the function g on the natural numbers 0, 1, 2, .. as in (1) below, then it will be the case that for each n g(n) is a node of T which has infinitely many successors in T:

(1) (a) g(0) = <>
(b) for all natural numbers n, g(n+1) = f(g(n))

It is evident that the range of g is an infinite sequence of nodes of t such that for each n g(n) is an initial segment of g(n+1). From this it follows immediately that if n and m are any natural numbers such that n < m, then g(n) is an initial segment of g(m). So the range of g is a linearly ordered subset of T, and, given that g is defined for all n, it is infinite. In fact, the set is a branch of T, since for each n the length of the sequence g(n) is n. So it is impossible to extend the set with an element s of T which does not yet belong to it without losing linearity. For s will of necessity be of some finite length n and so of the same length as the node g(n). It is clear, however, that for any two distinct sequences s_1 and s_2 of the same length neither is an initial segment of the other, i. e. we have neither s ≤ s' nor s' ≤ s. So no proper extension of Ran(g) with a further element of T will be a linear order. Hence Ran(g) is a maximal linear subset of T and thus a branch of T.

q.e.d.

N.B. The property which Lemma LA1 establishes for trees with at most binary branching - i.e. trees in which each node has at most two daughters - is a special case of a more general statement:

> Every infinite tree in which each node has finitely many daughters has an infinite branch.

Exercise: Show that any infinite tableau has an infinite branch.

For the remainder of this Appendix it will be convenient to introduce the following terminology. Suppose that <T,D> is a semantic tableau and that s is one of its stages (i.e. s ε T). We say that the formula A *occurs positively at* s iff A is the formula of a signed formula occurring in $D(s)_F$ whose sign is T. (That is, the signed formula is of the form <A,T,S> when A begins with a universal quantifier and in all other cases it equals <A,T>.)
Similarly, A *occurs negatively at* s iff A is part of a signed formula occurring in $D(s)_F$ whose sign is F.

Def. DA4

1. Suppose that <T,D> is a semantic tableau and Z a branch of T. Then we say that Z is *closed* iff there is a node s ε Z and an atomic sentence A which occurs both positively and negatively at s.

2. A semantic tableau <T,D> is *closed* iff every branch of it is closed.

Lemma LA2. Suppose that <T,D> is a semantic tableau and that Z is an infinite branch of T. Then Z is not closed.

Proof. This is immediate. Supppose that Z was closed. Then there would be an atomic formula A and a node s of Z such that <A,T> and <A,F> belong to $D(s)_F$. But in that case s would have no successors. (See (1) of Def. DA2.) So Z would be finite.

Corollary. If the semantic tableau <T,D> is closed, then T is finite.

Theorem TA1. (Soundness of the Tableau Method)

Suppose that the semantic tableau <T,D> for the argument $<A_1,..., A_n \mid B>$ is closed. Then $<A_1,..., A_n \mid B>$ is valid.

Proof. Assume $<T,D>$ is closed. From the Corollary it follows that T is finite. This entails that every branch of T consists of a finite set of nodes $<s_1, .., s_k>$.

We have to show that $A_1,..., A_n \models B$, i.e. that every model for the language L of $<A_1,..., A_n \mid B>$ which verifies the premises $A_1,..., A_n$ also verifies the conclusion B. Suppose that this is not so. Then there is a model M for L which verifies the premises but falsifies the conclusion.

We will construct a branch $<s_o,.., s_k>$ of nodes of T and a sequence $<M_o,.., M_k>$ of models where each pair (s_i,M_i) (i = 1,.., k) has the following three properties:

(P1) M_i is a model for the language $L_i = L \cup D(s_i)_C$.

(P2) If A occurs positively in $D(s_i)_F$, then $M_i \models A$.

(P3) If A occurs negatively in $D(s_i)_F$, then not $M_i \models A$.

N.B. the models M_i will all be *expansions* of the model M, i.e. they have the same universe U as M and the same interpretations for the non-logical constants of L. They differ from M only in providing denotations in U for the individual constants in the sets C_i. For the notion of 'model expansion' see Section 1.5 of this Chapter.

It should be clear that the combination of P1 - P3 leads to a contradiction. For it entails that P2 and P3 hold in particular for the final node s_k of the branch. But since s_k has no successors in T and its branch is closed, it must be the case that some sentence A occurs both positively and negatively at s. By P2 and P3 we then have that both $M_k \models A$ and not $M_k \models A$.

The construction of the pairs (s_i,M_i) proceeds by induction. For the basic step, which concerns the root node s_o, recall that $D(s_o) = D(<>) = < <\alpha_1,.., <\alpha_n, \beta>, <c_1> >$, where α_i is a signed formula with positive sign which contains premise A_i and β is a signed formula with negative sign which contains the conclusion B. In other words, the A_i occur positively at $<>$ and B negatively. Further, since $D(s_o)_C$ is the sequence $<c_1>$, we have that $L_o = L \cup \{c_1\}$, A model M_o for this language can be obtained from M by extending the interpretation function F_M of M to c_1. Since c_1 doesn't occur in either the A_i or B, it is immaterial how the interpetation of c_1 is chosen. That is, we can arbitrarily pick an element u of U_M and extend to the function $F_{M_o} = F_M \cup \{<c_1,u>\}$. If we then put: $M_o = < U_M, F_{M_o} >$, then clearly $M_o \models A_1,.., A_n$ and not $M_o \models B$.

Now suppose that s_i and M_i have been chosen, that (s_i,M_i) has the properties P1-P3 and that s_i has at least one successor in T. Then the one or two successors of s_i are the result of reducing one of the signed formulas in $D(s_i)_F$. The choice of s_{i+1} and M_{i+1} and the proof that they satisfy P1-P3 depends on what kind of reduction is involved.

We first consider those reductions which lead to one successor of s_i. And as regards these reductions, we begin by looking at the ones where the main operator of the reduced formula is a sentence connective. These are the cases where the signed formula to which the reduction applies has one of the following forms: $<\neg C,T>$, $<\neg C,F>$, $<C\&D,T>$, $<CvD,F>$ or $<C\rightarrow D,F>$. Once again we consider just one of these cases, and as before we focus on that of a conjunction occurring under TRUE, i.e. $<C\&D,T>$.

Suppose then that the transition from s to its immediate successor $s^\frown 0$ is the result of reducing the signed formula $<C\&D,T>$ belonging to $D(s_i)_F$. Since the reduction does not involve the introduction of a new parameter, we have in this case that $D(s_{i+1})_C$ is the same as $D(s_i)_C$. So $L_{i+1} = L_i$. This means that we can take M_{i+1} to be the same as M_i. So, since by assumption M_i satisfies P1, this will then also be the case for M_{i+1}. To verify P2 and P3 we need to show that the signed formulas in $D(s_{i+1})_F$ are true or false in M_{i+1} depending on whether their sign is T or F. For those signed formulas of $D(s_{i+1})_F$ that also belong to $D(s_i)_F$ this follows from the assumptions made about s_i and M_i. So the only signed formulas for which P2 and P3 have to be checked are those that have been added to $D(s_{i+1})_F$ in the transition from s_i to s_{i+1}. In the case at hand these are the positively signed formulas containing C and D. But since $<C\&D,T>$ belongs to $D(s_i)_F$ it follows by the induction assumption (more specifically, the assumption that P2 holds for s_i and M_i) that $M_i \models C \& D$. So by the clause for & in the Truth Definition, $M_i \models C$ and $M_i \models D$. Since $M_{i+1} = M_i$, the desired result follows.

Next, we consider cases where s_i leads to s_{i+1} through the reduction of a quantified formula. First suppose that the reduction involves a parameter that already belongs to $D(s_i)_C$. In this case the signed formula to which the reduction applies is either of the form $\langle(\forall x)E(x),T,S\rangle$ or of the form $\langle(\exists x)E(x),F,S\rangle$. We focus on the first possibility. Once more the immediate successor s_{i+1} is $s_i{}^\frown 0$. This entails that $D(s_{i+1})_C$ is identical to $D(s_i)_C$, so that once more $M_{i+1} = M_i$. Suppose further that the reduction of $(\forall x)E(x)$ consists in substituting for the free occurrences of x in E(x) the constant c^r from the list of parameters provided by $D(s_i)_C$. Thus the only signed formula in $D(s_{i+1})_F$ which does not belong to $D(s_i)_F$ is $\langle E[c^r/x],T\rangle$. So it is only necessary to verify P2 for this signed formula. By assumption $M_i \models (\forall x)E(x)$. This means that $[[(\forall x)E(x)]]^{M_i,\mathbf{a}} = 1$ for all assignments a in M_i, including in particular those assignments a such that $a(x) = F_i(c^r)$ (which in turn is equal to $[[c^r]]^{M_i,\mathbf{a}}$). So by the clause of the Truth Definition for the universal quantifier it follows that $[[E(x)]]^{M_i,\mathbf{a}} = 1$, where $\mathbf{a}$ is any assignment in M such that $a(x) = [[c^r]]^{M_i}$. And from this it follows by the Corollary to Lemma 2[38] that $[[E[c^r/x]]]^{M_i,\mathbf{a}} = 1$. Since $E[c^r/x]$ is a sentence, this amounts to the same thing as: $M_i \models E[c^r/x]$. This concludes the case under consideration.

