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 $f(\alpha) = \langle s,n \rangle$ 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 v 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 $(\forall v_i)$ ($(\exists v_i)$).

Def. 4 (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 v 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.

Def. 5 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, $[[t]]^{M,a}$, is defined thus:

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

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

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

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

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

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

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

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

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

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

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

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

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 $v_i \in X$, $\mathbf{a}(v_i) = \mathbf{b}(v_i)$. 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 $x \in X$ $\mathbf{a}(x) = \mathbf{b}(x)$, 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

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

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

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

1.1.3 Interlude about Proofs by Induction

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):

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

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".

Def.9

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: $\Gamma \models A$) 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 l 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.

Def. 10. (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 **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 $\mathbf{a}'$ 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 $j \neq i$. 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 $\mathbf{a}$ 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 $j \neq i$. 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 $\mathbf{a}$ 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 $d \in U_M$ $[[C[t'/v_i]\ ]]^{M,\mathbf{a}[d/v_j]} = 1$ (*)

By the induction assumption,

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

where $\mathbf{a}''$ is the assignment $\mathbf{a}[d/v_j]\ [\ [[t']]^{M,\mathbf{a}\ [d/v_j]}/v_i]$.
We now make use of the fact that v_j does not occur in t'.
Because of this $[[t']]^{M,\mathbf{a}\ [d/v_j]} = [[t']]^{M,\mathbf{a}}$. 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 $d \in U_M$ $[[C]]^{M,\mathbf{a}[\ [[t']]^{M,\mathbf{a}}/v_i][d/v_j]} = 1$ (**)

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

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

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

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

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:

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

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

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

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 v1 - 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.

Exercise:

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:

Def. 10:

(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:

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

(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: $\Gamma \vdash_L A$ for "there exists a proof in L of A from Γ".

Lemma 3: Supppose that $L \subseteq L'$ (i.e. the function L' extends the function 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.

Proof. 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):

1.2 Soundness and Completeness of the axiomatic proof system of Section 1.1.3

Theorem 1 (Soundness): If $\Gamma \vdash A$, then $\Gamma \models A$

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

Proof of Theorem 1. 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 $B \in \Gamma$ 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 $\langle C_1, ..., C_{n-1}, C_n \rangle$ 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) $C_n \in \Gamma$;
(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 $u \varepsilon U_M$. Let $u \varepsilon U_M$. 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 $u \varepsilon U_M$, $[[(\forall v_i)A]]^{M,a} = 1$.

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 $\bot$ be short for the following formula:

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

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

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

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

T10. $\neg\neg\bot \vdash' \bot$

T11. $\vdash' \neg\bot$

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

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 $\Gamma \cup \{\neg B\} \vdash' \bot$

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. $\neg A \rightarrow \bot \vdash A$

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 $P = < C_1, ..., C_n>$ is a proof of B from Γ U {A} in which there are no applications of UG. Note that for each $i < n$, the initial segment $< C_1, ..., C_i>$ is a proof (without UG) of C_i from Γ U {A}.
We transform P into a proof $<D_1, ..., D_{f(n)}>$ 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 $<C_1>$ 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 $< C_1, ..., C_n>$ 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.

Proof of T31.

⇒ As in the proof of T2.

⇐ 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)	$A \rightarrow (\forall v_i jD$	(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.

Def. 10' (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 $(\exists v_j)A'[v_j/v_k]$.

Remark 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.)

Lemmma. 3' 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 $(\exists v_r)A[v_r/v_i]$.

Exercise: 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.

Lemma 3''. 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 $\vdash A \leftrightarrow A$ by T33.

(ii) Suppose that $\vdash A \leftrightarrow A'$ and $\vdash B \leftrightarrow B'$. Then by the first two theorems liset under T34 $\vdash \neg A \leftrightarrow \neg A'$ and $\vdash (A \,\&\, B) \leftrightarrow (A' \,\&\, B')$. 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 $\vdash A \leftrightarrow A'$. 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:

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

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

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

By UG we can infer from this:

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

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.

1.2.2 The core of the Completeness Proof.

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 U $\{c_1, c_2, \ldots\}$.[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) $\Delta_0 = \Gamma$

$\Delta_i \cup \{A_{i+1}\}$ if Δ_i U $\{A_{i+1}\}$ is consistent and A_{i+1} is not of the form $(\exists v_j)B$

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

$\Delta_i \cup \{\neg A_{i+1}\}$ otherwise

Let $\Delta_\omega = U_{i \varepsilon \omega} \Delta_i$. 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

$$\begin{array}{l} C'_1 \\ C'_2 \\ \quad . \\ \quad . \\ \quad . \\ C'_{n-1} \\ B[v_k/v_j] \rightarrow \bot \end{array} \qquad (4)$$

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

$$(\forall v_k)(B[v_k/v_j] \rightarrow \bot) \qquad (5)$$

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

$$(\exists v_k)B[v_k/v_j] \rightarrow \bot \qquad (6)$$

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).

$$(\exists v_j)B \rightarrow \bot \qquad (7)$$

We now have a proof of $(\exists v_j)B \rightarrow \bot$ from $\Delta_n \cup \{(\exists v_j)B\}$. So by T31 we have a proof of $\bot$ 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

$$\Delta_n \cup \{A_{n+1}\} \vdash \bot \qquad (8)$$

and

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

From (12) we get, by T31 and T6

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

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 $B \varepsilon \Delta_\omega$ nor $\neg B \varepsilon \Delta_\omega$. 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

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

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

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

and

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

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 $B \varepsilon \Delta_\omega$ nor $\neg B \varepsilon \Delta_\omega$ 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 $M = \langle U,F \rangle$ 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)

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

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 $t_i = c'_i \in \Delta_\omega$. So by induction hypothesis,

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

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

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

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

$$g(c'_1, \ldots, c'_n) = c \in \Delta_\omega \quad (5)$$

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

Now suppose that $t = c \in \Delta_\omega$. Again choose c'_i (i = 1, ..., n) such that $t_i = c'_i \in \Delta_\omega$. 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

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

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.

Lemma 3'. (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

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

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

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

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

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

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

(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.) $[[A]]^{M,a[F(c)/v_i]} = 0$ iff $[[\ B\ ]]^{M,a[F(c)/v_i]} = 1$.

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

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 $B \in \Delta_\omega$. Then, by the construction of Δ_ω, $A[c_r/v_i] \in \Delta_\omega$ for some constant c_r. So, by induction hypothesis, $[[A[c_r/v_i]\ ]]^{M,a} = 1$. So, by Lemma 3', $[[A]]^{M,a[F(cr)/v_i]} = 1$. So there is some u in U_M such that

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

Now suppose that $[[B]]^{M,a[F(c)/v_i]} = 1$. Then, by the truth definition, there is some u in U_M such that $[[A]]^{M,a[u/v_i]} = 1$. but if $u \varepsilon U_M$, 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 $A[c/v_i] \varepsilon \Delta_\omega$. So, since $\vdash A[c/v_i] \rightarrow (\exists v_j)A$, $(\exists v_j)A \varepsilon \Delta_\omega$.

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_o} 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_o} we can choose: "(C falls under C_o iff) $\neg\exists$ (x) C(x)". Moreover, Frege realized that when the natural numbers 0, 1, 2, ... are identified with the concepts C_o, 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_o} and m_{C_o} 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_o} is true while the otherc is the extension of a concept that n_{C_o} 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 "¬ (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 "¬ (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 "¬ (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 ⊆ V ⊆ X. Then f[U] ⊆ f[V]; so Y\ f[V] ⊆ Y\ f[U]; so g[Y\ f[V]] ⊆ g[Y\ f[U]]; so X\g[Y\ f[U]] ⊆ 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 ∉ 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 ∉ 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 ≠ x'. We must show that h(x) ≠ h(x'). If x, x' ε W, then h(x) ≠ 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) ≠ h(x'). Lastly suppose x ε W, x' ε X\W. Then h(x) ε

f[W] and h(x') ε Y\f[W], so again h(x) ≠ 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 $\emptyset$ 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.

Def. 14 (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.)

Postulate. (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 $\{Y: Y \subseteq X \ \& \ \varnothing \in Y \ \& \ (\forall x)(x \in Y \rightarrow x \cup \{x\} \in Y\})$. Since $X \in Z$, 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.

Def. 15 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 $C \in \Gamma$, $M \models C$.

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

Cor.1 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.

Cor. 2. (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.

Exercise. 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 $\langle\Gamma,B\rangle$ will involve only fnitely many premsies from Γ, So the argument $\langle\Delta,B\rangle$, 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:

Thm. 5. 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 $h(\mathbf{a}(v_i))$.

Exercise: 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 $M \models A$ iff $M' \models A$. We will state this corollary using the concept of elementary equivalence:

Def. 17 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, $M \models A$ iff $M' \models A$.

Prop. 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 = <U,F> 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 $F(P)(<u_{i_1},..,u_{i_k}>) = 1$, 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 $F(P)(<u_{i_1},..,u_{i_k}>) = 0$. 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 $F(g)(<u_{i_1},.., u_{i_k}>) = u_j$ and all formulas $\neg (g(v_{i_1},..,v_{i_k}) = v_j)$, such that $F(g)(<u_{i_1},.., u_{i_k}>) \neq u_j$. 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) $(\exists v_1) ... (\exists v_n)((\Lambda_{i\neq j}\, v_i \neq v_j) \,\&\, (\forall v_{n+1})\, V_i(v_{n+1} = v_i) \,\&\, B)$

We will refer to the part of A_M which follows the initial block of existential quantifiers $(\exists v_1) ... (\exists v_n)$ 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(ecutive)-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 $Cl_L(\Gamma) = Cl_{L'}(\Gamma) \cap \{A: A \text{ is a sentence of } L\}$

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

---

20 Recall that $\bot$ 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 $T \cup T' \models \bot$.

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.

Def. 20 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 $\Gamma\backslash\{A\} \models A$.
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 $M = \langle U, F\rangle$ and $M' = \langle U', F'\rangle$ 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, $F'(P)(\langle a_1, ..., a_n\rangle) = F(P)(\langle a_1, ..., a_n\rangle)$
(iii) for each n-place functor f of L and elements $a_1, ..., a_n$ of U, $F'(f)(\langle a_1, ..., a_n\rangle) = F(f)(\langle a_1, ..., a_n\rangle)$

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

When $M = \langle U,F\rangle$ is a submodel of the model $M' = \langle U',F'\rangle$ for L, we sometimes denote M as "$M' \restriction U$".

If the language L does not contain any function constants, then there exists for every model $M' = \langle U', F'\rangle$ for L and non-empty subset U of U' a (unique) submodel $M = \langle U,F\rangle$ 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 $a_1, ..., a_n \in U$, such that $F'(f)(\langle a_1, ..., a_n\rangle)$ 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.

Def. 22 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: $M = M' \restriction L$, 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.

Prop. 6 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.)

Def. 23 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 $M' \models A$, then $M \models A$.

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

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 $\forall\exists$-formulas (formulas consisting of a block of universal quantifiers followed by a block of existential quantifiers followed by a quantifier-free part), nor $\forall\exists$-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' = <{a,b},F'> and its submodel M = <{a},F>, where F(P)(<a>) = F'(P)(<a>) = 0 and F'(P)(<b>) = 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.

Thm 8. 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:

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

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 $A_1, ...,A_n$ from G such that $\models (A_1 \&...\& A_n) \rightarrow B$. It is easily seen that the conjunction $A_1 \&...\& A_n$ of the purely universal sentences $A_1, ...,A_n$ is logically equivalent to a single purely universal sentence A. (First rename the bound variables of $A_1,...,A_n$ in such a way that they are all different, i.e. that no two quantifiers in $A_1 \&...\& A_n$ 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, ...,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 $D_1, ...,D_k$ from D(M') such that $\models (D_1 \&...\& D_k) \rightarrow \neg B$, or, equivalently,

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

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:

Thm. 9. 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 $r,s \varepsilon \mathbb{Q}$ and $i = \sqrt{-1}$)
   i. the set of all pairs $\langle n,m \rangle$ 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 $X \cap Z = \emptyset$.
(i) Show: $Y \cup Z \preccurlyeq X \cup Z$.
(ii) Show that the condition that $X \cap Z = \emptyset$ 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_M such that for every model M' for L:

M' $\models A_M$ 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:

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

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.

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

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

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:

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

(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 $<U, \text{dass } U, \underline{\cap}, \underline{\cup}, \underline{\neg}, \underline{0}, \underline{1}>$ 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:

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

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

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. $\mathbf{n}_o < c_1$, for n = 1,2,3, ... ,
where for any natural number n, $\mathbf{n}_o$ 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 $\mathbf{n}_o$ 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 $X \cap Z = \emptyset$.
(i) Show: $Y \cup Z \preccurlyeq X \cup Z$.
(ii) Show that the condition that $X \cap Z = \emptyset$ 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 $Z \cup (Y \setminus X)$ to Y. Since Z is a proper subset of X, $Z \cup (Y \setminus X)$ 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 $Y' \setminus \{f^{n+1}(u)\}$. 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 $x \in X'$, then g(x) is that element x' such that (i) $x' \in X$, (ii) $x' = f^n(x)$ for some n, and (iii) there is no positive $m < n$ such that $f^m(x) \in X$. Then g is a bijection from X' to the proper subset $X' \setminus \{u\}$ of X'. This contradicts the assumption that X is finite.

(The following 'solution' to (5) is not correct. What is the mistake?

'Solution' to (5). 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 $u \in X$.

First assume $f(u) \in X$. Note that $f[X] = (f[X] \cap X) \cup (f[X] \cap (Y \setminus X))$. So $X = f^{-1}[(f[X] \cap X)] \cup f^{-1}[(f[X] \cap (Y \setminus X))]$. Put $X_1 = f^{-1}[(f[X] \cap X)]$ and $X_2 = f^{-1}[(f[X] \cap (Y \setminus X))]$. Clearly, $X_1 \cap X_2 = \varnothing$ and $X_1 \cup X_2 = X$. define the function g on X as follows: (i) for $x \in X_1$, $g(x) = f(x)$; (ii) for $x \in X_2$, $g(x) = x$. Then $DOM(g) = X$, $RAN(g) \subseteq X$ and g is 1-1. Moreover, $u \in (X \setminus g[X])$; 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 $f(u) \in Y$. If $f[Y] \subseteq Y$, then we are done. For then f[Y] is a proper subset of Y, since $f(u) \in Y \setminus f[Y]$, 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 $f[Y] \subseteq Y$. So there is a $y \in Y$ such that $f(y) \in X$. 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 $v \in X \cup Y$ such that $v \neq u$ and $u \neq y$, $f'(v) = f(v)$.) Then we have that $f'[X \cup Y] = f[X \cup Y] = Z$, u does not belong to $f'[X \cup Y]$ and $f'(u) \in X$. This reduces the second case to the first. The cxase where $u \in Y$ is com,pletely parallel to that where $u \in X$.)

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 $<A_1,..., A_n \mid B>$ 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 $<A_1,..., A_n \mid B>$ 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 $<A_1,..., A_n \mid B>$ closes, then $<A_1,..., A_n \mid B>$ has no counter-model (and thus is semantically valid).

PR2. When the tableau for $<A_1,..., A_n \mid B>$ has an open branch, then $<A_1,..., A_n \mid B>$ 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	TRUE	FALSE	(¬,F)
	¬ C	||		|| ¬ C	
		|| C	C	||	

(v,T)	TRUE	FALSE	TRUE	FALSE	(v,F)
	C v D	||		|| C v D	
	C | D			C	
				D	

(→,T)	TRUE	FALSE	TRUE	FALSE	(→,F)
	C → D	||		|| C → D	
	| D	C |	C	D	

($\leftrightarrow$,T) TRUE FALSE

$$\frac{C \leftrightarrow D \;\|\; \_\_\_\_\_}{\begin{array}{l|l} C & C \\ D & D \end{array}}$$

($\leftrightarrow$,F) TRUE FALSE

$$\frac{\_\_\_\_\_ \;\|\; C \leftrightarrow D}{C \mid D \quad D \mid C}$$

($\forall$,T) TRUE FALSE

$$\frac{(\forall v_i)\, A}{A(t/v_i)} \;\|\; \_\_\_\_\_$$

(t an arbitraryclosed term)

TRUE FALSE ($\forall$,F)

$$\frac{\_\_\_\_\_ \;\|\; (\forall v_i)\, A}{A(c/v_i)}$$

(c a new parameter)

($\exists$,T) TRUE FALSE

$$\frac{(\exists v_i)\, A}{A(c/v_i)} \;\|\; \_\_\_\_\_$$

(c a new parameter)

($\exists$,F) TRUE FALSE

$$\frac{\_\_\_\_\_ \;\|\; (\exists v_i)\, A}{A(t/v_i)}$$

(t an arbitraryclosed term)

(=,Sub) TRUE FALSE

$$\frac{s = t \quad A}{A'} \;\|\; \_\_\_\_\_$$

TRUE FALSE

$$\frac{s = t}{\_\_\_\_\_ \;\|\; A} \qquad \frac{\_\_\_\_\_ \;\|\; A}{A'}$$

(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

$$\frac{\_\_\_\_\_}{t = t} \;\|\; \_\_\_\_\_$$

(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. $(\exists y)(Q(y) \ \& \ (\forall x)(P(x) \rightarrow R(x,y)))$

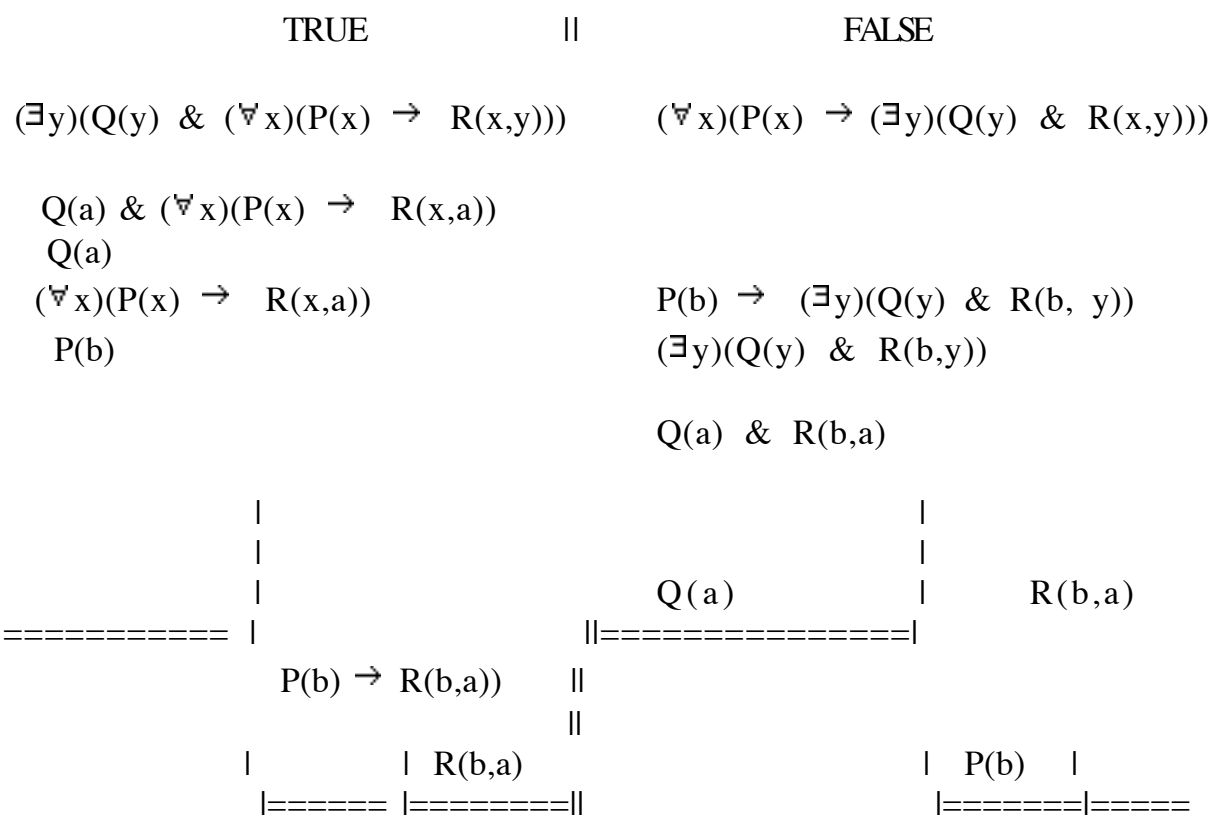
Thus the two arguments are:

(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. $< (\forall x)(P(x) \rightarrow (\exists y)(Q(y) \ \& \ R(x,y))) \mid (\exists y)(Q(y) \ \& \ (\forall x)(P(x) \rightarrow R(x,y))) >$

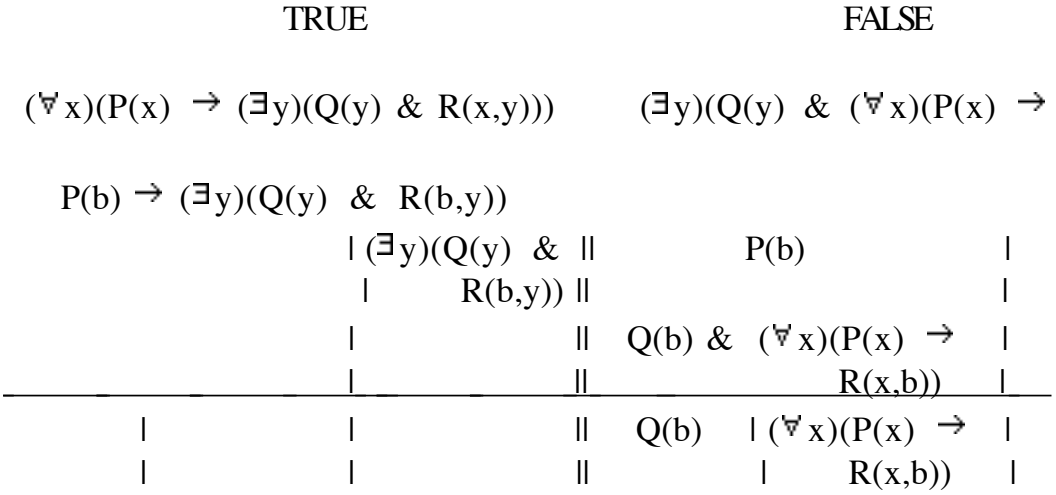
Of these (4.i) is valid and (4.ii) is not. The following two tableaus show this.

(5) (Tableau for (4.i))



Since this tableau closes, we conclude that (4,i) is valid.

(6) (Tableau for (4.ii))



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 $< (\forall x)(\exists y)R(x,y) \mid \varnothing >$).

   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.

Def. DA1. (Trees)

A tree *T* is a pair $<T,\leq>$, 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 $T = <T,\leq>$ 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 $<s_1, ..., s_i>$, then $s^\frown n$ is the sequence $<s_1, ..., s_i, n>$

[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 $<A_1,...,A_n \mid B>$, where the premises $A_1,...,A_n$ 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 $<A_1,...,A_n \mid B>$ can be of two kinds: either c occurs somewhere in $A_1,...,A_n$ 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 $<A_1,...,A_n \mid B>$, 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 $<A_1,..., A_n \mid B>$, 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.)

Def. DA2. (of *signed formula*)

A *signed formula* of L is either:

(i) a pair <A,π>, where A is a sentence of L, π ε {T,F}, and neither of the following two conditions (a), (b) holds:

(a) π = T and A is of the form (∀x)E(x)
(b) π = F and A is of the form (∃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 (∀x)E(x) under TRUE and (∃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 ε 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 $A_1,..., A_n$ 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 $<A_i,T,\varnothing>$; otherwise α_i has the form $<A_i,T>$. Analogously, β is a signed formula establishing B as occurring under FALSE. That is, β has the form $<B,F,\varnothing>$, if B begins with an existential quantifier and otherwise is equal to $<B,F>$.

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} $s \in T$ 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 $<(\forall x)E(x),T,S>$ or of the form $<(\exists x)E(x),F,S>$, 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: γ = <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 $C_{(\vee,T)}$ 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 $C_{(\vee,T)}$.

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 $\varepsilon = <E(c),T,\varnothing>$ 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 $<(\forall x)E(x),T,S>$.

$C_{(\forall,T)}$ Suppose the member α_i of $D(s)_F$ that is up for reduction has the form $<(\forall x)E(x),T,S>$, 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

$$D(s^\frown 0) = <<\alpha_1,., \alpha_{i-1}, \alpha_{i+1},., \alpha_n, \varepsilon, <(\forall x)E(x),T,S\cup\{c\}>>,D(s)_C>;$$

again ε is equal to $<E(c),T,\varnothing>$ 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

$$<(\exists y)(Q(y) \ \& \ (\forall x)(P(x) \rightarrow R(x,y))) \mid (\forall x)(P(x) \rightarrow (\exists y)(Q(y) \ \& \ R(x,y)))>$$

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 $s \in T$, f(s) is specified as follows:

$$f(s) = \begin{cases} \varnothing & \text{in case (i) (s has finitely many successors in T)} \\ s^\frown 0 & \text{in case (ii) } (s^\frown 0 \text{ has infinitely many successors in T)} \\ s^\frown 1 & \text{in case (iii) } (s^\frown 1 \text{ has infinitely many successors in T while } {}^\frown 0 \text{ has finitely many successors in T)} \end{cases}$$

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) $g(0) = \langle\rangle$
(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 $s \leq s'$ nor $s' \leq s$. 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 **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 **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.

Theorem TA2. (Completeness of the Tableau Method)

Suppose that the semantic tableau for the argument $<A_1,..., A_n \mid B>$ is not closed. Then $<A_1,..., A_n \mid B>$ is not valid.

Proof. Suppose that the premises and conclusion of $<A_1,..., A_n \mid B>$ belong to the language L and that the tableau <T,D> for $<A_1,..., A_n \mid B>$ 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,$\varnothing$> 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 $A_1,..., A_n$ are true in M and the conclusion B is false in M, which proves that $<A_1,..., A_n \mid B>$ 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 (= $U \otimes ..(n \text{ times}).. \otimes U$) into the set {0,1}:[40] $F_M(P)(c^1, ..,c^n) = 1$ iff $P(c^1, ..,c^n) \varepsilon PF(Z)$

(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, $M \models P(c^1, .., c^n)$. Now suppose that $P(c^1, ..,c^n) \varepsilon NF(Z)$. Then it is not the case that $P(c^1, .., c^n) \varepsilon PF(Z)$; for if this were the case, then there would be a node s of Z such that $<P(c^1, .., c^n),T>$ and $<P(c^1, .., c^n),F>$ 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 $M \models P(c^1, .., c^n)$.

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 ⊨ C and M ⊨ D. So by the clause for & of the Truth Definition, M ⊨ 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 ⊨ C. But then it also won't be the case that M ⊨ 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 $s \trianglerighteq s'$ 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 ⊨ $E[c_k/x]$ and from this by the Truth Definition that M ⊨ $(\exists x)E(x)$, i.e. M ⊨ 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.

Arguments with identity.

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):

$$c \equiv c' \text{ iff } c = c' \in PF(Z) \qquad (\equiv)$$

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 ≡)
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 ≡" stands for 'Congruence of ≡'. 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 ≡) states that ≡ 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 ≡) 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 ≡) 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 ≡ 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 $c \varepsilon C(Z)$ 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 $P(c^1,...,c^m) \varepsilon PF(Z)$. Then by (ii) above $F_M(P)([c^1]_{\equiv},.., [c^m]_{\equiv}) = 1$. So, in virtue of (iii), $M \models P(c^1,...,c^m)$.

Now suppose $P(c^1,...,c^m) \varepsilon NF(Z)$. To show that it is not the case that $M \models P(c^1,...,c^m)$ we need to show (***):

$$\text{For no } c'^1 \varepsilon [c^1]_{\equiv},..., c'^m \varepsilon [c^m]_{\equiv}, P(c'^1,...,c'^m) \varepsilon PF(Z) \qquad (***)$$

Suppose there were $c'^1 \varepsilon [c^1]_{\equiv},..., c'^m \varepsilon [c^m]_{\equiv}$ such that $P(c'^1,...,c'^m) \varepsilon PF(Z)$. Then by (Con $\equiv$) also $P(c^1,...,c^m) \varepsilon PF(Z)$. 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 $M \models P(c^1,...,c^m)$.

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 $P(c^1,...,c^m)$, 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 $c \varepsilon C(Z)$, $\langle E[c/x],T\rangle$ 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 $\langle E[c/x],T\rangle$ 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 $(\exists x)E(x) \varepsilon NF(Z)$, 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 $E[c/x] \varepsilon PF(Z)$ for no $c \varepsilon C(Z)$. Using the induction assumption we can infer that for no $c \varepsilon C(Z)$, $M \models E[c/x]$. 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.

Remark on the rule (=,Sub).

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 $\langle\Gamma| B\rangle$, 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 $\langle C_n,T\rangle$ (or $\langle C_n,T,\varnothing\rangle$, 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 $<A_1,..., A_n \mid B>$ 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 $<A_1,..., A_n \mid B>$ 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 $<A_1,..., A_n \mid B>$. We translate $<A_1,.., A_n \mid B>$ into an argument $<A'_1,..., A'_{n+k} \mid B'>$ 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

$(\forall x_1)..(\forall x_{n_j})(\exists y)(Qf_j(x_1,..,x_m,y)$ &
$(\forall y)(\forall y')(Qf_j(x_1,..,x_m,y)$ & $Qf_j(x_1,..,x_m,y') \rightarrow y = y'))$

(This sentence says that Qf_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 $\equiv$ is the relation between constants determined by this branch, then $\equiv$ 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 = <U,F> a model for L and let M' = <U,F'> be the model for L' which is obtained by putting:
(i) $F'(\alpha) = F(\alpha)$ for all $\alpha \in L \cap L'$,
(ii) $F'(Q_{f_i},)(\langle u_1,.. u_n, u_{n+1}\rangle) = 1$ 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 $[[Q(x_1,..., x_{m_j},y>]]^{M',a[w_j/y]} = 1$. 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 $\alpha \in L \cap L'$ 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 $\in$ 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 $P_{t1}(x_1),..,P_{tm}(x_m)$ in M under **a**. But since by assumption the satisfiers of $P_{t1}(x_1),..,P_{tm}(x_m)$ 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 $F(Q_f)$, 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) $(\exists x_1)..(\exists x_m)(P_{t1}(x_1)$ & .., & $P_{tm}(x_m)$ & $P(x_1,..,x_m))$

Suppose $[[A]]^{M,a} = 1$. Let **a**' = **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) $[[P_{ti}(x_i)]]^{M',\mathbf{a}'} = 1$ for i = 1,.., m. So the conjunction $P_{t1}(x_1)$ & .. & $P_{tm}(x_m)$ & $P(x_1,..,x_m)$ 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 $[[P_{t1}(x_1)$ & .., & $P_{tm}(x_m)$ & $P(x_1,..,x_m)]]^{M',a'} = 1$, where **a**' = **a**[$u_1/x_1$] .., [u_m/x_m]. From $[[P_{ti}(x_i)]]^{M',a'} = 1$ 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.

The Craig Interpolation Theorem.