Now consider those reductions of quantified formulas which involve the introduction of a new parameter into the tableau branch. These are the cases where the signed formula that is reduced is either of the form $\langle(\exists x)E(x),T\rangle$ or of the form $\langle(\forall x)E(x),F\rangle$. We focus on the first of these.

Again we put $s_{i+1} = s_i{}^\frown 0$. Suppose that the new parameter is c_k. Then $D(s_{i+1})_C$ consists of $D(s_i)_C$ together with c_k. This means that $L_{i+1} = L_i \cup \{c_k\}$, so this time M_{i+1} will have to be a proper expansion of M_i. Since $\langle(\exists x)E(x),T\rangle$ occurs in the first member of $D(s_i)$, by induction assumption $M_i \models (\exists x)E(x)$. So there is an element d in the universe U of M_i such that $[[E(x)]]^{M_i,\mathbf{a}} = 1$ where $\mathbf{a}$ is any assignment such that $\mathbf{a}(x) = d$. This means that we can make sure that (s_{i+1},M_{i+1}) satisfy P1-P3 by defining M_{i+1} to be that model for L_{i+1} which is like M_i as far as L_i is concerned and in addition interprets c_k as denoting d. (That is, $F_{i+1}(\alpha) = F_i(\alpha)$ for every non-logical constant α of L, and $F_{i+1}(c_k) = d$.) For then we have that $[[E(x)]]^{M_{i+1},\mathbf{a}} = 1$, provided $\mathbf{a}(x) = [[(c_k]]^{M_{i+1}}$. So, again by the Corollary to Lemma 2, $M_{i+1} \models E[c_k/x]$, which concludes the argument for this case.

---

38 See Section 1.1 of this Chapter.

This concludes the argument for all reductions which lead to a single successor. Now we consider those reductions which produce two successors. These are all reductions of formulas whose main operator is a sentence connective. To be precise, the types of signed formulas which lead to pairs of successor nodes are, as may be recalled from Def.DA2, <CvD,T>, <C&D,F>, <C$\rightarrow$D,T>, <C$\leftrightarrow$D,T> and <C$\leftrightarrow$D,F>. Once more we focus on the first of these.

Since we are dealing with a reduction in which no new parameter is introduced, we have, as in earlier cases of this kind, that $M_{i+1} = M_i$. But this time the choice that matters is that of the successor s_{i+1} to s_i. We know from the induction assumption that $M_i \models C \vee D$. This entails (by the Truth Definition clause for v) that either $M_i \models C$ or $M_i \models D$. Suppose that the first of these is true. Then we choose s_{i+1} to be $s_i{}^\frown 0$. The first member of $D(s_{i+1}) = D(s_i{}^\frown 0)$ differs from the first member of $D(s_i)$ only in having the additional signed formula <C,T>[39]. But by assumption $M_i \models C$. So, since $M_{i+1} = M_i$, $M_{i+1} \models C$. If it is not the case that $M_i \models C$, then $M_i \models D$. In this case we choose s_{i+1} to be $s_i{}^\frown 1$. Otherwise the reasoning is just as in the first case.

It should be stressed that since the entire tree T is finite (see the Corollary to Lemma LA2), there is a finite upper bound n to the possible length of the branch we are constructing. So after at the very most n steps the end node of this branch will be reached and with it the contradiction we have been aiming for.

This concludes the argument for our last case, and with it the proof of Theorem TA1. q.e.d.

<u>Theorem TA2.</u> (Completeness of the Tableau Method)

Suppose that the semantic tableau for the argument <A_1,..., A_n | B> is not closed. Then <A_1,..., A_n | B> is not valid.

<u>Proof.</u> Suppose that the premises and conclusion of <A_1,..., A_n | B> belong to the language L and that the tableau <T,D> for <A_1,..., A_n | B> is not closed. Then <T,D> has an open branch Z. Let C(Z) be the set of all individual constants c such that there is a node s in Z with c

---

[39] Or <C,T,∅> in case C begins with a universal quantifier. This qualification will be needed also in a number of further cases below. Since it should by now be clear when such cases arise, we will henceforth forgo drawing explicit attention to this qualification.

occurring in $D(s)_C$ and let L' be the language $L \cup C(Z)$. Furthermore, let PF(Z) be the set of those sentences of L' which occur positively at some stage of Z and let NF(Z) be the set of those sentencs which occur negatively at some stage of Z. We prove Th. TA2 by constructing a model M for L' in which the members of PF(Z) are all true and the members of NF(Z) are all false. This will entail that in particular the signed formulas that occur in D(<>) are true or false in M according to whether their sign is T or F. So the premises $A_1,\ldots, A_n$ are true in M and the conclusion B is false in M, which proves that $\langle A_1,\ldots, A_n \mid B\rangle$ is invalid.

M is defined as follows

(i) The universe U_M of M is the set C(Z).

(ii) Let P be an n-place predicate of L. Then the interpretation $F_M(P)$ of P in M is defined to be the following function from the Cartesian product U^n (= $U \otimes ..(\text{n times}).. \otimes U$) into the set $\{0,1\}$:[40]
$F_M(P)(c^1, ..,c^n) = 1$ iff $P(c^1, ..,c^n) \in PF(Z)$

(iii) c is a constant from C(Z). Then $F_M(c) = c$. (That is, we let c denote itself.)

To prove that M verifies the sentences in PF(Z) and falsifies the sentences in NF(Z) we proceed by induction on the syntactic complexity of formulas.

To show the base case, suppose first that the atomic sentence $P(c^1,..,c^n)$ belongs to PF(Z). Then, by the definition of F_M, $F_M(P)(c^1, .., c^n) = 1$. So by the Truth Definition, $M \vDash P(c^1, .., c^n)$. Now suppose that $P(c^1, ..,c^n) \in NF(Z)$. Then it is not the case that $P(c^1, .., c^n) \in PF(Z)$; for if this were the case, then there would be a node s of Z such that $\langle P(c^1, .., c^n),T\rangle$ and $\langle P(c^1, .., c^n),F\rangle$ both occur in $D(s)_F$, and then s would have been the final node of Z and Z would have been closed. So, by the definition of F_M, $F_M(P)(c^1, .., c^n) = 0$, and so it follows from the Truth Definition that it is not the case that $M \vDash P(c^1, .., c^n)$.

Second, assume that A is a complex sentence, that the induction assumption holds for all sentences of smaller complexity and that the main operator of A is a sentence connective. We only consider one

---

40 As usual, $U^1 = U$.

case, that where A is of the form C & D. We assume that the induction hypothesis holds for C and for D.

First suppose that A ε PF(Z). Then there must be some node s in Z such that A occurs positively at s and the signed formula containing A that belongs to $D(s)_F$ has been reduced in the transition from s to its (unique) successor $s^\frown 0$. (That there must be such an s follows from the fact that there is by definition of PF(Z) some s' in Z such that <A,T> belongs to $D(s')_F$. Since $D(s')_F$ is a finite sequence, and since with each reduction of an element of the sequence the signed formula containing A moves closer to a position in the sequence where it will be the formula up for reduction, its reduction is bound to take place either at s' itself or at some successor of s'. Note also in this connection that universal formulas under TRUE and existential formulas will be reduced at least once.) This means that $D(s^\frown 0)_F$ contains both <C,T> and <D,T>. From the induction assumption it then follows that $M \models C$ and $M \models D$. So by the clause for & of the Truth Definition, $M \models C \,\&\, D$.

Now suppose that A ε NF(Z). Then for some node s' in Z A occurs negatively at s'. As above, we infer that there must be a node s in Z such that a negatively signed formula containing A is reduced at s. In this case the reduction has led to two successors $s^\frown 0$ and $s^\frown 1$ of s, with <C,F> occurring in $D(s^\frown 0)_F$ and <D,F> occurring in $D(s^\frown 1)_F$. One of these successor nodes must belong to Z, for otherwise Z would not be a maximal linearly ordered subset of T and thus wouldn't be a branch. Let us assume that $s^\frown 0$ belongs to Z. Then we may conclude from the induction assumption that it is not the case that $M \models C$. But then it also won't be the case that $M \models C \,\&\, D$.

The remaining cases are sentences beginning with a quantifier. We will only consider the case of the existential quantifier. Suppose that A has the form $(\exists x)E(x)$. Once again we begin with the case where A belongs to PF(Z). This means that for some s' in Z <$(\exists x)E(x)$,T> occurs in $D(s')_F$. As before we may conclude that there is a node s in Z such that $s \geq s'$ and <$(\exists x)E(x)$,T> is reduced in the transition from s to $s^\frown 0$. In this case a new parameter c_k is introduced into Z and the signed formula <$E[c_k/x]$,T> belongs to $D(s^\frown 0)_F$. From the induction assumption it follows that $M \models E[c_k/x]$ and from this by the Truth Definition that $M \models (\exists x)E(x)$, i.e. $M \models A$.

The final case to be dealt with is that where $(\exists x)E(x) \in NF(Z)$. In this case there is a node s' in Z such that a signed formula $<(\exists x)E,F,S>$ is a member of $D(s')_F$. As we have seen, reductions of signed formulas of this kind do not result in elimination of the signed formula to which the reduction applies; instead the formula is put at the end of $D(s^\frown 0)_F$ each time that the formula is subjected to reduction in the transition from some node s in B to $s^\frown 0$. In fact, given the way in which we have defined the procedure for treating signed formulas of this type and putting them back in the queue, it is easy to verify that for each c in the parameter set C(Z) there will be a transition from some node s in Z to its successor $s^\frown 0$, in which c has been used to instantiate the quantifier $(\exists x)$ in $<(\exists x)E(x), F>$, with the effect that $<E[c/x],F>$ has been added to $D(s^\frown 0)_F$;. Thus $E[c/x] \in NF(Z)$. Therefore we can infer, using the induction assumption, that for each $c \in C(Z)$ it is not the case that $M \models E[c/x]$. Since $C(z) = U_M$, and for each $c \in C(Z)$, $F_M(c) = c$ it follows from the Corollary to Lemma 2 and the clause for $\exists$ of the Truth Definition that it is not the case that $M \models (\exists x)E(x)$. In other words, it is not the case that $M \models A$.

This concludes the proof of our last case, and therewith of Theorem TA2.

q.e.d.

<u>Arguments with identity.</u>

So far we have proved soundness and completeness under the assumption that = does not occur in $<A_1,..., A_n \mid B>$. We now drop this assumption. This means that the tableau for $<A_1,..., A_n \mid B>$ will in genral contain atomic formulas of the form '$c_i = c_j$'. When the sign of such a formula is T, then it can give rise to 'reduction' steps involving applications of the rule (=,Sub). And for such applications there is the same requirement as for other rules: all possible applications must be carried out at some stage. It might be thought that for applications of (=.Sub) this requirement presents a similar bookkeeping problem as for universally quantified formulas under TRUE and existentially quantified firmulas under FALSE, since in both cases the same formula will typically have to be subjected to repeated applications. (For instance, the formula $P(c_1,c_2)$ will have to be subjected to the rule in combination with each formula under TRUE that is either of the form '$c_1 = c_i$' or of the form '$c_2 = c_i$'.) In the case of the two types of quantified formulas that give rise to this problem we were forced to introduce a special device that keeps track of which instantations have already been carried out. Fortunately, however, in connection with (=,Sub) no new notational device is needed. The reason is that we have restricted the applications of (=,Sub) to atomic formulas. The result of applying (=,Sub) to an equation $c_i = c_j$ and an atomic formula $P(c_{i_1},..,c_{i_n})$ is again an atomic formula and no atomic formula is ever deleted from a tableau branch once it has become part of it. This means that whenever a given application of (=,Sub) is being considered, we can check whether the formula that would result from it already belongs to the given tableau branch. If that is so, then we do not carry out the application and pass to the rule application that is next in line.

It is still necessary to agree on a convention which ensures that all substitution results that can be obtained by applications of (=,Sub) are obtained, without the risk that other rule applications might remain in the queue forever. One convention that will do this is as follows: (i) apply (=,Sub) only when its signed identity premise $<c_i = c_j,T>$ occurs as first formula of $D(s)_F$. Then look at the first signed formula $<A,T/F>$ in $D(s)_F$ such that A has an occurrence of c_i. Consider the leftmost occurrence of c_i in A. If the result of applying (=,Sub) to $<c_i = c_j,T>$ and A,T/F> already occurs in $D(s)_F$, then pass to the next occurrence of c_i in A. If all results of substituting c_j for some occurrence of c_i in A already belong to $D(s)_F$, then pass to the next signed formula in which there is an occurrence of c_i; and so on. When all possible applications of (=,Sub) with $c_i = c_j$ as identity premise have been carried out - at any stage there can of course be only finitely many such applications - then $<c_i = c_j,T>$ is moved from the beginning to the end of $D(s)_F$.

There is one further matter connected with the rule (=,Sub) that must be raised at this point. Intuitively, applications of the rule with identity premise $<c_i = c_j,T>$ and second premise $<A,T/F>$ should not only allow replacements of c_i by c_j but also replacements of c_j by c_i. This is not the way in which we have formulated the rule, however, The reason why the formulation we have given, according to which an identity premise $<c_i = c_j,T>$ only allows for replacements of c_i by c_j, suffices is that tableau construction also allows for applications of the rule (=,Ref). These allow us to introduce signed formulas of the form $<c_i = c_i,T>$ whenever we need them. Such formulas can then serve as non-identity premises in applications of (=,Sub) to lead from $<c_i = c_j,T>$ to $<c_j = c_i,T>$.

In order to make sure that we get all the instances of $<c_i = c_i,T>$ that might ever be needed in our tableau branches we make the following provision. Each time a constant c_i gets introduced into a tableau branch at a stage s, we add $<c_i = c_i,T>$ to the end of $D(s^\frown 0)_F$ (the formula part of the decoration of hte unique successor of s).

We are now ready to modify the proofs of the Soundness and Completeness Theorems so that they also apply to arguments that contain =. For Soundness this is straightforward. Once again we assume that the tableau T for the argument $<A_1,.., A_n \mid B>$ is closed and suppose that there is a model M for L such that $M \models A_1,.., A_n$ while not $M \models B$. Agaon we prove by induction on n that there is a linearly ordered subset $<s_o,..,s_n>$ of T, with $s_o = <>$, and a sequence of models $<M_o,..,M_n>$, where M_i is a model for the language $L \cup D(s_i)_C$ such that the positive formulas of s_i are true in M_i and the negative formulas are not. As before, this then gives a contradiction with the assumption that T is closed, which entails that there is a uniform finite upper bound to the lengths of its branches. The proof that such a pair of sequences $<s_o,..,s_n>$ and $<M_o,..,M_n>$ can be built carries over without modification except that we must now also deal with the new rule applications, viz. those of (=,Sub) and (=,Ref).

The applications of (=,Ref) are adjoined to the applications of those rules that introduce new constants. None of these applications need worry us here, since formulas of the form $c_i = c_i$ are true in all models.

That leaves applications of (=,Sub). Suppose that we have constructed the pair of sequences $<s_o,..,s_n>$ and $<M_o,..,M_n>$ and that the rule application in s_n is an application of (=,Sub) with identity premise $<c_i = c_j,T>$ and second premise <A,T/F>. By assumption A is an atomic formula, so it is either of the form $P(c_{i_1},..,c_{i_n})$ or else an identity. The argument is the same for these two cases; let us assume, without loss of generality, that A has the form $P(c_{i_1},..,c_{i_n})$. We also assume, again without loss of generality, that the sign of <A,T/F> is T.