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:

Thm. (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. $(\exists v1)v1 \neq v1$) 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 $\neg\neg G$, 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 $<\neg\neg G,F,A>$ 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 $\neg\neg G$ and that REPR(A,$s^\frown 1$) has the additional conjunct $\neg\neg H$. So REPR(A,$s^\frown 0$) is logically equivalent to (REPR(A,s) & $\neg\neg G$) and REPR(A,$s^\frown 1$) to (REPR(A,s) & $\neg\neg H$). We further note that REPR(A,s) has as one of its conjuncts the formula $\neg\neg(G\&H)$ and finally that $L_{AB,s^\frown 0} = L_{AB,s^\frown 1} = L_{AB,s}$. Let the sentence C associated with s be $(C_0 \vee C_1)$. Then, since (REPR(A,s) & $\neg\neg G$) $\models C_0$, (REPR(A,s) & $\neg\neg G$) $\models C_0 \vee C_1$, and by an analogous argument (REPR(A,s) & $\neg\neg H$) $\models C_0 \vee C_1$. So (REPR(A,s) & $(\neg\neg G \vee \neg\neg H) \models C_0 \vee C_1$. But $\neg\neg G \vee \neg\neg H$ is logically equivalent to $\neg\neg(G \& H)$, 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,$s^\frown 0$) $\models \neg\neg C_0$ and REPR(B,$s^\frown 1$) $\models \neg\neg C_1$. So REPR(B,s) $\models \neg\neg C_0$ and REPR(B,s) $\models \neg\neg C_1$. Therefore REPR(B,s) $\models \neg\neg C_0 \& \neg\neg C_1$ and so REPR(B,s) $\models \neg\neg(C_0 \vee C_1)$, i,e, REPR(B,s) $\models \neg\neg C$. 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

$$REPR(A,s) \models G[c_k/v_i][v_r/c_k] \rightarrow C'[v_r/c_k] \qquad (i)$$

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).

$$REPR(A,s) \models (\forall v_r)(G[c_k/v_i][v_r/c_k] \rightarrow C'[v_r/c_k]) \qquad (ii)$$

$$REPR(A,s) \models (\exists v_r)G[v_r/v_i] \rightarrow (\exists v_r)C'[v_r/c_k] \qquad (iii)$$

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.

Chapter II. Mathematical Structures and their Descriptions in First Order Logic.

In this chapter we will look at a few well-known examples of first order theories. These examples are important in their own right, i.e. as formalisations of structures which arise in certain branches of mathematics and other scientific domains. But they will also serve as illustrations of certain general logical issues and we shall use them as opportunities to introduce and discuss those.

The kinds of structures which we will discuss fall into four main classes:

(i) orderings

(ii) certain classes of algebraic structures such as boolean and non-boolean lattices and groups

(iii) the structure of the natural numbers and that of the real numbers with their familiar arithmetical operations + and .

(iv) feature structures

The first order theories of these structures and structure classes we will present will serve as anchors for the discussion of such general issues as: incompleteness, completeness and categoricity of theories; theory extensions and Lindenbaum algebras; quantifier elimination; independence; implicit and explicit definability; equational logic as a subsystem of first order logic; and feature logic as an alternative to first order logic.

2.1 Orderings.

Our first examples concern the concept of order. Mathematically, order is most naturally represented in the form of a binary relation - either a *strict* ordering relation $<$ or a *non-strict* ordering relation $\leq$. (Strict ordering relations are irreflexive and non-strict orderings reflexive. Given a strict ordering $<$ we can define a corresponding non-strict ordering $\leq$ by: $a \leq b$ iff $a < b \vee a = b$; conversely, from a non-strict ordering $\leq$ we get a strict ordering $<$ via: $a < b$ iff $a \leq b \,\&\, a \neq b$.) Orderings can be classified in terms of the properties of $<$ (or, equivalently, of $\leq$). First, there is the distinction between *linear orders* and *partial orders*. In a linear order of a domain D any two distinct

elements a, b of D are ordered in the sense that either a stands in the ordering relation to b or b else stands in the relation to a. In partial orders this condition is in gene4ral not satisfied. (Thus the notion of a partial order is the more general one; linear orders are a special kind of partial order.)

A second important distinction is that between denseness and discreteness. In a dense order there is for any a and b such that $a < b$ an element c such that $a < c < b$; in a discrete order there exists for any a and b such that $a < b$ a c with the properties that (i) $a < c \leq b$ and (ii) there is no d such that $a < d < c$; and, similarly, if $b < a$ then there is a c such that (i) $b \leq c < a$ and (ii) there is no d such that $c < d < a$. (The element c in question is called the *immediate successor* (*predecessor*) *of* a *in the direction of* b.) It should be emphasised that denseness and discreteness are mutually exclusive (in the sense that no non-trivial ordering - i. e. no ordering which holds between at least two different elements - can be dense and discrete at the same time), but that they are not jointly exhaustive: An ordering may be neither dense nor discrete, for instance because it consists of one part which is dense and another which is discrete.

Here we will look at two distinct kinds of ordering structures:

(a) certain linear orders, among them the ordering of the rational numbers, that of real numbers (both dense orderings), that of the narural numbers and that of the integers (both discrete orderings);

(b) partial orders which have the additional property of being *lattices*. A *lattice* is a partial order in which for any two elements a and b there is a "smallest element above both of them" - i.e. an element c such that $a \leq c$ and $b \leq c$ and which is least with regard to this condition, i.e. if c' is any other element such that $a \leq c'$ and $b \leq c'$, then $c \leq c'$ - and, dually, there exists for any a and b a "greatest element that is $\leq$ a and $\leq$ b".)

Lattice-like orders have the important property that they c an be described not only in terms of their orderings, but also interms of the lattice operations $\cup$ and $\cap$, which can be defined in terms of the ordering ($a \cup b$ is the least element above a and b and $a \cap b$ the greatest element below a and b) and which in their turn allow definition of the

ordering relation (e.g. via the definition: a $\leq$ b iff a $\cup$ b = b). Thus lattices can also be viewed as *algebraic structures* or *algebras* - that is, structures consisting of a universe together with a number of functions defined on that universe. (In other words, an algebraic structure is a model for a language L all of whose non-logical constants are function constants.)

Of particular importance among the lattices that we will discuss are the *boolean lattices* (or *boolean algebras*, the term that is used to refer to them when they are pressented as structures involving functions). The logical importance of boolean algebras will no doubt be familiar: classical propositional logic with the connectives & and v has the structure of a boolean algebra.

The order in which we proceed in this section is as follows. We begin with the ordering theory T_{rat} of the rational numbers, presenting the conceptually and historically important theorem of Cantor's according to which any denumerable model of T_{rat} is isomorphic to the ordering structure of the rationals. This will be the basis for introducing the notion of a theory being categorical in a certain cardinality κ. Cantor's Theorem shows that T_{rat} is categorical in the cardinality of the denumerably infinite sets, but as it turns out not in any other infinite cardinality. The subsection closes with a brief discussion of Morley's Categoricity Theorem.

Next, in subsection 2.1.2, we proceed to lattices. We begin with axiomatic characterisations of the class of all lattices, first from the ordering perspective (i.e. formulating our axioms in the first order language $\{\leq\}$ whose only non-logical constant is the 2-place relation $\leq$, and then form the algebraic perspective, using the language $\{\cup,\cap\}$. We show that each of these two theories is definable within the other. We then extend these axiomatisations to obtain theories for the class of all boolean lattices and for that of all boolen algebras, respectively, theories that are again definable within each other. Section 2.1.3 is concerned with the variety of boolean algebras. It presents some particular boolean algebras and some properties in terms of which arbitrary boolean algebras can be classified. 2.1.4 presents the Cech-Stone Representation Theorem, according to which every boolean algebra is isomorphic to (and thus 'can be represented as') a set algebra - a boolean algebra consisting of sets, with the set inclusion relation as the partial order of the lattice. Representation theorems, which assert that every structure with certain abstract properties can be 'represented', or 'realised' as a structure of some more specific

kind, are of great importance in many areas in mathematics; the Cech-Stone Theorem can be regarded as the classical paradigm of theorems of this general form.

The theory of boolean algebras is incomplete, since among its models are boolean algebras that can be distinguished from each other by properties expressible in the language of the theory itself. Even more obvious is the incompleteness of the theory of all lattices, since among its models are on the one hand the boolean lattices, which are also models of the theory of boolean lattices, and on the other hand non-boolean lattices, which are not models of that theory. (Thus the theory of boolean lattices is a proper extension of the general theory of lattices, which proves the latter's incompleteness.) In Section 2.2 we look at incomplete theories from a more general and systematic perspective. The structure consisting of all theories of a given language l., and more generally that consisting of all theories of L which extend a given theory T, are both lattices (though in general not boolean lattices). Thus the study of these structures provides with a further application of lattice theory, as well as giving more insight in the structure of first order logic.

The lattice consisting of all extensions of a given theory T as well as a certain boolean sublattice of this structure, the so-called Lindenbaum algebra of T, are studied in 2.2.1. 2.2.2 contains a discussion of almost complete theories. here we return to linear orderings comparing theories of dense orderings with certain theories of discrete orderings.

2.1.1. The Theory of Dense Linear Orders without End Points.

We choose as our first task in this chapter that of formulating a first order theory that captures all truths about the ordering of the rational numbers. To this end we choose as our language, in which the theory will be formulated, the language $\{<\}$, whose only non-logical constant is the two-place predicate $<$. We will refer to $\{<\}$ also as $L_<$. Our task is thus to state a theory of $L_<$ whose theorems are all and only the truths expressible in $L_<$ about the structure $\langle Q,<_Q\rangle$, where Q is the set of rational numbers and $<_Q$ is the standard ordering of the rationals.

Here is our proposal: Let T_{rat} be the theory consisting of all logical consequences of the following set of axioms:

Def. 1 (Axioms of T_{rat})

L1. $(\forall x)(\forall y)\ (x < y \rightarrow \neg\ (y < x))$
L2. $(\forall x)(\forall y)(\forall z)\ ((x < y\ \&\ y < z) \rightarrow x < z)$
L3. $(\forall x)(\forall y)\ (x < y\ \ v\ \ x = y\ \ v\ \ y < x\ )$
L4 $(\forall x)(\forall y)\ (x < y \rightarrow (\exists z)\ (x < z\ \ \&\ \ z < y))$
L5 $(\forall x)(\exists y)\ (x < y)$
L6. $(\forall x)(\exists y)\ (y < x)$

T_{rat} is also known as the *theory of dense linear orders without endpoints*. The subtheory of T_{rat} that is axiomatised by L0-L3 is known as the *theory of linear orders* and that axiomatised by L0-L2 as the *theory of partial orders*.[1] WE will refer to the first as T_{lin} and to the second as T_{par}.

Some the properties of T_{rat} are stated in Theorem 10.

Theorem 1. (1) Every model of T_{rat} is infinite:
(2) (Cantor) Every two denumerably infinite models of T_{rat} are isomorphic.
(3) T_{den} is complete.

Proof.

(1) Note that because L0 is an axiom of T_{rat} any model of T_{rat} must have at least 2 elements. Secondly, suppose that $M = <U_M,<_M>$ is a finite model of T_{rat}, i.e. that U_M consists of elements $a_1, \ldots, a_n$, where n is some natural number. As just observed, n must be at least 2. Furthermore, since $<_M$ is a linear order, there must be among the elements $a_1, \ldots, a_n$ at the very least one pair of elements (a_i, a_j) such that $a_i < a_j$ and for no a_k, $a_i < a_k$ & $a_k < a_j$. But this contradicts L5.[2]

---

[1] Often axiom L0 is not included in axiomatisations of the theories of linear or partial orderings. leaving it out has the effect that among the models of the hteory one also includes structures of the form $<\{a\}, \varnothing>$, where $\{a\}$ is any singleton set and < is interpreted as the empty relation $\varnothing$. Whether such structures are included or not makes no real difference to what the theory says about the structures which really matter, viz. those in which the universe contains more than one element. In hte present context it has proved to be a little more convenient to exclude them from the start, and thus to include L0 among the axioms.

[2] Strictly speaking the existence of a pair (a_i, a_j) as just stated should be proved. In fact it is easy to prove, by induction on n, that every model of the theory of linear orders whose universe consists of n elements contains such a pair: Suppose this holds for n and let M be a model with universe $\{a_1, \ldots, a_n, a_{n+1}\}$.

(2) Let M, M' be denumerable models of T_{rat} with universes $U_M = \{a_1, a_2, \ldots\}$ and $U_{M'} = \{b_1, b_2, \ldots\}$. We refer to the interpretations $<_M$ and $<_{M'}$ in respectively M and M' of the predicate < as $<$ and $<'$. We construct, by induction on n, partial isomorphisms h_n from M to M' with domains $\{a^1, \ldots, a^n\}$ and ranges $\{b^1, \ldots, b^n\}$. In this notation we assume that $a^1 < \ldots < a^n$ and $b^1 <' \ldots <' b^n$ (and thus that h_n is defined by: $h_n(a^i) = b^i$, for i = 1, .., n. Moreover, the h_n will be constructed in such a way that, putting $h = \bigcup_n h_n$, h is an isomorphism from M to M'.

We proceed as follows. Suppose that the elements $a^1, \ldots, a^n$ and $b^1, \ldots, b^n$ have alredsy been chosen. We distinguish between the case where n is odd and that where n is even.

(a) Suppose n is odd. Then we pick the first element a_i from the enumeration $\{a_1, a_2, \ldots\}$ which does not occur among $\{a^1, \ldots, a^n\}$. For the position of a_i with respect to the $a^1, \ldots, a^n$ there are three possibilities:

(i) $a_i < a^1$;
(ii) $a^n < a_i$;
(iii) $a^k < a_i < a^{k+1}$, for some $k < n$.

(i) Because M' is a model of T_{rat} and T_{rat} contains L4, we know that there is a b among $\{b_1, b_2, \ldots\}$ such that $b <' b^1$. Let b_j be the first such b and let $h_{n+1} = h_n \cup \{<a_i, b_j>\}$. Then h_{n+1} is an isomorphism with $DOM(h_{n+1}) = \{a_i, a^1, \ldots, a^n\}$ and $RAN(h_{n+1}) = \{b_j, b^1, \ldots, b^n\}$.

(ii) This case is just like (i): We know that there is a b in $\{b_1, b_2, \ldots\}$ such that $b <' b^n$, etc.

(iii) This time we make use of L5. Because T_{rat} contains L5 that we may infer that $\{b_1, b_2, \ldots\}$ contains a b such that $b^k <_{M'} b <' b^{k+1}$. Again we let b_j be the first such b. Putting, as before, $h_{n+1} = h_n \cup$

Then consider the restriction M' of M to the set $\{a_1, \ldots, a_n\}$, i.e. the model with universe $\{a_1, \ldots, a_n\}$ in which the interpretation of < is the restriction of the interpretation of < in M to $\{a_1, \ldots, a_n\}$. Since M' has n elements, there is by assumption a pair (a_i, a_j) $(i,j \leq n)$ such that there is no a_k in M' with $a_i < a_k$ & $a_k < a_j$. If it is not the case that $a_i < a_{n+1}$ & $a_{n+1} < a_j$ then (a_i, a_j) is a pair for M of the required kind. If $a_i < a_{n+1}$ & $a_{n+1} < a_j$ then a pair of the desired kind is (a_i, a_{n+1}).

$\{<a_i,b_j>\}$, we conclude that h_{n+1} is an isomorphism with Domain $\{a^1, \ldots, a^k, a_i, a^{k+1}, \ldots, a^n\}$ and Range $\{b^1, \ldots, b^k, b_j, b^{k+1}, \ldots, b^n\}$.

(b) n is even. In this case, let b_i be the first element from the enumeration $\{b_1, b_2, \ldots\}$ which does not occur among $\{b^1, \ldots, b^n\}$ and find, in each of the cases (i) - (iii), an a_j in M which is "similarly situated" with respect to $\{a^1, \ldots, a^n\}$. We put $h_{n+1} = h_n \cup \{<a_j,b_i>\}$.

It is not hard to verify that the union h of all the h_n has for its Domain all of $\{a_1, a_2, \ldots\}$ (because of the steps in the construction for odd n) and that it has for its Range all of $\{b_1, b_2, \ldots\}$ (because of the steps for n even). Moreover, it is obvious from the construction that if a, a' are elements of M and
$a < a'$, then $h(a) <' h(a')$. From linearity (Axiom L3!) it then follows that for all a, a' from M, $a < a'$ iff $h(a) <' h(a')$.