Since applications of (=,Sub) produce no splitting, s_n will have a single successor $s_n{}^\frown 0$ in T. This fixes the next node s_{n+1} of the sequence as $s_n{}^\frown 0$. Also, since no new constants are introduced by applications of (=,Sub), we can take the model M_{n+1} to be the same as M_n. By induction assumption (i) $M_{n+1} \models c_i = c_j$ and (ii) $M_{n+1} \models P(c_{i_1},..,c_{i_n})$. Let c_{i_k} be the occurrence of c_i in $P(c_{i_1},..,c_{i_n})$ which gets replaced in the given application of (=,Sub) by c_j; the result is the T-signed formula $P(c_{i_1},..,c_{i_{k-1}},c_j,c_{i_{k+1}},..,c_{i_n})$. It follows directly from the clause for atomic formulas in the Truth Definition together with (i) and (ii) above that $M_{n+1} \models P(c_{i_1},.., c_{i_{k-1}}, c_j, c_{i_{k+1}},.., c_{i_n})$.

So much for the modification of the proof of TA1. To adapt the proof of TA2 a little more is needed. To see this suppose for instance that the sentence c = c' is a sentence from PF(Z), where c and c' are distinct constants from C(Z). Then it should be the case that $M \models c = c'$. But according to the Truth Definition this will be so only if $[[c]]^{M,\mathbf{a}} = [[c']]^{M,\mathbf{a}}$ (where **a** may be any assignment whatever). But that won't be the case if $F_M(c) = c$ and $F_M(c') = c'$, since by assumption $c \neq c'$.

We adopt the standard solution to this difficulty, which consists in taking U_M not to consist of the constants in C(Z) themselves, but of equivalence classes of these constants, which we obtain by "dividing" the set C(Z) by a certain equivalence relation. This relation is generated by the set of all sentences of the form c = c' that belong to PF(Z). To be precise, we define the following relation ≡ between constants in C(Z):

$$c \equiv c' \text{ iff } c = c' \, \varepsilon \, PF(Z) \qquad (\equiv)$$

But is this relation ≡ really an equivalence relation? It is, but a few remarks are in order to show why that is so. First, Reflexivity of ≡ holds because our tableau construction makes sure that c = c gets added to PF(Z) for every constant c that gets introduced into Z. Secondly, that ≡ is symmetric follows from our observation above: Suppose that c ≡ c'. Then c = c' belongs to PF(Z). We know already that c' = c' also belongs to PF(Z). But that means that c' = c also to PF(Z). For if this formula doesn't enter Z in some other way, then some application of (=,Sub) in Z, in which the identity premise c = c' is used to replace the second occurrence of c' in c' = c' by c, will have added it. Thirdly, ≡ is transitive, for much the same reason that it is symmetric. Suppose that c ≡ c' and c' ≡ c''. Then c = c' and c' = c'' both belong to PF(Z). But then c = c'' will also belong to PF(Z), either through an application of (=.Subj) in which c' = c'' is used as identity premise and c = c' as A, or in some other way.

Along these same lines we can also show that PF(Z) has the following property

Let P be any m-place predicate of L. (Con $\equiv$)
If $P(c^1,..., c^m)$ and $c^1 = c'^1, ..., c^m = c'^m$ belong to PF(Z), then $P(c'^1,..., c'^m)$ also belongs to PF(Z).

Remark 1 "Con $\equiv$" stands for 'Congruence of $\equiv$'. A binary relation R is called a *congruence relation with respect to* some m-place relation S (where m can be any natural number) iff for any two m-tuples $<a_1, ..., a_m>$ and $<b_1, ..., b_m>$, if $<a_1, ..., a_m> \varepsilon$ S and $<a_i,b_i> \varepsilon$ R for $i = 1, ..., m$, then $<b_1, ..., b_m> \varepsilon$ S. So (Con $\equiv$) states that $\equiv$ is a congruence relation with respect to the m-place relation S which holds between entities $a_1, ..., a_m$ (here the entities are the constants in C(Z)) iff the sentence $P(a_1,..,a_m)$ belongs to PF(Z).

Remark 2 Note that (Con $\equiv$) includes cases where for one or more $i \leq m$ c'^i is the same constant as c^i. In these cases "replacement of one or more occurrences of c^i by c'^i" amounts to leaving those occurrences just as they were. Since any self-identity formula $<c = c,T>$ will belong to $D(s)_C$ from the stageat which c has made its entry into the given tableau branch, (Con $\equiv$) also covers cases where only some of the constants in $P(c^1,..., c^m)$ are replaced by other constants. And of course, many applications of (=,Sub) will be of this kind. For as we have formulated (=,Sub), it is always applied to only one constant occurrence at a time. So whenever the head of the atomic formula that plays the part of A in the application is a predicate of 2 or more places, then the application will leave some constant occurrences unchanged.

The properties which have been shown to hold for $\equiv$ entail that an open tableau branch Z can be converted into the following counter-model M. (We denote the equivalence class generated within the set C(Z) by a constant c ε C(Z) as "$[c]_\equiv$".)

(i) $U_M = \{[c]_\equiv : c \ \varepsilon \ C(Z)\}$

(ii) $F_M(P)([c^1]_\equiv,.., [c^m]_\equiv) = 1$ iff there are $c'^1 \ \varepsilon \ [c^1]_\equiv,...,$ $c'^m \ \varepsilon \ [c^m]_\equiv$ such that $PP(c'^1,...,c'^m) \ \varepsilon$ PF(Z)

(iii) $F_M(c) = [c]_\equiv$

The proof that all sentences in PF(Z) are true in M and all sentences in NF(Z) false in M involves the same steps as the proof of Theorem TA2 given earlier. Most of the steps carry over without change. The steps that deserve a closer look are those for atomic sentences and those for quantified formulas.

(1) Atomic sentences.

First suppose that $P(c^1,\ldots,c^m) \varepsilon PF(Z)$. Then by (ii) above $F_M(P)([c^1]_\equiv,.., [c^m]_\equiv) = 1$. So, in virtue of (iii), $M \models P(c^1,\ldots,c^m)$.

Now suppose $P(c^1,\ldots,c^m) \varepsilon NF(Z)$. To show that it is not the case that $M \models P(c^1,\ldots,c^m)$ we need to show (***):

$$\text{For no } c'^1 \varepsilon [c^1]_\equiv,\ldots, c'^m \varepsilon [c^m]_\equiv , P(c'^1,\ldots,c'^m) \varepsilon PF(Z) \qquad (***)$$

Suppose there were $c'^1 \varepsilon [c^1]_\equiv,\ldots, c'^m \varepsilon [c^m]_\equiv$ such that $P(c'^1,\ldots,c'^m) \varepsilon PF(Z)$. Then by (Con $\equiv$) also $P(c^1,\ldots,c^m) \varepsilon PF(Z)$. But then Z would be closed, contrary to assumption. So (***) holds; so by (ii) of the definition of M $F_M(P)([c^1]_\equiv,.., [c^m]_\equiv) = 0$; so it is not the case that $M \models P(c^1,\ldots,c^m)$.

It is to be noted that we now also have to deal with a type of atomic sentence which did not play a role in our earlier proof of Lemma 6 under the restrictions there assumed, viz. sentences of the form $c = c'$. However, this case is just like the case of atomic formulas of the form $P(c^1,\ldots,c^m)$, of which we have just shown that they behave in the required way. It is left to the reader to verify this.

(2) Quantified sentences.

Again we only consider the case of an existential sentence $(\exists x)E(x)$. First suppose that $(\exists x)E(x) \varepsilon PF(Z)$. Then there is a node s in Z such that, for some $c \varepsilon C(Z)$, <E[c/x],T> belongs to $D(s)_F$. So by the induction assumption $M \models E[c/x]$. By Corollary 1 to Lemma 2 this entails that $[[E(x)]]^{M,\mathbf{a}} = 1$ for any assignment $\mathbf{a}$ such that $\mathbf{a}(x) = [[c]]^{M,\mathbf{a}} = F_M(c) = [c]_\equiv$. So it follows from the Truth Definition that $M \models (\exists x)E(x)$.

Second, assume that $(\exists x)E(x) \varepsilon NF(Z)$. Then for no $c \varepsilon C(Z)$ $E[c/x] \varepsilon PF(Z)$. For suppose that $E[c/x] \varepsilon PF(Z)$. Then <E[c/x],T> would belong to $D(s)_F$ for some node s in Z. But then c would have had to be a member of $D(s)_C$. Since $(\exists x)E(x)$ ε NF(Z), it may be assumed without loss of generality that $D(s)_F$ also contains $<(\exists x)E(x),F>$. So either at s or at some later stage of Z $<E[c/x],F>$ would have become part of the decoration as well. But then Z would have been closed, contrary to assumption. So it follows that E[c/x] ε PF(Z) for no c ε C(Z). Using the induction assumption we can infer that for no c ε C(Z), $M \models E[c/x]$. Relying once more on Corollary 1 of Lemma 2, we conclude that for no element $[c]_\equiv$ of U_M, $[[E(x)]]^{M,a[c]_\equiv/x} = 1$. So it follows from the Truth Definition that it is not the case that $M \models (\exists x)E(x)$.

This completes the modifications that are needed in the proof of Theorem TA2.

<u>Remark on the rule (=,Sub).</u>

The version of (=,Sub) we have assumed involves the restriction that replacement of constants is allowed only in atomic formulas. There is also a stronger version of the rule, according to which replacements of constants are permitted in arbitrary formulas. That the more general version of the rule is over all no more powerful than the restricted version is something that may not be immediately obvious. But one corollary of our completeness proof is that this must be so: Since applications of the general version are valid, they must be provable by means of the tableau method in which only the restricted version of the rule is used. Ao any proof in which there are applications of the generalised version of (=,Sub) can be replaced by a proof of the same argument in which there are only applications of the restricted version.

(=,Sub) also allows for another generalisations, according to which several occurrences of the same constant c in A can be replaced at once. Our proof showsthat this generalisation doesn't add real deductive power either. However, in this case it is obvious in any case that the generalisation doesn't buy us more than the version which permits only one replacement at a time. For, evidently, any case of simultaneous replacement can be mimicked by a succession of applications of (=,Sub), in which each application involves replacement of just one of these occurrences.

As noted at the outset of this Appendix (see also Section 1.1.3 of this Chapter), the Soundness and Completeness proofs we have given are still not quite as general as they might have been, since we have assumed that the language L contains no function constants. Extending the formal treatment of the tableau method, and exact proofs of soundness and completeness based on it, to this more general case is a routine exercise. But the exercise is awkward and cumbersome, and doesn't bring anything to light that is of real interest. On the other hand, as we noted earlier on, we can generalise the results we have obtained to languages with function constants by translating arguments in which function constants occur into arguments in which those constants have been replaced by corresponding predicate constants. The reader can find out how this works by going through Exercise EA2 below.

One final observation on the tableau method in the context of this Chapter. In Section 1.4 we made use of the fact that for arbitrary sets of sentences Γ (i.e. infinite as well as finite sets) satisfiability coincides with consistency. This result is established in the proof of the Completeness Theorem given in Section 1.2, but strictly speaking it has not been established by the tableau-related proof we have given in this Appendix. The problem is that we have developed our algorithmic version of the tableau method only for arguments with finite sets of premises. We still need to establish that the method can be extended so that it also covers infitnite premise sets.

As a matter of fact, with the mathematical tools available to us at this point this result can be proved only for sets that are at most denumerably infinite. Given how we have defined first order predicate logic this doesn't constitute a real limitation, as our definition admits only denumerable sets of sentences anyway. But since our formalism does allow for denumerable sets and since these will play an important role throughout, the tableau method should be modified so that at least denumerable premise sets can be handled.

As a matter of fact extending the construction algorithm to this effect isn't difficult. Suppose that we want to construct a tableau for the argument $<\Gamma| B>$, where Γ is denumerably infinite and C_1, C_2, .. is a complete enumeration of Γ. Then we can modify the tableau construction as it was defined hitherto as follows: We reserve certain construction stages s for the introduction of a new premise from our list C_1, C_2, .. . (For instance we could reserve for this purpose those stages whose length is a prime number.) Each time when such a node s is reached, (e.g, when length(s) = p_n, where p_n is the n-th prime number), we add the pair $<C_n,T>$ (or $<C_n,T,\varnothing>$, depending on the form of C_n) to the end of $D_F(s)$. Since this is an operation that does not produce a tableau split, so s has only one successor $s^\frown 0$. No other modifications are needed. So apart from the points where new premises are brought into play, everything proceeds as before. Moreover, if at a given stage of a tableau branch construction no reduction rules can be applied, then the next premise is "loaded" at that point.

It should be clear that an open branch B of a completed tableau constructed according to the new specification will have occurrences under the column 'TRUE' of all the premises in Γ. (i. e. $\Gamma \subseteq PF(B)$. So the model we construct from B will verify all sentences in Γ. It is also easy to see that notwithstanding the extra construction steps that are now required for the introduction of the premises in Γ, the length of B will be at most denumerably infinite.

Exercise EA2.

a. Let L be a language of First Order Predicate Logic with finitely many function constants f_1, ..., f_k and let $\langle A_1,..., A_n \mid B\rangle$ be an argument of L. Let for each i = 1, ..., k n_i be the number of argument places of f_i.

We form a new language L' which contains all the predicates of L which occur in $\langle A_1,..., A_n \mid B\rangle$ and which furthermore has for each i = 1, ..., k a distinct predicate Q_{f_i} of $n_i + 1$ places which does not occur in $\langle A_1,..., A_n \mid B\rangle$. We translate $\langle A_1,.., A_n \mid B\rangle$ into an argument $\langle A'_1,..., A'_{n+k} \mid B'\rangle$ of L' as follows:

(i) With any term t of L we associate formulas $P_t(x)$ of L' with distinguished free variable x. $P_t(x)$ is defined by induction on the complexity of t.

(a) If t is the variable v_i, P(t) is the formula x = t, where x is a variable not occurring in t.

(b) Supppose that $t = f(t_1,...,t_m)$, and that $P_{t_1}(x)$, .., $P_{t_m}(x)$ have been defined. Choose distinct variables $x_1,...,x_m$ not occurring in t and let $P_t(x)$ be the formula

$$(\exists x_1)..(\exists x_m)(P_{t_1}(x_1) \,\&\, .. \,\&\, P_{t_m}(x_m) \,\&\, Q_f(x_1,..,x_m,x)).$$

(ii) Each of the sentences $A_1,..., A_n$, B is translated as follows. (In the description of the translation we focus on A_1 but the same procedure applies to all other sentences of the argument)

(a) Let α be an occurrence of the atomic formula $P(t_1,..., t_n)$ in A_1. Then we replace this occurrence by the formula

$$(\exists x_1)..(\exists x_n)(P_{t_1}(x_1) \ \&\ .., \ \&\ P_{t_n}(x_n) \ \&\ P(x_1,..,x_n)),$$

where $x_1,..,x_n$ are variables not occurring in A_1.

(b) Let α be an occurrence of the atomic formula $t_1 = t_2$ in A_1. Then we replace this occurrence by the formula

$$(\exists x_1)(\exists x_2)(P_{t_1}(x_1) \ \&\ P_{t_2}(x_2) \ \&\ x_1 = x_2).$$

where x_1, x_2 are variables not occurring in A_1.

(iii) The translation of the argument $<A_1,..., A_n \mid B>$ is the argument $<A'_1,..., A'_{n+k} \mid B'>$, where

(a) for I = 1, ..., n A'_i is the translation of A_i as described under (ii);

(b) B' is the translation of B as described under (ii); and

(c) for j = n+1, ..., n+k A'_j is the sentence

$$(\forall x_1)..(\forall x_{n_j})(\exists y)(Q_{f_j}(x_1,..,x_m,y) \ \&$$
$$(\forall y)(\forall y')(Q_{f_j}(x_1,..,x_m,y) \ \&\ Q_{f_j}(x_1,..,x_m,y') \rightarrow y = y'))$$

(This sentence says that Q_{f_j} behaves like an m-place function with the function value represented by its last argument.)

Show: $A_1,..., A_n \models B$ iff $A'_1,..., A'_{n+k} \models B'$ (1)

b. Show (1) for the case where L has infinitely many function constants.

Exercise EA3.

Suppose that $<A_1,..., A_n \mid B>$ is an argument in which = does not occur. We can then still apply the tableau construction as described for arguments which do contain occurrences of =. Show that when a tableau for $<A_1,..., A_n \mid B>$ that is constructed according to this method has an open branch and ≡ is the relation between constants determined by this branch, then ≡ is the identity relation on the set C of constants occurring in this branch (i.e. $\equiv\ = \{<c,c>: c \in C\}$.

Exercise EA4.

We can prove the correctness and completeness results for argumentts with constants also by modifying the tableau construction algorithm directly. This is not difficult in principle, but it requires careful bookkeeping. For as soon as we have to deal with function constants of 1 or more argument places, the number of terms that have to be substituted for univeral quantifiers under True and existential quantifiers under False explodes. (Even with one 1-place function constant f and one indvidiual constant c we get an infinite number of such terms: f(c), f(f(c)), f(f(f(c))) and so on. Since we cannot allow for any of the possible substitutions to be "missed" by the algorithm, some kind of "pecking order" among the terms has to be defined, so that via the right kind of rotation system each pair consisting of (i) a term that can bebuilt form the function constants and the individual constants that have been introduced and (ii) a formula that can be instantiated by the term gets its turn.