(3) This follows almost directly from (2). Note that if T_{rat} were not complete, then there would be a sentence A from the language $L_<$ such that $\neg(A \in T_{rat})$ and $\neg(\neg A \in T_{rat})$. So it follows that $T_{rat} \cup \{A\}$ and $T_{rat} \cup \{\neg A\}$ are both consistent and thus each of them has a model. Let M_1 be a model of $T_{rat} \cup \{A\}$ and M_2 a model of $T_{rat} \cup \{\neg A\}$. By (i) both models are infinite. So by the downward Skolem-Löwenheim Theorem there are denumerably infinite models M'_1 and M'_2 such that $M'_1 \equiv M_1$ and $M'_2 \equiv M_2$. So A is true in M'_1 and false in M'_2. But by (ii) $M'_1 \equiv M'_2$: contradiction. We conclude that T_{rat} is complete.

q.e.d.

The centre piece of Theorem 1 is part (2). This result is generally known as 'Cantor's Theorem' (or more fully 'Cantor's Theorem about Dense Linear Orders', in order to distinguish this theorem from the equally famous theorem of Cantor that the cardinality of the power set of a given set X always exceeds that of X). The proof of this theorem has, like Cantor's proof of his power set theorem, been a milestone in the development of our understanding of what constitutes valid mathematical reasoning. At first, many mathematicians were very sceptical with regard to the soundness of these proofs. Precisely because their initially controversial status, Cantor's arguments were a major input to the debates over the foundations of mathematics that became a vital concern in the second half of the nineteenth Century and which in its turn provided much of the impetus to the development of formal logic as a fool-proof framework for doing mathematics. (Recall the interlude on Set Theory in Chapter I.)

As opposed to part (2) of Theorem 10, which is specifically about dense linear orderings. the purport of part (3) is much more general. The general statement. known as 'Vaught's test', is this:

Prop. 1 (Vaught's Test)

> Whenever T is a theory which (i) has only infinite models and (ii) is such that for some infinite cardinaltiy κ any two models of T of caridnality κ are isomorphic, then T is complete.

Complete theories are the closest we can get to characterising the properties of a given mathematical structure, when we want to do this by dscribing them within some logical language L. We have already seen some general limits to what can be achieved along these lines, viz. those imposed by the Skolem-Löwenheim Theorems presented in Chapter I. But in fact, for many structures, the best that can be achieved is even farther from the ideal (characterisation of the structure up to isomorphism) than the Skolem-Löwenheim Theorems would in principle allow for.

Let us be more exact. In order that a theory T of a first order language L can be considered a characterisation of some given structure *A*, two conditios must be satisfied. First, all the structural properties of *A* must be expressible in L. That is, we must be able to represent *A* as a model $M_O = <U_O,F_O>$ of L such that each relation that is relevant to the structure of *A* is either given as the interpretation $F_O(\alpha)$ of some non-logical constant α of L or else must be definable in terms of one or more relations $F_O(\alpha)$ with $\alpha \varepsilon$ L. (For a general discussion of notions of definability see Section 2.3.) Second, all sentences of L that are true in M_O must be iderivable from T as theorems (and thus, because T is closed under logical consequence,, must be members of T).

Assume that we have succeeded in choosing a language L such that the first condition is fulfilled - i.e. that we can represent *A* as some particular model M_O of L. In that case there exists - trivially - a unique theory T of L which verifies all and only the sentences of L that are true in M_O, viz, $Th(M_O)$. That the set $Th(M_O)$ always exists follows from general principles of set-theory (which will be spelled out in Ch. 3). But from the general principles which guarantee the existence of $Th(M_O)$ nothing follows that has anything to do in particular with the structure *A* whose properties $Th(M_O)$ describes. What we eally want is a non-trivial characterisation of $Th(M_O)$ that reveals some of the special

properties of $Th(M_o)$, and that ideally gives us some insight into them that might have eluded us without them. A natural way to go about this is to try to find 'axioms' for $Th(M_o)$ - sentences belonging to the theory which on the one hand can be readily verified as true in M_o and on the other as entailiing all other sentences that are true in M_o. It seems particularly desirable form this perspective to find a finite set of axioms for the theory. As we saw in Chapter I, this is always possible whenn *A*, and therewith M_o, are finite. But for infinite structures *A* the situation is very different. For instance, it is an interesting and surprising consequence of Gödel's Incompleteness Theorems that for many infinite structures *A* no finite axiomatisation of $Th(M_o)$ exists. (In fact, the situation is even worse in that there isn't even an infinite recursively enumerable set of axioms for $Th(M_o)$; for 'recursively enumerable' see Ch. ??.)

These negative results hold in spite of the fact that iby requiring only that our theory captures all the truths about *A* that are expressible in L we haven't pitched our aims necessarily very high. There is also another, stricter sense in which one can define complete characterisation of *A* by T

Any model M of T is isomorphic to M_o

(where again M_o is represented as model for the language L of T) Again, when *A* is finite, then, as established by Thm. 6 in Chapter I, a theory T satisfying this requirement can always be found (and when L is also finite, then this theory is finitely axiomatisable, e.g. by the single axiom described in th proof of Thm. 6). But the Skolem-Löwenheim Theorems tell us that this desideratum is never met iwhen *A* is infinite. For as soon as *A* is infinite, $Th(M_o)$ will have models of different infinite cardinalities and these can never be isomorphic to each other. The best we can hope for is that models of $Th(M_o)$ are isomorphic to each other so long as they are of the same cardinality. But even this weaker condition is only seldomly fulfilled and holds only for rather uninteresting structures *A*, with largely trivial structural properties.

In fact, even for the ordering structure $<Q,<_Q>$ of the rationals this weaker requirement is not fulfilled, Cantor's Theorem notwithstanding. For while , as the Thm states, any two *denumerable* models of $Th(<Q,<_Q>)$ (= T_{rat}) are isomorphic, this is not so for non-denumerable models - see Exercise ??.

For easier formulations during the remainder of this section we introduce some further terminology.

Def. 2 A theory T in a first order language L is called *categorical in* a cardinality κ, or also κ-*categorical*, iff any two models of T of cardinality κ are ismorphic.

Using this defintion we can restate what has just been said about T_{rat}) as:

(i) T_{rat} is ω-categorical (where ω is the cardinality of the denumerable sets and structures)

(ii) For any non-denumerable cardinality κ, T_{rat} is not κ-categorical.

Another way to describe these two facts makes use of the notion of the *categoricity spectrum* of a (complete) theory T. By the *categoricity spectrum of* T, CS(T), we understand that function which maps an infinite cardinality κ to 1 iff any two models of T of cardinality κ are ismorphic, and otherwise maps κ to 0. In terms of categoricity spectra the characterisation of T_{rat} is as follows:

(i) $CS(T_{rat})(\omega) = 1$;
(ii) $CS(T_{rat})(\kappa) = 0$, if κ non-denumerable.

From what little has been said so far, we should be prepared for all sorts of categoricity spectra - functions CS(T) according to which the collection of infinite cardinalities κ such that $CS(T)(\kappa) = 1$can take a wide variety of different forms. But as a matter of fact this is not so. It was shown in the early sixties by Morley - arguably the first truly deep result in general model theory - that for categoricity spectra CS(T) there are altogether only four possibilities: :

i. $CS(T)(\kappa) = 1$ for all infinite cardinalities κ;
ii. $CS(T)(\omega) = 1$; $CS(T)(\kappa) = 0$ for κ non-denumerable;
iii. $CS(T)(\omega) = 0$; $CS(T)(\kappa) = 1$ for κ non-denumerable;
iv. $CS(T)(\kappa) = 0$ for all infinite cardinalities κ.

As indicated above, case (i) turns out to be very rare and arises only for essentially trivial structures. (An example is the theory T_{inf} of the language {} which says that there are infinitely many indviduals.) An example of case (ii) is, as we have seen, our theory T_{rat}, but there

aren't many other interesting examples in this category, involving structures that are familiar on independent grounds. Examples of case (iii) are also rare; one - very surprising - example is the first order theory of the arithmetic operations + and . on the real numbers (see Section 2.4.2).

The bulk of mathematically important structures gives rise to theories falling under (iv). Among these structures there are in particular all those which contain the arithmetical structure of the natural numbers (i.e. the natural numbers with the operations of + and .)as a definable substructure. (Trivially, this includes in particular to the arithmetical structure of the natural numbers itself. For that structure contains itself as an (improper) substructure, definable by means of identity maps.)

All these negative results are indications of the limits of first order logic as a toll for characterising non-trivial mathematical structure.

Morley's Theorem is usually stated in the following form[3]:

Theorem 2 (Morley).

> Suppose that T is a theory of some first order language L and that T is κ-categorical for some non-denumerably infinite cardinality κ. Then T is κ-categorical for all non-denumerably infinite cardinalities κ.

2. 1.2 Lattices, as Partial Orders and as Algebras.

We noted in 2.1 that lattices can be viewed in two different ways. On the one hand they can be described as partial orderings with certain special properties (any two elements a and b have a least element above them (the *supremum* of a and b) and a greatest element below them (the *infimum* of a and b). But they can also be described as algebraic structures, characterised by two binary operations $\cup$ and $\cap$, which

---

[3] We do not prove Morley's theorem in these Notes. The proof is of this theorem is hard (much harder than any proof presented in these Notes) and would detain us for far too long. Proofs can be found in several textbooks on model theory, for instance in Chang & Keisler, *Model Theory*. or Hodges *Model Theory*.

assign to any pair of elements a, b their supremum $a \cup b$ and their infimum
$a \cap b$.

We first present lattices as partial orders with the mentioned properties; that is, we formulate an axiomatic theory T_{lato} ('lato' stands for 'lattice order') in the language L_{lato} (the language whose only non-logical constant is the 2-place predicate $\leq$ and for which the canonical reference would be '{$\leq$}') whose models are all and only the partial ordering that are lattices. We then show how the operations $\cup$ and $\cap$ can be defined in this theory and form a new theory T'_{lato} in the language $\{\leq,\cup,\cap\}$ by adding the proposed definitions of $\cup$ and $\cap$ to the given axioms of T_{lato}. From the axioms of T'_{lato} (which, remember, include the definitions of $\cup$ and $\cap$ in terms of $\leq$) we derive a certain set of theorems which are phrased strictly in terms of $\cup$ and $\cap$ (and thus do not contain $\leq$). These theorems can serve in their turn as axioms of a theory T_{lata} in the language $L_{lata} = \{\cup,\cap\}$. In this theory it is now possible to define $\leq$ (either in terms of just $\cup$ or in terms of just $\cap$). And these definitions are the reverse of the definitions of $\cup$ and $\cap$ in terms of $\leq$ in that adding them to the axioms of T'_{lata} yields a theory T'_{lato}:

(1) $\quad T'_{lata} = T'_{lato}$

Equation (1) captures the ultimate equivalence of the two directions from which lattice structure can be approached.

After having obtained this result we proceed to the theories of boolean lattices and boolean algebras. These theories - T_{bl} and T_{ba} (for 'boolean lattices' and boolean algebras', respectively) - are obtained by adding further axioms to T_{lato} and T_{lata}. It is easy to show that T_{lato} and T_{lata} stand in the same relation of definitional equivalence as T_{lato} and T_{lata}.

As implied by what was said in the introductory remarks to this section, it is convenient to o axiomatise the theory of lattice-like partial orderings using as primitive relation not the strict ordering relation <

but rather the corresponding weak ordering relation $\leq$.[4] In other words we start with the language $L_{lato} = \{\leq\}$. Let T_{lato} be the theory axiomatised by the following sentences of this language.

Des. 3 (Axioms for T_{lato})

$Ax_{lato}.1$ $(\forall x)(\forall y)(x \leq y \ \& \ y \leq x \leftrightarrow x = y)$
$Ax_{lato}.2$ $(\forall x)(\forall y)\forall z)(x \leq y \ \& \ y \leq z \rightarrow x \leq z)$
$Ax_{lato}.3$ $(\forall x)(\forall y)((\exists z)(x \leq z \ \& \ y \leq z \ \& \ (\forall u)(x \leq u \ \& \ y \leq u \rightarrow z \leq u))$
$Ax_{lato}.4$ $(\forall x)(\forall y)((\exists w)(w \leq x \ \& \ w \leq y \ \& \ (\forall u)(u \leq x \ \& \ u \leq y \rightarrow u \leq w))$

Note that $Ax_{lato}.1$ says that $\leq$ is both reflexive and antisymmetric. Thus $Ax_{lato}.1$ and the transitivity axiom $Ax_{lato}.2$ together state that $\leq$ is a partial ordering. $Ax_{lato}.3$ and $Ax_{lato}.4$ assert the existence of suprema and infima.

Our first task is to show that the suprema and infima whose existence is asserted by $Ax_{lato}.3$ and $Ax_{lato}.4$ are unique. We will argue the case for suprema; the case of infima is analogous.

We argue informally. (Here as elsewhere the argument could be turned without a formal derivation without any real difficulties, but such formal derivations tend to be lengthy and cumbersome and to obscure the idea of the argument.) Let x and y be any elements. Suppose that z and z' have the properties stated in (2) and (3)

(2) $(x \leq z \ \& \ y \leq z) \ \& \ (\forall u)((x \leq u \ \& \ y \leq u) \rightarrow z \leq u)$
(3) $(x \leq z' \ \& \ y \leq z') \ \& \ (\forall u)((x \leq u \ \& \ y \leq u) \rightarrow z' \leq u)$

Then we have, instantiating u to z' in (2),

(4) $(x \leq z' \ \& \ y \leq z') \rightarrow z \leq z'$

Since the antecedent of (3) is a conjunct of (2), we get $z \leq z'$ by MP. In the same way we get $z' \leq z$ by instantiating u to z in (2). From $Ax_{lato}.1$ we then get $z = z'$.

---

[4] As noted in the opening remarks to this Chapter the choice between $<$ and $\leq$ is strictly one of convenience. If we choose $<$ as primitive, then we can define $\leq$ in terms of it via $x \leq y \equiv_{df} x < y \vee x = y$; and if we choose $\leq$, then $<$ can be defined via $x < y \equiv_{df} x \leq y \ \& \ x \neq y$.

Exercise: Derive the sentence

$$(\forall x)(\forall y)(\forall z)(\forall z')((x \leq z \,\&\, y \leq z \,\&\, (\forall u)(x \leq u \,\&\, y \leq u \rightarrow z \leq u)) \,\&\, x \leq z' \,\&\, y \leq z' \,\&\, (\forall u)(x \leq u \,\&\, y \leq u \rightarrow z' \leq u))) \rightarrow z = z')$$

from T_{lato}. (The easiest way to do this is to construct a Semantic Tableau. Constructing a derivation in some system of Natural Deduction is also quite doable. An axiomatic derivation is (here as in most other cases) much harder.)

Given that T_{lato} entails the existence and uniqueness of suprema and infima, we can define the operations $\cup$ and $\cap$ in T_{lato} in terms of $\leq$ as in Def($\cup$,{$\leq$}) and Def($\cap$,{$\leq$})below. These definitions correctly determine the interpretations of $\cup$ and $\cap$ in any model of T_{lato}. Also, they can be added to T_{lato} without undesirable 'side effects', i.e. without adding new theorems that can be expressed in the language L_{lato} of T_{lato}.[5]

Def($\cup$,{$\leq$}) $(\forall x)(\forall y)(\forall z)(x \cup y = z \leftrightarrow (x \leq z \,\&\, y \leq z \,\&\, (\forall u)(x \leq u \,\&\, y \leq u \rightarrow z \leq u)))$

Def($\cap$,{$\leq$}) $(\forall x)(\forall y)(\forall z)(x \cap y = z \leftrightarrow (z \leq x \,\&\, z \leq y \,\&\, (\forall u)(u \leq x \,\&\, u \leq y \rightarrow u \leq z)))$

Let, as already indicated in the introduction to this section, T'_{lato} be the theory in the language $L_{lat} = \{\leq, \cup, \cap\}$ that is obtained by adding the definitions Def($\cup$,{$\leq$}) and Def($\cap$,{$\leq$})as new axioms to the axiom set {$Ax_{lato}.1$-$Ax_{lato}.4$} of T_{lato}. It is not hard to show that the following sentences are all theorems of T'_{lato}:

---

[5] If existence and/or uniqueness could not be proved fromT_{lato}, then adding the definitions would also add the non-derivable statement or statements of T_{lato} which expressing existence and uniqueness, respectively to the theory. The reason is that the left hand sides of the biconditionals in the definitions Def($\cup$,{$\leq$}) and Def($\cap$,{$\leq$}) (e.g. $x \cup y = z$ for the first of these) entail existence and uniqueness of z simply because that is part of the general logical properties of function constants. The fact that T_{lato} entails the existence and uniqueness conditions associated with the right hand sides of the biconditionals guarantees that addition of the two definitions is what is called a *conservative* extension of T_{lato}, i. e. an extension which has exactly the same theorems as T_{lato} in its original language L_{lato}. For more on conservativity and other properties of formal definitions see Section 2.3.

$Th_{lata}.1$ $(\forall x)\ x \cup x = x$

$Th_{lata}.2$ $(\forall x)x \cap x = x$

$Th_{lata}.3$ $(\forall x)(\forall y)\ x \cup y = y \cup x$

$Th_{lata}.4$ $(\forall x)(\forall y)\ x \cap y = y \cap x$

$Th_{lata}.5$ $(\forall x)(\forall y)(\forall z)\ (x \cup y) \cup z = x \cup (y \cup z)$

$Th_{lata}.6$ $(\forall x)(\forall y)(\forall z)\ (x \cap y) \cap z = x \cap (y \cap z)$

$Th_{lata}.7$ $(\forall x)(\forall y)\ (x \cup y) \cap x = x$

$Th_{lata}.8$ $(\forall x)(\forall y)\ (x \cap y) \cup x = x$

Exercise: Show that these are theorems of T'_{lato}.

The theorems $Th_{lat}.1$ - $Th_{lat}.8$ can now be used in their turn as axioms of a theory T_{lata} of the language $L_{lata} = \{\cup, \cap\}$. In this new capacity we refer to them as $Ax_{lata}.1$ - $Ax_{lata}.8$. T_{lata} allows us to define $\leq$ in terms of the non-logical constants $\cup$ and $\cap$ of its language L_{lata}. In fact, as adumbrated earlier, we need only one of $\cup$ and $\cap$ in such a definition. Two such definitions, one in terms of $\cup$ and one in terms of $\cap$, are given below as $Def(\leq,\{\cup\})$ and $Def(\leq,\{\cap\})$.

$Def(\leq,\{\cup\})$ $(\forall x)(\forall y)(x \leq y \leftrightarrow x \cup y = y)$

$Def(\leq,\{\cap\})$ $(\forall x)(\forall y)(x \leq y \leftrightarrow x \cap y = x)$

Adding either $Def(\leq,\{\cup\})$ or $Def(\leq,\{\cap\})$ as a new axiom to the set $\{Ax_{lata}.1,.., Ax_{lata}.8\}$ of axioms of T_{lata} yields an extension in the language L_{lat} from which the our original axioms $Ax_{lato}.1$ - $Ax_{lato}.4$ can be derived in their turn. For the sake of definiteness let us assume that the definition that is added is $Def(\leq,\{\cup\})$ and that the resulting extension of T_{lata} is the theory T'_{lata} of the language L_{lat}. As we noted in the introduction, it turns out that this theory is the very same theory as the theory T'_{lata} which we obtained by approaching the characterisation of lattice structure from the perspective of partial orderings. That is, we have the equality (1).

(1) $T'_{lata} = T'_{lato}$.

Exercise: Show the equality (1) is true. This requires showing -in addition to what has already been asked of the reader in earlier exercises from this section:

(i) $T'_{lato} \models Def(\leq,\{\cup\})$;

(ii) $T'_{lata} \models Ax_{lato}.i$ for i = 1,..,4;

(iii) $T'_{lata} \models Def(\cup,\{\leq\})$ and $Def(\cap,\{\leq\})$.

2. 1.3 Lattices based on sets and Boolean Lattices

Prominent among the models of T_{lato} are *power set inclusion structures*. These are models of the form $<P(X), \subseteq>$, where $P(X)$ is the power set of some set X and $\subseteq$ is the set inclusion relation (restricted to $P(X)$). Similarly a prominent subclass of the models of T_{lata} is that consisting of models of the form $<P(X), \cup, \cap>$, where $\cup$ and $\cap$ are the operations of set-theoretic union and intersection, again restricted to $P(X)$. What we have seen in general terms in the last section - viz. that $\cup$ and $\cap$ are definable in terms of $\leq$ and that $\leq$ is conversely definable in terms of $\cup$ or $\cap$ - is reflected by the well-known fact that set-theoretic union and intersection are definable in terms of $\subseteq$ and conversely. In fact, for any given X we can combine the structures $<P(X), \subseteq>$ and $<P(X), \cup, \cap>$ into a single structure $<P(X), \subseteq, \cup, \cap>$), which is a model of the theory which we have denoted either as T'_{lato} or as T'_{lata}.

But models of this kind are speial not only in that they are based on set-theoretic relations and operations. They are also special in that they all verify some additional conditions, which are expressible in the languages of our theories but are not derivable from those theories.

Among these conditions are in particular the so-called *distribution laws* for $\cup$ and $\cap$. Formulations of these laws are given in BA9 and BA10.

DISTR.1 $(\forall x)(\forall y)(\forall z)\ (x \cup y) \cap z = (x \cap z) \cup (y \cap z)$

DISTR.2 $(\forall x)(\forall y)(\forall z)\ (x \cap y) \cup z = (x \cup z) \cap (y \cup z)$

It follows from the results in the last section that DISTR.1 and DISTR.2 can be expressed in the language $\{\leq\}$. (In fact, one way to obtain such

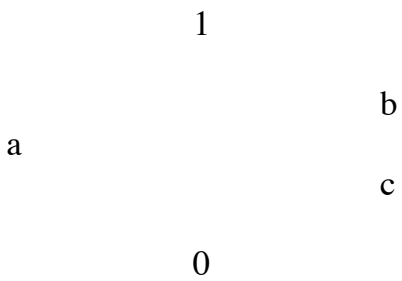
formulations is to translate DISTR.1 and DISTR.2 into formulas of L_{lat} using definitions Def($\cup$,{$\leq$}) and Def($\cap$,{$\leq$}) of $\cup$ and $\cap$ in terms of $\leq$.)

Exercise: Carry out this translation for DISTR.1 and prove that $Cl(T'_{lata} \cup \{DISTR.1\}) = Cl(T'_{lato} \cup \{DISTR'.1\})$, where DISTR'.1 is the translation of DISTR.1.

Lattices satisfying DISTR.1and DISTR.2 (or, what comes to the same thing, satisfying their translations into L_{lato}) are called *distributive* lattices. The following simple example shows that not all lattices are distributive. Let M be hte following model for the language L_{lato}:

> M = <{0,a,b,c,1}, $\leq$>, where $\leq$ is the following set of ordered pairs: {<0,0>,<0,a>,<0,b>,<0,c>,<0,1>,<a,a>,<a,1>, <c,c>,<c,b>, <c,1>,<b,b>,<b,1>,<1,1>}.

More perspicuously, M can be represented as the following directed graph[6]:



In this lattice we have: a $\cup$ c = a $\cup$ b = 1 and a $\cap$ c = a $\cap$ b = 0. So (a $\cap$ b) $\cup$ c = 0 $\cup$ c = c and (a $\cup$ c) $\cap$ (b $\cup$ c) = 1 $\cup$ b = b, falsifying DISTR.2.

Exercise: Show that the structure M described above also falsifies DISTR.1.

---

6 A directed graph G is a structure <U,R> where U is a set (the *nodes* of the graph G) and R is some binary relation on U. The pairs (a,b) of elements of U that belong to R are the *(directed) edges* of G. The edge (a,b) *goes from* a *to* b. Certain directed graphs, in which R is antisymmetric and either reflexive or irreflexive, can be used to represent partial orderings. When a graph G is used in this way, its node set represents the universe of the ordering,while the ordering relation itself is the transitive closure of R. Thus the ordering relation holds between two nodes a and b iff there is a path (i.e. a chain of edges) from a to b.

When a lattice is finite, it always has a smallest element - keep taking infima of pairs of elements - first taking the infimum c of two arbitrarily chosen elements a an b, then taking the infimum of c and some element d chosen arbitrarily from the elements not yet considered, and so on - until you have used up all elements of the lattice's finite universe - and a largest element (obtainable by taking suprema until the universe has been exhausted). Infinite lattices - i.e. infinite models of our theory T_{lato} - do not necessarily a smallest element (an element a such that for all other elements b in the lattice a $\leq$ b - or a largest element. (A counterexample is any unbounded linear order, such as, for instance, the orderings of the integers, the rationals or the reals.[7]) For the remainder of this section, however, we will focus on lattices which do have a smallest and a largest element.[8] We will refer to these as *the 0 of* the lattice and *the 1 of* the lattice, respectively. We will also use '0' and '1' as individual constants to denote these elements. We further limit our attention to distributive lattices. Thus - stated in terms of the language L_{lato} - we will be dealing with models of the theory $T_{d,0,1}$, whose axioms are, besides those of T_{lato}, translations into L_{lato} of the axioms DISTR.1 and DISTR.2 as well as the following two axioms, which assert the existence of a smallest and a largest element:

ExO $(\exists z)(\forall u)\ z \leq u$

Ex1 $(\exists z)(\forall u)\ u \leq z$

It is easy to see that $T_{d,0,1}$ entails that both the smallest and the largest element are uniquw. (This follows from ExO and Ex1, respectively, together with the fact that the models of T_{lato} are partial orderings.) This means that we can, for the same reason that this was possible earlier for $\cup$ and $\cap$, and following the same procedure, introduce individual constants 0 and 1 into the language L_{lato} by definitions obtained from the existence axioms ExO and Ex1. For the sake of explicitness the two definitions are given below.

Def(0,{$\leq$}) $(\forall z)(0 = z \leftrightarrow (\forall u)\ z \leq u)$

Def(1,{$\leq$}) $(\forall z)(1 = z \leftrightarrow (\forall u)\ u \leq z)$

---

7 Every linear order is a lattice. Exerice: prove that this is so.

8 The notion of a lattice is sometimes d*efined* as including the existence of a smallest and a largest element. This is not the practice we have adopted here.

Note that all set inclusion algebras are distributive lattices with a 0 and 1. On the other hand, as we already noted, linear orderings are distributive lattices, but they need not have a 0 or 1.

From here on it will be convenient to work in a language which contains all the constants considered so far - the 2-place predicate $\leq$, the two 2-place operations $\cup$ and $\cap$ and the individual constants 0 and 1. For the moment this is the language which contains just these five constants, i.e. $\{\leq,\cup,\cap,0,1\}$. Let $T'_{d,0,1}$ be the theory of this language whose axioms are:

(i) $Ax_{lato}.1$ - $Ax_{lato}.4$,
(ii) Def($\cup$,{$\leq$}) and Def($\cap$,{$\leq$}),
(iii) DISTR.1 and DISTR.2
(iv) Def(0,{$\leq$}) and Def(1,{$\leq$})

The theory $T'_{d,0,1}$ provides a suitable basis for the introduction of yet another operation, the 1-place operation of *complement*. To pave the way for the introduction of this operation we proceed once as we did before in the case of $\cup$, $\cap$, 0 and 1, viz. by first adopting a new axiom which asserts the existence of suitable values for the operation, then proving that these values are unique, and then, on the basis of this result introducing hte operation by means of a definition that is derived directly from the existence axiom.

Our existence axiom, COMP, asserts that for every element x there is an element y such that (a) the supremum of x and y is the 1 of the lattice and (b) the infimum of x and y is the 0 of the lattice:

COMP $(\forall x)(\exists y)(x \cup y = 1 \ \& \ x \cap y = 0)$

From the combination of $T'_{d,0,1}$ and COMP it is possible to prove that the element y mentioned in COMP is uniquely determined in relation to x. We argue as follows. First we observe that the sentences (i) and (ii) are theorems of $T'_{d,0,1}$. (The proof of this is left to the reader.)

(i) $(\forall u)\ u \cap 1 = u$

(ii) $(\forall u)(u \cup 0 = u$

Assume that y_1 and y_2 both satisfy the matrix (= the quantifier-free part) of COMP for some given x, i.e. that

(a) $x \cup y_1 = 1$ (b) $x \cup y_2 = 1$
(c) $x \cap y_1 = 0$ (d) $x \cap y_2 = 0$

Then, since $x \cup y_1 = 1$, $(x \cup y_1) \cap y_2 = 1 \cap y_2 = y_2$, by (i). By DISTR.1 $(x \cup y_1) \cap y_2 = (x \cap y_2) \cup (y_1 \cap y_2)$ and $(x \cap y_2) \cup (y_1 \cap y_2) = 0 \cup (y_1 \cap y_2) = y_1 \cap y_2$, by (ii) and assumption (d). So $y_1 \cap y_2 = y_2$. Similarly we show that $y_2 \cap y_1 = y_1$. So $y_1 = y_2 \cap y_1 = y_1 \cap y_2 = y_2$.

The definitions Def($\cup$,{$\leq$}) and Def($\cap$,{$\leq$}) enable us to translate the axioms DISTR.1, DISTR.2 and COMP into sentences DISTR.1($\leq$), DISTR.2($\leq$) and COMP($\leq$) of the language {$\leq$}. Consider the theory T_{bl} that we obtain when these translations to the theory T_{lato}. (The subscript 'bl' stands for 'boolean lattice'.) The models of T_{bl} are called *boolean lattices*. In view of the existence and uniqueness of complements in such models we can, in the same way in which we extended the theory of lattice orderings with definitions for the supremum and infimum functions and those for the '0-place functions' 1 and 0, now add a definition of the complement function. We denote this function as '-'. (That is, -x is the complement of x.)

The definition Def(-,{$\cup$,$\cap$}) of - can, as we already said, be directly obtained form the corresponding existence axiom COMP.

Def(-,{$\cup$,$\cap$}) $(\forall x)(\forall y)(y = -x \leftrightarrow (x \cup y = 1 \ \& \ x \cap y = 0))$

'-' and Def(-,{$\cup$,$\cap$}) are our final additions to language and theory, respectively. Let L_{bla} be the language {$\leq$,$\cup$,$\cap$,0,1,-} and let T_{bla} be the extension of $T'_{d,0,1}$ with COMP and Def(-,{$\cup$,$\cap$}). The models of T_{bla} are on the one hand, because of the properties of their partial ordering relation, boolean lattices, while on the other hand they have, because of the properties of their operations $\cup$,$\cap$,0,1 and -, the structure of *boolean algebras*.

To amplify this last statement: We have seen that the theory of lattices can be formulated in terms of the operations $\cup$ and $\cap$. (This was the theory T_{bla}.) We can extend this theory with existence axioms and

definitions for 0,1 and - all couched in terms of $\cup$ and $\cap$. It is not hard to show that the theory that we obtain this way, and which belongs to the language $\{\cup,\cap,0,1,-\}$ is identical with the restriction of T_{bla} to the sentences of this language. This theory is known as the theory of boolean algebras and its models as *boolean algebras*. So as to fit in with this terminology we refer to the language $\{\cup,\cap,0,1,-\}$ as L_{ba} and to the theory of this language which we have just described as T_{ba}.

For further reference we list once more the set of axioms for T_{ba} which has emerged in the course of this discussion. In this list we have combined the existence axioms which guarantee the legitimacy of the corresponding definitions we used to introduce the new operation symbols into single axioms, in which the operation symbols take the place of the existentially quantified variables in the existence axioms.