Think of a modification of the construction algorithm which guarantees that every possible substitution of every closed term for the quantifiers of such formulas is executed at some point in the course of the construction of every infinite (open) branch of a non-closing tableau.

Solution to Ex. EA4.

The result that needs showing is that

$$A_1,..., A_n \models B \text{ iff } A'_1,.., A'_n, A'_{n+1}.., A'_{n+k} \models B' \qquad (*)$$

where the first argument belongs to a language L with function constants f_1,.., f_k, the second argument belongs to the language L' which has instead of each n-place function constant f_i of L a new p(n+1)-place predicate Q_{f_i}, A'_1,.., A'_n are the translations of A_1,..., A_n and A'_{n+1}.., A'_{n+k} are the axioms that state that the new predicates are functional in their last arguments. (*) followss from the following statement (1)

(1) Let C be any formula of L, M = $\langle U,F\rangle$ a model for L and let M' = $\langle U,F'\rangle$ be the model for L' which is obtained by putting:
(i) $F'(\alpha) = F(\alpha)$ for all $\alpha \varepsilon L \cap L'$,
(ii) $F'(Qf_i,)(\langle u_1,.. u_n, u_{n+1}\rangle) = 1$ iff $F'(f_i,)(\langle u_1,..., u_n\rangle) = u_{n+1}$.
Then for any assignment **a** in M, $[[C]]^{M,a} = [[C]]^{M',a}$

We first show that (1) entails (*). First suppose that $A'_1,.., A'_n, A'_{n+1}.., A'_{n+k} \models B'$. Let M be a model for L and **a** an assignment in M such that $[[A_i]]^{M,a} = 1$ for i = 1,..,n. Let M' be the model for L' that is obtained from M in the way described under (1). Then the following two statements hold:

(i) $[[A'_i]]^{M',a} = 1$ for i = 1,..,n, because of (1)
(ii) $[[A'_i]]^{M',a} = 1$ for i = n+1,..,n+k, because of the way M' is constructed from M.

Since by assumption $A'_1,.., A'_n, A'_{n+1}.., A'_{n+k} \models B'$, it follows that $[[B']]^{M',a} = 1$. So by (1) $[[B]]^{M,a} = 1$. Since this holds for arbitrary M and **a** we conclude that $A_1,..., A_n \models B$.

Now suppose that $A_1,..., A_n \models B$. Let M' be a model for L' such that $[[A'_i]]^{M',a} = 1$ for i = 1,..,n+k. Note that since $[[A'_{n+j}]]^{M',a} = 1$ for j = 1,..,k, there is for each j = 1,..,k and each m_j-tuple $\langle u_1,..., u_{m_j}\rangle$ (where m_j is the arity of the function constant f_j) a unique object w_j in U such that $[[Q(x_1,..., x_{m_j},y\rangle]]^{M',a[w_j/y]} = 1$. This means that we can define the model M for L from M' by keeping its universe U and the interpretations $F'(\alpha)$ for all $\alpha \varepsilon L \cap L'$ while defining the interpretations $F(f_j)$ of the function constants f_j of L by the clause:

for every m_j-tuple $\langle u_1,.,u_{m_j}\rangle$ of objects ε U, $F(f_j)(\langle u_1,..., m_j\rangle) = w_j$, where w_j is the object that is uniquely determined by $\langle u_1,..., m_j\rangle$ in the way indicated above.

It is easily seen that because of the way in which we have defined the interpretations of the function constants of L in M, M and M' are related as in (1). So by (1) we get that $[[A_i]]^{M,a} = 1$ for i = 1,..,n. Since by assumption $A_1,..., A_n \models B$, it follows that $[[B]]^{M,a} = 1$. So by (1)

$[[B']]^{M',a} = 1$. Again we can conclude because of the generality of the reasoning that this holds for arbitrary models M' for L', so that $A'_1,..,$ A'_n, $A'_{n+1}..$, $A'_{n+k} \models B'$.

This concludes the proof that (1) entails (*). To prove (1) we have to proceed in two steps. The second step consists in proving (1) by induction on the complexity of formulas. But before we can do that, we first have to prove another fact by induction on the complexity of terms. This fact consists in each term t having the following property (1.a):

(1.a) If M and M' are related in the manner of (1) and **a** is any assignment in M, then $[[t]^{M,a}$ = is the unique element w_j of U such that$[[P_t(x)]^{M',a[w_j/x]} = 1$.

In the proof of (1.a) we can keep M and M' fixed.

(i) If t is the variable v_i, then $P_t(x)$ is the formula $x = v_i$. In this case there is obviously only one element in U such that $[[x = v_i]^{M',a[w_j/x]} = 1$, namely the element that **a** assigns to v_i.

(ii) Now suppose that $t = f(t_1,..., t_m)$ and that (1.a) has been proved for $t_1,.., t_m$. Let $u_1,.., u_m$ be the objects denoted in M under **a** by $t_1,..,$ t_m, respectively (i.e $u_i = [[t_i]^{M,a}$ for i = 1,..,m. By induction assumption we have that u_i is the unique element of U such that $[[P_{t_i},(x)]^{M',a[u_i/x]} = 1$.

Note further that in this case $P_t(x)$ is the formula

(2) $(\exists x_1)..(\exists x_m)(P_{t_1}(x_1)$ & .. & $P_{t_m}(x_m)$ & $Q_f(x_1,..,x_m,x))$.

Let u be the value of the term t in M under the assignment **a**, i.e. $u = [[t]^{M,a}$. First we show that u satisfies $P_t(x)$ in M' under **a**. This follows from the fact that u is the value which F(f) returns for the arguments $u_1,.., u_m$, since these satisfy the predicates $P_{t_1}(x_1),.., P_{t_m}(x_m)$ in M under **a**. Since by the definition of M' $<u_1,..,u_m,u>$ belongs to the extension of $F(Q_f)$, it follows that u satisfies (2) in M' under **a**, i.e, $[[P_t(x)]]^{M,a[u/x]} = 1$.

Now suppose that u' is an object that satisfies (2) in M' under **a**. We have to show that u' = u. Since u satisfies (2) there are objects $u'_1,..,$

u'_m which satisfy $P_{t_1}(x_1),..,P_{t_m}(x_m)$ in M under **a**. But since by assumption the satisfiers of $P_{t_1}(x_1),.., P_{t_m}(x_m)$ are unique, it follows that for i = 1,..., m, $u'_i = u_i$. From the definition of M' it follows that there is just one object w such that $u_1,.., u_m$,w> belongs to the extension of F(Qf), We already know that u has this property. So if u' has this property too, then u' = u.

To prove (1) we proceed by induction on the complexity of formulas of L. First, let A be an atomic formula of L. Then A is either of the form $P(t_1,..., t_m)$ or of the form t = s. Suppose that A is of the form $P(t_1,..., t_m)$. Then $P(t_1,..., t_m)'$ is of the form

(3) $(\exists x_1)..(\exists x_m)(P_{t_1}(x_1)$ & .., & $P_{t_m}(x_m)$ & $P(x_1,..,x_m))$

Suppose $[[A]]^{M,a} = 1$. Let $\mathbf{a}' = \mathbf{a}[\ [t_1]]^{M,\mathbf{a}}/x_1\ ]..[\ [t_m]]^{M,\mathbf{a}}/x_m\ ]$. By Lemma 2 of p. 18 $[[P(x_1,..,x_m)]]^{M,\mathbf{a}'} = 1$ and so, using the fact that F'(P) = F(P), $[[P(x_1,..,x_m)]]^{M',\mathbf{a}'} = 1$. By property (1,a) $[[P_{t_i}(x_i)]]^{M',\mathbf{a}'} = 1$ for i = 1,.., m. So the conjunction $P_{t_1}(x_1)$ & .. & $P_{t_m}(x_m)$ & $P(x_1,..,x_m)$ is satisfied in M' by **a**'. Therefore (2) is satisfied in M' by **a** (using the clause for the existential quantifier in the Truth Definition).