Def. 4 (Axioms for the theory T_{ba} of boolean algebras.[9])

$Ax_{ba}.1$ $(\forall x)\ x \cup x = x$

$Ax_{ba}.2$ $(\forall x) x \cap x = x$

$Ax_{ba}.3$ $(\forall x)(\forall y)\ x \cup y = y \cup x$

$Ax_{ba}.4$ $(\forall x)(\forall y)\ x \cap y = y \cap x$

$Ax_{ba}.5$ $(\forall x)(\forall y)(\forall z)\ (x \cup y) \cup z = x \cup (y \cup z)$

$Ax_{ba}.6$ $(\forall x)(\forall y)(\forall z)\ (x \cap y) \cap z = x \cap (y \cap z)$

$Ax_{ba}.7$ $(\forall x)(\forall y)\ (x \cup y) \cap x = x$

$Ax_{ba}.8$ $(\forall x)(\forall y)\ (x \cap y) \cup x = x$

$Ax_{ba}.9$ $(\forall x)(\forall y)(\forall z)\ (x \cup y) \cap z = (x \cap z) \cup (y \cap z)$

$Ax_{ba}.10$ $(\forall x)(\forall y)(\forall z)\ (x \cap y) \cup z = (x \cup z) \cap (y \cup z)$

$Ax_{ba}.11$ $(\forall u)\ u \cap 1 = u$

$Ax_{ba}.12$ $(\forall u)(u \cup 0 = u$

$Ax_{ba}.13$ $(\forall x)(x \cup -x = 1 \ \&\ x \cap -x = 0))$

---

[9] Here $Ax_{ba}.1$ - $Ax_{ba}.8$ are the theorems $Th_{lat}.1$ - $Th_{lat}.8$ of Section 2.1.2; $Ax_{ba}.9$ and $Ax_{ba}.10$ are the earlier DISTR.1 and DISTR.2; $Ax_{ba}.11$ and $Ax_{ba}.12$ - mentioned earlier as (i) and (ii) in the proof of uniqueness of complements, are the results of combining the existence axioms Ex0 and Ex1 for the lattice One and the latice Zero with the definitions of the individual constants 0 and 1 in terms of $\cup$ and $\cap$ - these definitions we did not actually give, but they can be obtained from the defintions Def(0,{$\leq$}) and Def(1,{$\leq$}) we did give by translating them into sentences of L_{ba} using the definition of $\leq$ in terms of $\cup$; $Ax_{ba}.13$ results form combining the axiom COMP with Definition Def(-,{$\cup$,$\cap$}).

This concludes our general account of lattices, lattice algebras, boolean lattices and boolean algebras. The route we have followed, with all the switching back and forth between partial orderings and operations, may appear rather round-about and hard to follow, certainly on a first reading. But I believe that this is a price worth paying. The central methodological point of the last two sections has been to show, by means of the example that lattices and the corresponding algebras provide, how two at first sight very different perspectives on structure - here that of structure in the form of partial order and structure in the form of a number of connected operations - can nevertheless prove to be concerned with what is essentially the same structure after all. In order to bring out how and why this convergence arises in the case in question, switching between the two perspectives was essential. That does of course require a greater effort, both on the part of the presenter and that of the reader, than a simple presentation of lattices *just* as ordered structures or of lattice algebras and boolean algebras *just* in terms of their operations.

There is also a practical spin-off to the presentation of lattices as being describable either as partial orders or as algebras: Now that we have explored the nature of this correspondence thoroughly, we can, with the benefits of that investigation, join the wide-spread practice of switching betgween the two perspectives in discussions of such structures if and when this proves convenient. We will make use of this freedom in particualr in the next sections.

In the next two sections we focus exclusively on boolean algebras. 2.1.4 presents a number of distinct types of boolean algebras and defines certain properties in terms of which they can be distinguished from each other and classified. 2.1.5 is devoted to the Stone Cech Theorem, according to which every boolean algebra is isomorphic to a structure in which the operations $\cup$, $\cap$ and - are set-theoretic union, intersection and subtraction, respectively.

2.1.4 Some Examples of Boolean Algebras.

As compared with lattices in general, boolean lattices form a quite special category. But even so there is much variety even within this special domain. One important subtype is that identified by the power set inclusion lattices $<P(X),\subseteq>$ that were mentioned earlier. These are distinguished by two properties: they are (i) *atomic* and (ii) *complete*.

Before we define these two properties, first, in Prop. 2, an obvious observation about the power set lattices, viz that they are determined up to isomorphism by their carrier sets X:

Prop. 2 If $|X| = |X'|$, then $<P(X),\subseteq> \cong <P(X'),\subseteq>$.

Proof: It suffices to note that a bijection between X and X' induces a bijection between $P(X)$ and $P(X')$ and carries the inclusion relation restricted to $P(X)$ into the inclusion relation restricted to $P(X')$.

Next the definitions of atomicity and completeness. The first of these presupposes the notion of an element being an *atom*, which is important in its own right.

Def. 5 (a) Let $BL = <U,\leq>$ be a boolean lattice, b an element of BL. b is an *atom* of BA iff

(i) $a \neq 0$ and

(ii) there is no c in BL such that $0 < c < a$ (where $<$ is the strict partial order corresponding to the lattice ordering $\leq$.

(b) BL is *atomic* iff for every b in BL there is an atom a of BA such that $a \leq b$.

Def. 6 A boolean lattice $BL = <U,\leq>$ is *complete* iff for every subset V of U there is a least element c in U such that for all $v \in V$, $v \leq c$. More formally:

For each $V \subseteq U$ there is a c in U such that

(i) $(\forall v \in V)\ v \leq c$, and

(ii) $(\forall c')((\forall v \in V \rightarrow v \leq c') \rightarrow c \leq c')$

To show that power set inclusion lattices are atomic and complete is once again very easy to show and we record the fact as another proposition.

Prop. 3 Every power set inclusion lattice is atomic and complete.

Proof: Let PSIL = $<P(X),\subseteq>$ be any 0power set inclusion lattice. Note that the 0 of PSIL is the empty set $\varnothing$. So the atoms of PSIL are the singleton sets {x}, where x is any element of X. Suppose that Y is any element of PSIL distinct from 0. Then Y is a subset of X and $Y \neq \varnothing$. So Y contains at least one element $x \in X$. But then we have $\{x\} \subseteq Y$, i.e. the lattice ordering relation holds between the atom {x} and Y. Thus PSIL is atomic.

To see that PSIL is complete, let V be any subset of $P(X)$. Then $\cup V$ is a subset of X and thus a member of $P(X)$. It is easy to verify (i) that for all $V \in V$, $V \subseteq \cup V$ and (ii) if W is any other element of $P(X)$ such that for all $V \in V$, $V \subseteq W$, then $\cup V \subseteq W$. Thus PSIL is complete.

But not all boolean lattices are either atomic or complete, In fact, there are boolean lattices that are the extreme opposite of atomic in that they have no atoms at all. And there are also boolean lattices that are the extreme opposite of complete in that they have the following property:

> Every set V of elements is either *essentially finite* or else V does not have a supremum.

Here by *essentially finite* we mean the following: V is *essentially finite* iff there is a finite subset W of V such that $(\forall v \in V)(\exists w \in W)\ v \leq w$. (Note that in this case the supremum of W, which must exist since W is finite, is also the supremum of V.)

But besides boolean lattices which occupy the opposite end of the spectrum from the power set inclusion lattices with regard to either atomicity or completeness or both, there are also many which display less extreme forms of non-atomicity or incompleteness. For instance there are boolean lattices which do contain some atoms but which nevertheless do not have enough of them to make them atomic.

Our first example of a boolean lattice that is not like the power set inclusion lattices, BL1, differs from them in being not complete,

although it shares with them the property of being atomic. The example also illustrates another important fact, the true significance of which will become clear when we turn to the Stone Cech Representation Theorem in the next section. This is because it is a boolean lattice whose ordering relation is, just like it is for the power set inclusion lattices, set-theoretic inclusion. The only, but crucial difference with the power set inclusion lattices is that in our example the universe is no longer a full power set $P(X)$, but rather some proper subset of such a power set. (The Stone Cech Theorem says that just by varying the universes of inclusion lattices all possible properties of boolean lattices can be exemplified.)

In the case of BL1 the universe is defined as the set of all fintie and all cofinite subsets of the set N of natural numbers. Here a cofinite subset of N is a subset Y of N such that N \ Y is finite. In other words, if U is the set of all finite and cofinite subsets of N, then BL1= $\langle U,\subseteq\rangle$, where is the relation of set-theoretic inclusion restricted to U.

Before we show that BA1 has the mentioned properties, i.e. that it is atomic but not complete, we first have to show that it is a boolean lattice- in other words, that it is a lattice and that is it boolean. To this end we make use of the possibility of switching back and forth between lattices and the corresponding algebras. To start, note that the restriction of $\subseteq$ to any set of sets will always be a partial order. To show that in the case at hand this order is a lattice we note that U is closed under the set-theoretic operations $\cup, \cap$. To see that the union X $\cup$ Y of two subsets X and Y of U belongs to U, we have to distinguish between two cases: (i) if X, Y are both finite, then X $\cup$ Y is finite and thus in U; (ii) if at least one of X, Y is cofinite, then X $\cup$ Y is cofinite and thus also in U. In the same way one shows that U is closed under $\cap$. From the fact that U is closed under $\cup$ and $\cap$ it follows that $\langle U,\subseteq\rangle$ is a lattice. For if, say, the union of the sets X and Y from U is again a member of U, then it will be the supremum of X and Y in $\langle U,\subseteq\rangle$; likewise, since the intersection of subsets X and Y of U again belongs to U it must be the infimum of X and Y. Thus $\langle U,\subseteq\rangle$ is a lattice.

Furthermore, $\varnothing$ and N both belong to $\langle U,\subseteq\rangle$, since is a finite and N a cofinite subset of N. But then it is obvious that these are the smallest and largest element, respectively, of $\langle U,\subseteq\rangle$. So $\langle U,\subseteq\rangle$ has a 0 and a 1. We also note that set-theoretic union and intersection satisfy the distributivity laws DISRT1 and DISTR2. So $\langle U,\subseteq\rangle$ is a distributive lattice

with a 0 and a 1. To see that <U,⊆> is boolean, we note that U is closed under the operation of complementation relative to N (that is, the operation of subtracting a given X from N, denoted as N \ X). For the relative complement of a finite subset of N is a cofinite subset and vice versa. Using the same reasoning as above, we conclude that the relative complement is the operation we obtain when we apply Def(-,{∪,∩}) (see section 1.2.3) to the supremum and infimum operations of <U,⊆>, which, as we have already shown, are nothing but the operations of set-theoretic union and intersection. Moreover the relative complement operation of set theory does satisfy, in conjunction with union and intersection, the laws $Ax_{ba}.11$ and $Ax_{ba}.12$. So <U,⊆> is a boolean lattice.

We next show that BA1 is atomic. This is easy. All singleton sets {n}, where the n ε N, are finite and thus belong to U. Clearly they are the atoms of <U,⊆>. And if X is a member of U that is different from the 0 of U, i.e. X ≠ ∅, then there must be some n such that n ε X and thus {n} ⊆ X; so there is an atom between 0 and X.

Finally BA1 is not complete. For let A be a subset of N such that both A and N\A are infinite. (For instance, we could take for A the set of even numbers.) Let *A* be the set {{n}: n ε A}. Then *A* has no supremum in <U,⊆>. For if Y is any element in U with the property that (∀Z)(Z e *A* → Z ⊆ Y), then A ⊆ Y. The only elements of U with this property are the cofinite subsets of N which include A. But among these there is no smallest element: Take any such Y. Then Y\A is non-empty (in fact it is infinite). Let m ε Y\A and Y' = Y\{m}. Then Y' ε U, A ⊆ Y' and Y' is a proper subset of Y. So there is no smallest member of U which includes all members of *A*.

Our second example, BA2, is presented as a boolean algebra. And it is not a set algebra. Once again the set N of natural numbers is our starting point. But this time we begin by defining an equivalence relation on the subset of N:

$$X \equiv Y \text{ iff}_{\text{def.}} X - Y \text{ is finite.}$$

Here "X - Y" denotes the *symmetric difference* between X and Y, i.e. X - Y = (X\Y) ∪ (Y\X).

The first thing to observe is that $\equiv$ is a *congruence relation* with respect to the set-theoretic operations $\cup$, $\cap$ and \. That is, if the arguments of the operations stand to each other in the relation $\equiv$, then so are the results of those operations. For instance, suppose that $X \equiv X'$ and that $Y \equiv Y'$. Then also $(X \cup Y) \equiv (X' \cup Y')$. That this must be so is not hard to see. On the one hand $(X \cup Y) \setminus (X' \cup Y') \subseteq (X \setminus X') \cup (Y \setminus Y')$. This entails that if the term on the right of $\subseteq$ is finite, so is the one on the left. Analogously $(X' \cup Y') \setminus (X \cup Y)$ is finite. So $(X \cup Y) \setminus (X' \cup Y')$ is finite. It follows that $(X \cup Y) \equiv (X' \cup Y')$. Likewise for the other two operations.

Let V be the set $\{[X]_{\equiv} : X \subseteq N\}$. (N.B. during the remainder of our discussion of BA2 we will leave out the subscript $\equiv$.) The congruence of $\equiv$ w.r.t. $\cup$, $\cap$ and \ entails that we can define the following operations on V:

Def. For arbitrary $X, Y \subseteq N$,

(i) $[X] \underline{\cup} [Y] = [X \cup Y]$
(ii) $[X] \underline{\cap} [Y] = [X \cap Y]$
(iii) $\underline{\neg} [X] = [N \setminus X]$

Now let BA2 be the structure $\langle V, \underline{\cup}, \underline{\cap}, \underline{\neg}, [\varnothing], [N] \rangle$. That this is a Boolean algebra follows straightforwardly from the Boolean nature of the set-theoretical operations $\cup$, $\cap$ and \, in terms of which we have defined the operations $\underline{\cup}$, $\underline{\cap}$, $\underline{\neg}$. Note that the lattice ordering $\underline{\leq}$ of this structure holds between any two members [X] and [Y] of V iff $X \setminus Y$ is finite. To see this, recall that $\underline{\leq}$ can be defined in terms of $\underline{\cup}$ by: $[X] \underline{\leq} [Y]$ iff $[X] \underline{\cup} [Y] = [Y]$. This entails that $(X \cup Y) \setminus Y$ is finite. But $(X \cup Y) \setminus Y$ is the same set as $X \setminus Y$.

We first observe that BA2 is atomless, and thus not atomic. Suppose that $[X] \neq [\varnothing]$. Then X is infinite. But then we can split X into two infinite subsets Y and $X \setminus Y$ But in that case we have $\varnothing < [Y] < [X]$, where $<$ is the strict order in relation corresponding to the lattice ordering $\leq$ of BA2. So [X] is not an atom.

BA2 is also not complete. To see this, let A be a denumerably infinite set of infinite mutually disjoint subsets of N whose union is N. (That is, if $X \varepsilon A$, then X is infinite, if $X, Y \varepsilon A$ and $X \neq Y$, then $X \cap Y = \varnothing$ and $\cup A$ = N.) Then there is no element in V which is the supremum of A. For suppose that [Z] were the supremum of A. Then for each $X \varepsilon A$, $[X] \leqq [Z]$. So, by the remark at the end of the penultimate paragraph $X \setminus Z$ is finite. Consequently, since X infinite and $X = (X \setminus Z) \cup (X \cap Z)$, $Z \cap X$ must be infinite and thus $\neq \varnothing$. So we can for each X in A pick an element n_X from $X \cap Z$. Note that if $Y \varepsilon A$ and $Y \neq X$, then by assumption Y is disjoint from X and therefore n_X is not an element of Y. So each n_X belongs to exactly one element of A. That is, if $B = \{n_X : X \varepsilon A\}$, then for each $X \varepsilon A$, $X \cap B = \{n_X\}$. Now let $Z' = Z \setminus B$. Since $B \subseteq Z$, $Z \setminus Z' = B$ and thus $Z \setminus Z'$ is infinite. So $[Z'] \leq [Z]$. On the other hand, for any $X \varepsilon A$, $Z' \setminus X = (Z \setminus X) \cup \{n_X\}$, and this set is finite, since $Z \setminus X$ is finite. So, by the remark at the end of the one-but-last paragraph, $[X] \leqq [Z']$. It follows that [Z] is not the supremum of A.

There are also boolean lattices that are complete but not atomic. [An example of such a lattice can be found in the exercises.]

2.1.5 The Stone-Cech representation Theorem

One of the most famous and most fundamental results in the theory of boolean algebras is the *Stone-Cech Representation Theorem,* which says that every boolean algebra is isomorphic to (and thus 'can be represented as') a set algebra; that is, it is isomorphic to a structure $\langle U, \cup, \cap, -, 0, 1\rangle$ in which the elements of U are subsets of some set X, the operations $\cup, \cap, -$, are set-theoretic union, intersection and complementation relative to X, 0 is the empty set and 1 the set X. (Once more, note well that U will in general not consist of all subsets of X.)

The proof of the Stone Cech Theorem involves the notion of an *ideal* of a boolean lattice, or, alternatively, that of a *filter*. So we begin by defining these notions as well as a few others connected with them.

Def. 7. Let BL = <U,≤> be a boolean lattice.

1. A subset V of U is an *ideal of* BL iff (i) V ≠ ∅; (ii) V ≠ U; (iii) if b ε V and a ≤ b, then a ε V; and (iv) if a, b ε V, then a ∪ b ε V.

2. A subset V of U is a *filter of* BL iff (i) V ≠ U; (ii) V ≠ ∅; (iii) if b ε V and b ≤ a, then a ε V; and (iii) if a, b ε V, then a ∩ b ε V.

3. Let b ε BL, b ≠ 1. The *ideal* of BL *generated by* b is the set {a ε U: a ≤ b}
An ideal is called a *principal* ideal if it is generated by some b ε U such that b ≠ 1.

   Likewise, if b ≠ 0, the *filter* of BL *generated by* b is the set {a ε U: b ≤ a}; and a filter is called a *principal* filter if it is generated by some b ε U such that b ≠ 0.

4. An ideal V of BL is called a *prime* or *maximal* ideal of BL iff for each b ε U either b ε V or else -b ε V.

   Likewise, a filter V of BL is called a *prime* or *maximal* filter of BA iff for each b ε U either b ε V or else -b ε V.

Prop. 4. 1. If V is an ideal of a boolean lattice BL = <U,≤>, then -V = {-b: b ε V} is a filter of BL, and conversely.

2. If V is a principal ideal {a ε U: a ≤ b} of BL, then -V is the principal filter {a ε U: -b ≤ a} of BL, and conversely.

3. If V is a prime ideal of BL, then -V is a prime filter of BL, and conversely.

Lemma. 1. (Boolean Prime Ideal Theorem for Boolean Lattices)

Let V be an ideal of some BL <U,≤>. Then there exists a prime ideal V' of BL such that V ⊆ V'.

A general proof of the Prime Ideal Theorem, which applies to lattices of arbitrary cardinality, is not possible at this stage, since it requires set-theoretic assumptions and methods that are not available to us as yet. We can only prove the theorem for BA's which are at most infinitely denumerable. For this case the argument goes as follows.

If $<U,\leq>$ is denumerable, then we can assume an enumeration $u_1, u_2, \ldots$ of all elements of U and extend V stepwise, first with u_1 or $-u_1$, then with u_2 or $-u_2$, and so on, obtaining in the limit an extension of V which is a prime ideal. We just sketch the first step, in which V is extended with either u_1 or $-u_1$. (The other steps are completely analogous.)

With regard to V and u_1 we distinguish two cases:

(a) For all finite $W \subseteq V$, $\sup(W) \cup u_1 \neq 1$.
(b) For some finite $W \subseteq V$, $\sup(W) \cup u_1 = 1$.

In case (a) $V_1 = \{u \in U: (\exists W)(W \subseteq V \ \& \ W \text{ finite} \ \& \ u \leq \sup(W) \cup \{u_1\}\}$;
in case (b) $V_1 = \{u \in U: (\exists W)(W \subseteq V \ \& \ W \text{ finite} \ \& \ u \leq \sup(W) \cup \{-u_1\}\}$

We begin by showing that in case (a) V_1 is an ideal. First, note that $u_1 \in V_1$. This is so since the empty set $\varnothing$ is a subset of V and $u_1 \leq 0 \cup u_1 = \sup(\varnothing) \cup u_1$. Second, suppose $b \in V_1$ and $a \leq b$. Then there is a finite $W \subseteq V$ such that $b \leq \sup(W) \cup u_1$. But then also $a \leq \sup(W) \cup u_1$, so $a \in V_1$. Third, suppose that $V_1 = U$. Then $1 \in V_1$. This means that there is a finite $W \subseteq V$ such that $1 \leq \sup(W) \cup u_1$, which is equivalent to: $\sup(W) \cup u_1 = 1$. This contradicts the assumption of case (a) and we conclude that $V_1 \neq U$. Lastly, let $a, b \in V_1$. Then there are finite subsets W_a, W_b of V such that $a \leq \sup(W_a) \cup u_1$ and $b \leq \sup(W_b) \cup u_1$. If W_a and W_b are both finite subsets of V, then so is $W_a \cup W_b$. Also, $\sup(W_a) \leq \sup(W_a \cup W_b)$, so $a \leq \sup(W_a \cup W_b) \cup u_1$. Similarly, $b \leq \sup(W_a \cup W_b) \cup u_1$. So $a \cup b \leq \sup(W_a \cup W_b) \cup u_1$. Our final observation is that $V \subseteq V_1$. Suppose that $u \in V$. Then $u \leq \sup(\{u\}) \cup u_1$, with $\{u\}$ a finite subset of V. So $u \in V_1$.

From all this we conclude for case (a): V_1 is an ideal which extends V and contains one of u_1 and $-u_1$.

Now consider case (b). We show:

(*) for all finite $W' \subseteq V$, $\sup(W') \cup -u_1 \neq 1$.

Suppose this is not so. Then there is a finite $W' \subseteq V$ such that $\sup(W') \cup -u_1 = 1$. By assumption of case (b) there also is a finite $W \subseteq V$ such that $\sup(W) \cup u_1 = 1$. Let $W'' = W \cup W'$. Then W'' is a finite subset of V. Let $w = \sup(W'')$. Then $w \in V$, so, since V is an ideal, $w \neq 1$ (for otherwise we would have that $V = U$). Furthermore, $\sup(W) \leq w$. So, since $\sup(W) \cup u_1 = 1$, $w \cup u_1 = 1$. Similarly $w \cup -u_1 = 1$. So $(w \cup u_1) \cap (w \cup -u_1) = 1 \cap 1 = 1$. But $(w \cup u_1) \cap (w \cup -u_1) = ((w \cup u_1) \cap w) \cup ((w \cup u_1) \cap -u_1) = w \cup ((w \cap -u_1) \cup (u_1 \cap -u_1)) = w \cup ((w \cap -u_1) \cup 0) = w \cup (w \cap -u_1) = w$. So $w = 1$, contrary to what we established above. This proves (*).

We can now show as in case (a) that V_1 is an ideal which extends V and contains $-u_1$. So it follows in either case that V_1 is an ideal which extends V and contains one of u_1 and $-u_1$.

In this way we coinstruct a denumerable sequence $V_1, V_2, \ldots$ of ideals extending V such that for each n V_n will contain, for $i = 1, \ldots, n$, one of u_i and $-u_i$.

Now let $V' = \bigcup_n V_n$. Then it is easy to show that V' is an ideal. (In particular V' does not cootain 1. For is it did than 1 would be an element of some V_n, contradicting the already established fact that V_n is an ideal. From the construction of V' it is also clear that V' is maximal.

q.e.d.

Corollary. Let u be an element of some BL $\langle U, \leq \rangle$ such that $u \neq 1$. Then there exists a prime ideal V' of BL such that $u \in V'$.

Proof. Suppose that u is as described in the statement. Then $V_u = \{v \in U: v \leq u\}$ is an ideal. (Show this. N. B. ideals of this form, which consist of all elements $\leq$ some given element, are called *principal ideals*.) So, accroding to Lemma 3 there is a prime ideal V such that $V_u \subseteq V$. Clearly u e V.

q.e.d.

We now turn to the Stone-Cech Theorem itself.

Theorem. 3 (Stone-Cech Theorem for Boolean Lattices)

Let $M = \langle U, \leq \rangle$ be any boolean lattice. Then there is a set inclusion lattice M* - i.e. a structure $\langle U^*, \subseteq \rangle$ in which U* is a subset of some power set $P(X)$ and $\subseteq$ is the set-theoretic inclusion relation on U* - such that $M \cong M^*$.

Proof. For any $u \in U$ let u* be the set consisting of all maximal ideals V of M such that not $u \in V$: $u^* = \{V: V \text{ is a prime ideal of M and not } u \in V\}$. Let $U^* = \{u^*: u \in U\}$. Then $U^* \subseteq P(P(U))$; so $U^* \subseteq P(X)$ for some X. We show that * is 1-1 map from U onto U*. That * is onto follows from the definition of U*. To show that * is 1-1 we argue as follows. First suppose that $u, u' \in U$ and that $u \neq u'$. Then either not $u \leq u'$ or not $u' \leq u$. Assume that not $u \leq u'$. (The other case is analogous.) Then $u' \cup -u \neq 1$. For if $u' \cup -u = 1$, then $u \cap u' = u' \cap u = (u' \cap u) \cup 0 = (u' \cap u) \cup (-u \cap u) = (u' \cup -u) \cap u = 1 \cap u = u$, so $u \leq u'$, contrary to assumption. Since $u' \cup -u \neq 1$, there is according to the Corollary to Lemma 3 a maximal ideal V containing $u' \cup -u$. Since $-u \in V$ it is not the case that $u \in V'$, For otherwise $u, -u \in V'$, so $u \cup -u \in V'$ and V' wouldn't be an ideal. So by teh definition of *, $V' \in u^*$. On the other hand, $u' \in V'$. So it is not the case that $V' \in u'^*$. So $u^* \neq u'^*$.

Next we prove that $u \leq u'$ iff $u^* \subseteq u'^*$. First assume $u \leq u'$. Then, as can easily be shown, $-u' \leq -u$. Let V be any maximal ideal in u*. Then, since not $u \in V$, $-u \in V$. So, since $-u' \leq -u$, $-u' \in V$. So it is not the case that $u \in V$, and therefore $V \in u'^*$. So $u^* \subseteq u'^*$. Conversely assume that $u^* \subseteq u'^*$. Suppose it is not the case that $u \leq u'$. Then, as we saw above, there is a maximal ideal V' such that $u' \in V'$ but not $u \in V'$. So $V' \in u^*$, but not $V' \in u'^*$, contrary to the assumption that $u^* \subseteq u'^*$.

q.e.d.

The Stone-Cech Representation Theorem for Boolean Algebras is a paradigm for a type of result that has proved of great value in mathematics and logic in a number of distinct contexts. Results of this type are generally called 'representation theorems'. Informally speaking, a *representation theorem for* a theory T of a language L is a statement to the effect that a certain class *M'* of models for some

extension L' of L is representative for T's models. Putting the matter more formally, representation theorems take the following general form:

> Let T be a theory in some language L. Let M be the class of all models of T. Let L' be an extension of L and let M' be a class of models for L' such that for each M ε M' the reduction M$\lceil$L of M to L is a member of M. Then M' *is representative of the models of* T iff for each M ε M there is an M' ε M' such that M $\cong$ M'$\lceil$L.

The use and importance of representation theorems is in most cases that they provide a clearer view of the range of variation among the models of a given theory T and/or a way of studying this variation. In order to obtain a picture of the different (isomorphism) types of models of T it is enough to study the variation within the representing class M'. And in many cases this latter investigation is helped by the fact that the models within this class are of a special kind, e.g. in that they have additional properties which do not apply to models of T in general. (Normally this is because these properties are not expressible within the language L of T, but only in the extended language L' of the models in M'.

The Stone Representation Theorem for boolean lattices is a good example of this: There are ways to explore the possible structure of set inclusion lattices which are not directly available for arbitrary boolean lattices. On the other hand, however, the very fact that the Stone Theorem is true is an indication of how much variation can be found among set inclusion lattices. To take just one example, our algebra BAII was not a set inclusion lattice as we defined it. Stone's Theorem tells us that there is a set inclusion lattice isomorphic to BAII, and also gives us a method for how to construct such a lattice. But the resulting lattice is not a set inclusion structure that one would easily have thought of off the bat. Should one have expected that a set inclusion lattice with this structure actually exists? That would of course depend on our general knowledge of set theory, but at the very least the answer is not obviously 'yes'.

In fact, one way to look at the Stone-Cech Theorem is as a statement telling us how much variation can be obtained by starting from the narrowly circumscribed notion of a power set lattice $<P(X), \subseteq>$ - recall: any such lattice is atomic and complete and it is fully determined by the cardinality of the carrier set X - and then to broaden this notion by allowing for variation in just one respect: the universe U

need not be all of $P(X)$, but may also be some proper subset of it. All variety, in other words, can be located in the choice of U.

2.1.6 Booelan Algebra and Logic.

We noted at the outset of this Chapter that boolean algebras are of particular importance for logic; some the most prominent structures that are studied in formal logic have the properties of such algebras.[10] The simplest (and arguably most central) example is the 'algebra of propositions', in which the disjunction p v q of two propositions p and q interpreted as the supremum of p and q, their conjunction p & q as their infimum and the negation -p of p as its complement. Exactly what boolean algebra this will give us depends on how we decide to characterise propositions. When we identify 'propositions' with the Fregean denotations of sentences - 'the True', or '1', and 'the False', or '0' - then we get a boolean algebra whose universe is the two-element truth value space {0,1}, in which the boolean operations are as follows:

(i) $1 \cup 1 = 0 \cup 1 = 1 \cup 0 = 1, 0 \cup 0 = 0;$
(ii) $1 \cap 0 = 0 \cap 1 = 0 \cap 0 = 0, 1 \cap 1 = 1;$
(iii) $-1 = 0, - 0 = 1.$

Note that this algebra results as the image of any language L of propositional logic under any classical valuation Suppose that V is a classical valuation of the set of propositional letters of L (classical in the sense that it assigns each letter one of the classical truth values 0 and 1). Then V will map each formula of L into {0,1} according to the familiar truth table rules:

---

10 Boolean algebras and lattices owe their name to one of the founders of modern logic, the 19-th century mathematician George Boole (1815-1864). Boole was together with his compatriot Augustus de Morgan, the first to look at logic from an algebraic perspective, according ot which the logical connectives &, v, ¬, etc. are seen as operators, or functors, which can be used to obtain propositions out of other propositions (e.g. the conjunction 'A & B' from the propoisitions A and B). Boole tried to formulate the laws of logic (his 'Laws of Thought') in algebraic terms, i.e. as equations that express logical equivalences that hold between propositions in virtue of their logical structure, such as e.g.

(i) A & B = B & A
(ii) (A & B) & C = A & (B & C)

to express the commutativity and associativity of conhunction. Eventually such equations became the axiomatic foundation of the definition of the concept of a boolean algrebra, see our axioms Ax_{ba}.1 - 13 on p. 21 of this Chapter.

(i) $V(A \vee B) = V(A) \cup V(B)$;

(ii) $V(A \& B) = V(A) \cap V(B)$;

(iii) $V(\neg A) = 1$ if $V(A) = 0$ and $V(\neg A) = 0$ if $V(A) = 1$.

More interesting is the kind of algebra that we get when propositions are characterised *intensionally*, viz. as sets of possible worlds. Let W be the set of all possible worlds. Then each proposition p determines a subset of W, consisting of those worlds in which p is true. According to the *intensional* theory of propositions this set - or, if one prefers, the division of W into two parts that comes with it, the part of those possible worlds in which p is true and those in which it is false - fully identifies the proposition p; in other words, propositions *are* sets of possible worlds; and on the assumption that the set of all possible worlds is W, they are subsets of W. The logical operations of disjunction, conjunction and negation now turn into set-theoretic operations. For instance, the conjunction p & q of the propositions (i.e. subsets of W) p and q is the set of worlds of W in which both p and q are true, i.e. the worlds which belong both to the subset p of W and to the subset q of W. Thus p & q is the set-theoretic intersection of p and q. Similarly, p v q becomes the union of p and q and $\neg$p the set-theoretic difference W\p. Furthermore, the 0 of the proposition algebra thus defined is the empty set $\varnothing$ ('the contradictory proposition') and its 1 the entire set W ('the tautologous proposition').