Conversely, assume that $[[(2)]]^{M',a} = 1$. Then there are $u_1,.., u_m$ in U such that $[[P_{t_1}(x_1)$ & .., & $P_{t_m}(x_m)$ & $P(x_1,..,x_m)]]^{M',a'} = 1$, where $\mathbf{a}' = \mathbf{a}[u_1/x_1\ ]\ ..,\ [u_m/x_m\ ]$. From $[[P_{t_i}(x_i)]]^{M',a'} = 1$ we can infer, using (1.a), that $u_i = [[t_i]]^{M,a}$. Since also $[[P(x_1,..,x_m)]]^{M',a'} = 1$, this entails that $F'(P)(<[t_1]]^{M,a}\ ,..,\ [t_m]]^{M,a}>) = 1$, and, using once more that F'(P) = F(P), we conclude that $F(P)(<[t_1]]^{M,a},..,\ [t_m]]^{M,a}>) = 1$, which comes to the same thing as $[[P(t_1,..,t_m)]]^{M,a} = 1$, i.e. $[[A]]^{M,a} = 1$.

The case where A is the formula t = s can be dealt with in essentially the same way.

What remains are the inductive steps in the proof of (1). These are largely routine. Suppose - to take one of the least uninteresting steps - that A is the formula $(\exists v_i)B$. In this case A' will be the formula $(\exists v_i)B'$, where B' is the translation of B.

Suppose that $[[A]]^{M,\mathbf{a}} = 1$ By the Truth Definition there is a u in U such that $[B]]^{M,\mathbf{a}[u/v_i]} = 1$. Then, by the Induction Hypothesis, $[B']]^{M',\mathbf{a}[u/v_i]} = 1$. So by the Truth Def. $[(\exists v_i)B']]^{M',\mathbf{a}} = 1$. But $(\exists v_i)B'$ =

$((\exists v_i)B)' = A'$. So $[A']]^{M',\mathbf{a}} = 1$. The converse direction is proved analogously.

All other inductive steps of the proof of (1) are similar to this one, or even simpler. This concludes the proof of (1) and thus of the exercise.
q.e.d.

<u>The Craig Interpolation Theorem.</u>

First order predicate logic has several properties which seem very plausible and natural, but which do not obtain for systems of formal logic in general. One of these is the *interpolation property*. A formalism (such as first order predicate logic) is said to have this property if the following holds:

(ip) Suppose A and B are formulas such that $A \models B$. Then there is a formula C in the common vocabulary of A and B such that $A \models C$ and $C \models B$.

Explicating what is meant by "in the common vocabulary of A and B" depends in general somnewhat on the specification of the logical system in question. But for the case of first order predicte logic the explication is straightforward: C *is in the common vocabulary of* A *and* B iff every non-logical constant occurring in C occurs both in A and in B.

Another way to put this is as follows. Let L_A be the language whose non-logical constants are those occurring in A, let L_B be defined analogously and let L_{AB} be the language $L_A \cap L_B$. Then C is in the common vocabulary of A and B iff C is a formula of the language L_{AB}.

The claim that first order predicate logic has the interpolation property can thus be stated as follows:

<u>Thm.</u> (Craig Interpolation Theorem)

Let A and B be sentences of first order predicate logic such that $A \models B$. Then there is a sentence C of the language $L_{AB} = L_A \cap L_B$, such that $A \models C$ and $C \models B$.

Proof. The proof of this theorem is surprisingly easy when we build upon the completeness proof given in this Appendix, in which correctness and completeness have been proved for the method of proof by semantic tableau construction. This is the way we will proceed here. (Another proof of the Interpolation Theorem, which builds on the completeness proof given in the main part of this chatper, can be found in Ch. 2)

Before we start with the proof itself, first a trivial but useful observation. We can rephrase the interpolation property as in (1)

(1) Let A and B be sentences of first order predicate logic such that $A \models B$. Then there is a sentence C of the language $L_{AB} = L_A \cap L_B$, such that $A \models C$ and $\neg B \models \neg C$.

Suppose that $A \models B$. Then, by the Completeness Theorem, the semantic tableau for the argument <A | B> will close. This closed tableau will be finite and thus in particular it will have finitely many end nodes. An end node s of a closed tableau always means that closure has been obtained in the step that led to the construction of s; in other words, $D_F(s)$ contains a pair of signed formulas <E,T> and <E,F> (i.e. a pair with the same formula E but opposite signs) that are responsible for closure of the branch of which s is the last node, i.e. two signed formulas with opposite signs but the same formula E. Each of these formulas is either obtained via 0 or more of successive reductions from A, or else is obtained in this way from B. The end nodes that are of special interest for the construction of the inerpolating sentence C are those where one of the two signed formulas that produce closure comes from A and the other from B. In that case E will belong to L_{AB}, and can be used as a piece in the construction of C. Moreover, we can then show that the formulas in the given branch which stem from A entail E while the formulas in the branch stemming from B entail $\neg$ E, or vice versa. (Details follow presently.) The other two types of end nodes - (i) both signed formulas stem from A or (ii) both signed formulas stem from B - must be handled in a slightly different way. For instance, suppose that both signed formulas that produce closure stem from A. That means that the set of all formulas stemming from A in the branch of which s is the end node entail a contradiction. This means that we can choose a contradictory sentence $\perp$ from L_{AB} (e.g. $(\exists v_1)v_1 \neq v_1$) to give us the piece for the construction of C contributed by this node. The formulas occurring in the branch that stem from A entail $\perp$ in this case whereas those stemming B (trivially) entail $\neg \perp$. The case where both signed formulas that produce closure stem from B can be handled analogously.

In this way we can associate with each end node a pair of formulas (C, ¬ C) from L_{AB}. We can then work our way up from the end nodes to the root, constructing at each step a pair for formulas (C, ¬ C) for the given mother node on the basis of such assignments to her daughter node or nodes. In the end we arrive at such a pair for the root <>. The C of that pair will then be the interpolating formula we are looking for.

To make this precise we must begin by defining the notion "stemming from". This is quite simple. Given an argument <A | B>, we can annotate every formula that gets produced in the course of the tableau construction with "A" or "B", depending on whether it comes from the first or the second of these formulas. The simplest way to do this is to extend the signature of a formula with an additional slot, to be occupied by either "A" or "B". Thus a signed formula will now have the form of a triple[41] <E,T/F,A/B>, where E is a formula, the second slot is filled with either a "T" or an "F" depending on whether the formula is meant to be true or false, and the third slot has an "A" or a "B" depending on whether the signed formula stems from A or from B. The premise A and the conclusion B are of course marked as "stemmng from themselves"; that is, D_F(<>) = <<A,T,A>,<B,F,B>>. Furthermore, the "stemming from" information is simply passed on from each signed formula to the one or two that result(s) from its reduction. (For instance, when the formula <G&H,T,A> is reduced at node s, then the new formulas added to D_F(s) in the transition to $D_F(s^\frown 0)$ are <G,T,A> and <H,T,A>.)

Given this information about the origin of the formulas which occur in the sequences D_F(s) it is possible to associate with a node s a formula that "conjoins" all the formulas that are part of the decoration of s or any of its predecessors. Let DESC(A,s) be the set of all formulas E such that <E,T,A> occurs in the decoration of s or in that of some predecessor of s, and of all formulas ¬ E, such that <E,F,A> occurs in the decoration of s or in that of some predecessor of s; and let REPR(A,s) be the conjunction of all the formulas in DESC(A,s); similarly for DESC(B,s) and REPR(B,s).

---

41 As before, universally quantified formulas marked "T" and existentially quantified formulas marked "F" involve as an additional component of their signatures the set of constants with which their quantifiers have already been instantiated. So in the case of such formulas signed formulas are now 4-tuples..

In order to formulate the precise hypothesis that we will be able to pull through the mentioned backwards induction, there is one more matter we need to address. Tableau construction involves the introduction of new constants. We have built a mechanism for recording which constants have been introduced by the time a tableau node s has been reached, viz. by including the sequence $D_C(s)$ in the decoration of s. The constants in $D_C(s)$ can occur in the formulas that occur within $D_F(s)$ and that is so in particular for those formulas associated with an end node s which produce the closure of the branch of which s is the last node. This means that in such cases we cannot assume that the formula C we want to construct for s belongs to the language L_{AB}. Rather, we will only be able to assume that it belongs to the language we will call $L_{AB,s}$, the language whose non-logical constants are those of L_{AB} together with the constants in $D_C(s)$.

We are now ready to formulate the hypothesis we will be able to prove by "backwards induction" on the nodes of the closed tableau <T,D> for <A | B>:

(2) For each node s of the tree T for < A | B> there is a sentence C from the language $L_{AB,s}$, such that REPR(A,s) $\models$ C and REPR(B,s) $\models \neg$ C.

That we can find a C of the required kind for each of the end nodes of T has already been shown. (Now that we have defined REPR(A,s) and REPR(B,s) explicitly, it is easy to verify that the claims we made abut the three types of end nodes earlier are true in the precise formal sense of (2).) To prove the inductive steps of the argument we once again consider only a few representative cases.

(i) Suppose that the formula reduced at the node s is $\neg$ G, that G stems from A and that the sign of G is T. We assume that a sentence C from the language $L_{AB,s^\frown 0}$ has already been associated with s's one successor node $s^\frown 0$ and that (2) holds for $s^\frown 0$ and this C. The difference between REPR(A,s) and REPR(A,$s^\frown 0$) is in this case merely that REPR(A,$s^\frown 0$) contains a conjunct corresponding to the signed formula <G,F,A>. But this conjunct is just $\neg$ G, and that formula is also part of the conjunction REPR(A,s) because of the presence of <$\neg$ G,T,A> in $D_F(s)$. So REPR(A,s) and REPR(A,$s^\frown 0$) are logically equivalent. Moreover, we have in this case that $L_{AB,s^\frown 0} = L_{AB,s}$. So we can take for the sentence associated with s C itself. Then REPR(A,s) $\models$ C; and since REPR(B,$s^\frown 0$) is identical with REPR(B,s), also REPR(B,s) $\models \neg$ C.

(ii) Suppose now that the formula reduced at the node s is $\neg$ G, that G stems from A, but that the sign of G is F. Again we assume that a sentence C from the language $L_{AB,s^\frown 0}$ has been assigned to $s^\frown 0$. In this case the difference between REPR(A,s) and REPR(A,$s^\frown 0$) is that REPR(A,$s^\frown 0$) has the additional conjunct G. However, REPR(A,s) has $\neg\neg$G as a conjunct (because of the signed formula <$\neg$G,F,A> in the decoration of s). So again REPR(A,s) and REPR(A,$s^\frown 0$) are logically equivalent and (2) follows for s.

(iii) Now consider the case where the reduction of s involves the signed formula <G&H,F,A>. Then s has two successors $s^\frown 0$ and $s^\frown 1$. Suppose that for both of these we have sentences C_0 and C_1 satisfying (2). Note that in this case REPR(A,$s^\frown 0$) has, as compared to REPR(A,s), the additional conjunct $\neg$G and that REPR(A,$s^\frown 1$) has the additional conjunct $\neg$H. So REPR(A,$s^\frown 0$) is logically equivalent to (REPR(A,s) & $\neg$G) and REPR(A,$s^\frown 1$) to (REPR(A,s) & $\neg$H). We further note that REPR(A,s) has as one of its conjuncts the formula $\neg$(G&H) and finally that $L_{AB,s^\frown 0} = L_{AB,s^\frown 1} = L_{AB,s}$. Let the sentence C associated with s be (C_0 v C_1). Then, since (REPR(A,s) & $\neg$G) $\models$ C_0, (REPR(A,s) & $\neg$G) $\models$ C_0 v C_1, and by an analogous argument (REPR(A,s) & $\neg$H) $\models$ C_0 v C_1. So (REPR(A,s) & ($\neg$G v $\neg$H) $\models$ C_0 v C_1. But $\neg$G v $\neg$H is logically equivalent to $\neg$(G & H), and that formula is a conjunct of REPR(A,s). So again REPR(A,s) and REPR(A,$s^\frown 0$) are logically equivalent, and it follows that REPR(A,s) $\models$ C.

We further note that REPR(B,$s^\frown 0$) = REPR(B,$s^\frown 1$) = REPR(B,s) in this case. by induction assumption we have that REPR(B,$s^\frown 0$) $\models$ $\neg C_0$ and REPR(B,$s^\frown 1$) $\models$ $\neg C_1$. So REPR(B,s) $\models$ $\neg C_0$ and REPR(B,s) $\models$ $\neg C_1$. Therefore REPR(B,s) $\models$ $\neg C_0$ & $\neg C_1$ and so REPR(B,s) $\models$ $\neg$(C_0 v C_1), i,e, REPR(B,s) $\models$ $\neg$C. This concludes the proof of case (iii).

(iv) Now suppose the reduction at s is of the signed formula <($\exists v_i$)G,T,A>. In this case a new constant c_k has been introduced in the transition from s to $s^\frown 0$, i.e. $L_{AB,s^\frown 0} = L_{AB,s} \cup \{c_k\}$. REPR(A,$s^\frown 0$) now has besides the formulas from REPR(A,s) as new conjunct the formula G[c_k/v_i]. So we have by induction assumption: REPR(A,s) & G[c_k/v_i] $\models$ C', where C' is the sentence from $L_{AB,s^\frown 0}$ that has been associated with

$s^\frown 0$. We can rewrite this as REPR(A,s) $\models G[c_k/v_i] \rightarrow C'$. Since c_k does not occur in REPR(A,s), it follows that

$$\text{REPR(A,s)} \models G[c_k/v_i][v_r/c_k] \rightarrow C'[v_r/c_k] \qquad \text{(i)}$$

where v_r is a variable not occurring in either G or C'. (Here, as always, $C'[v_r/c_k]$ is the result of replacing all occurrences of c_k in C' by v_r and, similarly, $G[c_k/v_i][v_r/c_k]$ the result of replacing all occurrences of c_k in $G[c_k/v_i]$ by v_r. Note that $G[c_k/v_i][v_r/c_k]$ has free occurrences of v_r in all and only those positions in which G has free occurrences of v_i. So we may write "$G[c_k/v_i][v_r/c_k]$" also as "$G[v_r/v_i]$".

From (i) we can infer (ii) and from (ii) we infer (iii) since the right hand side of (iii) follows logically from the right hand side of (ii).

$$\text{REPR(A,s)} \models (\forall v_r)(G[c_k/v_i][v_r/c_k] \rightarrow C'[v_r/c_k]) \qquad \text{(ii)}$$

$$\text{REPR(A,s)} \models (\exists v_r)G[v_r/v_i] \rightarrow (\exists v_r)C'[v_r/c_k] \qquad \text{(iii)}$$

It is easy to verify that $(\exists v_i)G \models (\exists v_r)G[v_r/v_i]$. ($(\exists v_i)G$ and $(\exists v_r)G[v_r/v_i]$ are alphabetic variants; see Section 1.1 of this chapter.) Moreover, $(\exists v_i)G$ is a conjunct of REPR(A,s). We now choose as sentence C associated with s the sentence $(\exists v_r)C'[v_r/c_k]$. Note that c_k does not occur in C, so that C belongs to $L_{AB,s}$. From what has been argued it is clear that REPR(A,s) $\models$ C. On the other hand, by induction assumption REPR(B,$s^\frown 0$) $\models \neg C'$. Since the reduction step which leads from s to $s^\frown 0$ does not involve a formula stemming from B we have once more that REPR(B,s) = REPR(B,$s^\frown 0$). So REPR(B,$s^\frown 0$) has no occurrences of c_k. Therefore, it follows from the Induction Hypothesis that REPR(B,$s^\frown 0$) $\models \neg C'[v_r/c_k]$. So REPR(B,$s^\frown 0$) $\models (\forall v_r)\neg C'[v_r/c_k]$. Since $(\forall v_r)\neg C'[v_r/c_k]$ is logically equivalent to $\neg(\exists v_r)C'[v_r/c_k]$, we conclude that REPR(B,s) $\models \neg$C.This concludes the proof of case (iv).

(v) Finally suppose the reduction at s is a reduction of the signed formula <$(\exists v_i)$G,F,A>. In this case the reduction step involves instantiating the quantifier of $(\exists v_i)G$ by a constant c_k that belongs to $L_{AB,s}$. So $L_{AB,s^\frown 0} = L_{AB,s}$. Again, let C' be the sentence associated with $s^\frown 0$. The new conjunct of REPR(A,$s^\frown 0$) is now $\neg G[c_k/v_i]$, whereas $\neg(\exists v_i)G$ is a conjunct of REPR(A,s). Since $\neg(\exists v_i)G \models \neg G[c_k/v_i]$ and since by induction assumption REPR(A,s$^\frown$0) $\models$ C', it follows that REPR(A,s) $\models$ C'. So we can take for the sentence associated with s simply this same C.

All other inductio steps are closely similarr to one of those we have presented. So we may consider the proof of (2) as completed.

Applying (2) to the root <> we obtain a sentence C in the language $L_{A\,B}$ such that A $\models$ C and $\neg$ B $\models$ $\neg$ C. This proves the theorem.

q.e.d.