This 'intensional' proposition algebra is the model-theoretic fundament of the currently most popular developments of modal and intensional logic, in which logical relations are defined in terms of a 'Kripkean'[11] model-theoretic semantics, propositions are interpreted as sets of possible worlds and modalities are analysed in terms of relations between such worlds. It is also the model-theoretic foundation of the system of Higher Order Intensional Logic, the logical formalism that was introduced by Montague[12] in his seminal work on the semantics of natural languages - work that, in various guises has served as the formal basis for the formal semantics of natural languages since the early seventies.

---

[11] Saul Kripke (1940 -) is the founder of modern modal logic. He did his astoundingwork in this area at the astoishingly young age of 16, while still in high school.

[12] Richard Montague (1930 - 1971). Founder of the model-theoretic approach to the analysis of meaning of natural languages. Montague was the first to see that it was possible and illuminating to apply the model-theoretic methods developed by his teacher Tarski for the formal languages of mathematical logic, such as, in particular, the predicate calculus..

There is a further connection between the semi-formal ideas expressed above and Montague's conception of the semantics of (formal and natural) languages. Montague thought of the way in which the syntactic structure of a sentence determines its meaning as generally taking the form of a *homomorpism* from syntactic structures to meanings (or 'semantic values'). In the context of the present discussion of boolean structure this idea can be explained rather succinctly. Doing so, moreover, will give an opportunity to introduce the general notion of a homomorphism and its systematic connections to the already familiar notions of an equivalence relation and that of one relation being congruence relation wrt. some other relation. And finally it throws an illuminating light on the ideas that Boole and De Morgan were after but that can be stated fully transparently only now that we know how to draw a clear distinction between sentences of a language as symbol strings with a syntactic structure and the samentic values ('propositions') they denote.

In the more formal discussion that follows we focus on first order languages as we have been doing hitherto. This will also allow for a natural transition to the topic of the next two sections.

Central to the discussion will be the language of boolean algebra, i.e. the language L_{ba} whose logical constants are $\cup$,$\cap$,0,1and -. Let L be any first order language. We can use the set S_L consisting of the sentences of L to define the following model M_L for L_{ba}: the universe is S_L and the interpretations of the non-logical constants of L_{ba} are given by the following function F_L:

$F_L(\cup)(A,B) = (A \vee B)$; $F_L(\cap)(A,B) = (A \& B)$; $F_L(0) = \neg(\forall v_1)\ v_1 = v_1$; $F_L(1) = (\forall v_1)\ v_1 = v_1$; & B); $F_L(-)(A) = \neg A$, where A and B are arbitrary sentences of L.

In other words, the 'boolean' operator symbols $\cup$ etc. are interpreted in as *syntactic* operations of the sentences of L. For instance, $\cap_L$ operates on arbitrary sentences (that is, arbitrary well-formed symbol strings) A and B of L and maps such a pair to the symbol string (A & B).

Now let $\mathbb{M}$ be some class of models for L. Then each sentence A of L can be said to express wrt $\mathbb{M}$ a 'proposition' $[[A]]^{\mathbb{M}}$, consisting of those models M in $\mathbb{M}$ for which $M \models A$: $[[A]]^{\mathbb{M}} = \{M \in \mathbb{M}: M \models A\}$. (It is reasonable to refer to $[[A]]^{\mathbb{M}}$ as the 'proposition expressed by A wrt. $\mathbb{M}$' insofar as $[[A]]^{\mathbb{M}}$ tells us for each $M \in \mathbb{M}$, and thus for each of the

'possible worlds' described by models of $\mathbb{M}$, whether or not A (or the proposition A expresses) is true in that world or model.)

From the propositions $[[A]]^{\mathbb{M}}$ expressed by A wrt. $\mathbb{M}$ it is possible to construct another model for L_{ba}, to which we refer as $M_{\mathbb{M}}$. The universe of this model is the set $\{[[A]]^{\mathbb{M}}$: A is a sentence of L} anf its interpretation function $F_{\mathbb{M}}$ is defined by:

$$F_L(\cup)([[A]]^{\mathbb{M}},[[B]]^{\mathbb{M}}) = ([[A]]^{\mathbb{M}} \cup [[B]]^{\mathbb{M}};$$
$$F_L(\cap)([[A]]^{\mathbb{M}},[[B]]^{\mathbb{M}}) = ([[A]]^{\mathbb{M}} \cap [[B]]^{\mathbb{M}};$$
$$F_L(0) = \varnothing;\ F_L(1) = \mathbb{M};\ F_L(-)([[A]]^{\mathbb{M}}) = \mathbb{M}\backslash[[A]]^{\mathbb{M}}.$$

(Here we have used bold face $\cup$ and $\cap$ to distinguish the set-theoretical union and intersection from the function constants $\cup$ and $\cap$ of the language L_{ba}.)

It follows directly from what we seen in the last section that $M_{\mathbb{M}}$ is a boolean algebra. On the other hand the model M_L is not, for one thing because syntactic disjunction and conjunction, the functions which interpret the function constants $\cup$ and $\cap$ in M_L, are not commutative. (For instance, in general, (A&B) is not the same string as (B & A); in particular, $(\neg(\forall v_1)\ v_1 = v_1\ \&\ (\forall v_1)\ v_1 = v_1)$ is not the same string as $((\forall v_1)\ v_1 = v_1\ \&\ \neg(\forall v_1)\ v_1 = v_1)$; and so on.) This means that the function $[[\ ]]^{\mathbb{M}}$ maps the non-boolean model for L_{ba} onto the boolean model $M_{\mathbb{M}}$.

Given a first order language L many different classes of models $\mathbb{M}$ are possible and for each such choice we get a different function $[[\ ]]^{\mathbb{M}}$. The possible choices of $\mathbb{M}$ are bounded on the one side by the smallest such choices- those where $\mathbb{M}$ is a singleton set {M} - and on the other side by the maximal choice, where $\mathbb{M}$ is the class of all models for L. When $\mathbb{M}$ = {M}, then the universe of the model $M_{\mathbb{M}}$ consists of just two elements, the set {M} itself and the empty set $\varnothing$. We can think of these two elements as 'true in $M_{\mathbb{M}}$' and 'false in $M_{\mathbb{M}}$' and replace them by 1 and 0. This gives us the 2-element boolean algebra, whose universe is the set {0,1} and whose operations are the familiar connectives of classical propositional logic, given by the classical truth tables. (For example, the interpretation of $\cap$ in this model is the 2-place function $\underline{\&}$ defined by $\underline{\&}(1,1) = 1$; $\underline{\&}(1,0) = \underline{\&}(0,1) = \underline{\&}(0,0) = 0$, and so on.) In this case the

notion of a 'proposition wrt $M_{\mathbb{M}}$' reduces to that of a mere truth value. $[[\ ]]^{\mathbb{M}}$ throws together any two sentences that have the same truth value in M and we end up with just two 'bags' one for the sentences of L that are true in M and one for the false sentences.

At the other extreme, where $\mathbb{M}$ is the class of all models for L, we get a maximal diversity of bags. Now two sentences A and B end up in the same bag only iff tghey are logically equivalent: $[[A]]^{\mathbb{M}} = [[B]]^{\mathbb{M}}$ iff for every model M for L, $M \models A$ iff $M \models B$.

The function $[[\ ]]^{\mathbb{M}}$ is an example of a *homomorphism*. Homomorphisms are maps from one structure into another which are structure-preserving. In general such maps are not 1-1. And that is true also for $[[\ ]]^{\mathbb{M}}$, since any two different logically equivalent sentences will be mapped onto the same value. For instance, we have for any sentences A and B that $[[(A \ \& \ B)]]^{\mathbb{M}} = [[(B \ \& \ A)]]^{\mathbb{M}}$, even though the two conjunctions (A & B) and (B & A) are, as we have just observed, in general distinct. In fact, the point of a homomorphism is often that it isolates those aspects of a given type of structure that are relevant from a certain perspective while abstracting from all remaining features. It does this by 'throwing into the same bag' any two elements for which the structural features that are relevant from the given perspective are the same and that thus only differ in respects that do not matter. Thus $[[\ ]]^{\mathbb{M}}$ identifies, by mapping them onto the same value, any two sentences whose structure guarantees that they have the same truth value in all models of $\mathbb{M}$.

We will define the notion of a homomorphism only for algebraic structures - that is, for models of algebraic languages. (There is a way of generalising the notion to arbitrary first order languages, some or all of whose non-logical constants are predicates, but since we won't need this generalisation here or later, we will limit ourselves to the case of algebraic languages only.

Def. 8

a. Let L be any algebraic language, M, M' models for L, h a function from U_M into $U_{M'}$. h is a *homomorphism from* M *into* M' iff for every non-logical constant f^n of L and every n elements $d_1, ..., d_n$ from U_M:

$$h(f^n{}_M(d_1,..,d_n)) = f^n{}_{M'}(h(d_1),..,h(d_n))$$

Special cases are those where a homomorphism h from M into M'is (i) onto M' and (ii) where h is 1-1. It is immediate that if h is both 1-1 and onto, then it is an isomorphism from M onto M'. We already noted that $[[\ ]]^{\mathbb{M}}$ is onto $M_{\mathbb{M}}$. Usually $M_{\mathbb{M}}$ is a proper submodel of the set inclusion lattice $M' = \langle P(\mathbb{M}), \subseteq \rangle$, and when that is so, $[[\ ]]^{\mathbb{M}}$ is a homomorphism into, but not onto, M'. (Exercise: For which combinations of of a first order language L and a class $\mathbb{M}$ of models for L is $M_{\mathbb{M}}$ a proper submodel of M'?)

There is an important general connection between homomorphisms and congruence relations. Again we use our 'syntax-semantics interface function' $[[\ ]]^{\mathbb{M}}$ to illustrate the matter. As a preliminary recall that there is a general correlation between functions and equivalence relations: (i) Let f be a function defined on some set X. Then f induces an equivalence relation $\sim$ on X, defined by:

(1) for any $x,y \in X$, $x \sim y$ iff $f(x) = f(y)$.

Conversely, any $\sim$ equivalence relation on a set X induces a function on X which maps each $x \in X$ onto the equivalence class $[x]_{\sim}$ it generates under $\sim$. Moreover, when (1) is applied to this function, it gets us back to the relation $\sim$.

This correlation holds in particular for functions that are homomorphisms. In particular, when h is a homomorphism from one structure M into another structure M', then there will be a corresponding equivalence relation $\sim$ on U_M induced by h via (1). In this case, however, $\sim$ has additional properties, which reflect the fact that h is a homomorphisms (and not just any function): is a *congruence relation wrt* each of the operations of M. We recall the notion of a congruence relation: Suppose that f is an n-place function defined on some set X, i.e. both the arguments and the values of f belong to X, and that $\sim$ is a binary relation on X. Then $\sim$ is a *congruence relation wrt* f iff for any $x_1, ..., x_n, x'_1, ..., x'_n$ from X such that $x_1 \sim x'_1, ..., x_n \sim x'_n$, $f(x_1,.,x_n) \sim f(x'_1,.,x'_n)$.

It is easily verified that when h is a homomorphism from a model M for an algebraic language L into some other model M' for L, then the relation $\sim$ induced by h via (1) is congruence relation wrt. all interpretations in M of function constants of L. Moreover, the converse

also holds in this case: If $\sim$ is an equivalence relation on U_M which is a congruence relation on the interpretations in m of all the non-logical constants of L, then the function which maps any element d of U_M onto its equivalence class $[d]_\sim$ is a homomorphism from M into the model M' whose universe is the set of equivalence classes $[d]_\sim$ and which interprets each n-place function constant f of L via the definition:

$$f_{M'} = \{<[d_1]_\sim, ..,[d_n]_\sim,[d]_\sim>: d_1, ..,d_n,d \varepsilon U_M \ \& \ f(d_1,..,d_n) = d\}$$

(This definition is legitimate because $\sim$ is a congruence relation wrt f.)

Returning to $[[\ ]]^{\mathbb{M}}$ we recall that this function is a homomorphism in that this function preserves the interpretations of all the function constants of L_{ba}. (For instance, $[[\ ]]^{\mathbb{M}}$ converts the syntactic conjunction operation & into the 'propositional conjunction' which maps the model sets $[[A]]^{\mathbb{M}}$ and $[[B]]^{\mathbb{M}}$ onto their intersection.) It follows from the general connection between homomorphisms and congruence relations we have described above that the relation which holds between sentences A and B iff they have the same truth values in each of the models of $\mathbb{M}$ is a congruence relation wrt to the syntactic operations that interpret the function constants of L_{ba} in $M_{L_{ba}}$. This is the formal justification for looking at the connectives of classical propositional logic as algebraic operations on 'sentence meanings'.

As noted in footnote **??**, the conception of the way in which meaning is determined by form as a homomorphism that maps syntactic strings onto meanings, thereby identifying any two strings whose structures make them identical in meaning, is a central assumption in the approach to meaning in natural languages developed by Montague in the late sixties and early seventies and now generally known as 'Montague Grammar'. The idea is that the syntax of any language - natural languages no less than the formal languages of logic and computer science (including in particular the first order languages that are the topic of these Notes) - can always be characterised by a set of syntactic operations which build complex expressions from constituents, and that to each such syntactic operation corresponds a rule which combines the semantic values of the constituents into the semantic value of the expression that is the output of the syntactic operation. It became clear soon that (except for very restricted fragments) the strictest implementation of this conception comes at a cost of assumptions about the syntax of natural languages that are quite artificial, and are ill supported by intrinsically syntactic evidence, of the

kind that linguists do, and should, take seriously. Nevertheless, the attempt to develop a syntax-semantics interface that is based on an independently plausible syntax and yet keeps as closely to Montague's original conception has proved a principle of immense methodological value in the development of semantics over the past 40 years.

The model $M_{\mathbb{M}}$ for L_{ba} that we obtain when $\mathbb{M}$ is the class of all models for L is known as the *Lindenbaum algebra of* L. Lindenbaum algebras will play an important part in the next section, be it in the different guise of structures whose elements are the finitely axiomatisable deductive theories of a given first order language L. (There is an obvious 1-1 correspondence between the finitely axomatisable theories of L and the classes $[[A]]^{\mathbb{M}}$ into which $[[\ ]]^{\mathbb{M}}$ partitions the set of all sentences of L and that make up the universe of $M_{\mathbb{M}}$ when $\mathbb{M}$ contains all models for L. For on the one hand, if t is a finitely axiomatisable theory of L, then there is a single sentence A of L such that $T = Cl_L(\{A\})$. On the other hand, when two sentences A and A' belong to the same class, i.e. if $[[A]]^{\mathbb{M}} = [[A']]^{\mathbb{M}}$, then A and A' are logically equivalent and thus axiomatise the same theory: $Cl_L(\{A\}) = Cl_L(\{A'\})$. Thus each finitely axiomatisable theory of L corresponds to exactly one element of the universe of $M_{\mathbb{M}}$.)

2.2 Incomplete Theories and their Extensions.

In section 2.1 we saw that complete theories do not always do what one might have expected of them, and for which they are often designed: describe a given structure uniquely up to isomorphism. A complete theory always succeeds in doing this, we observed, when the structure it is meant to describe is finite. (See Thm. 6 of Ch.1.) But for theories with infinite models the picture is much more complicated. We know that if a complete theory has an infinite model, then all its models are infinite (see exercise ??). But the differences between these infinite models may still be considerable. Not only will the theory always have models that are not isomorphic for the simple reason that their universes are of different cardinality - recall that the Skolem-Löwenheim Theorems tell us that theories with infinite models always have models of every possible infinite cardinality -, there exist complete theories that have non-isomorphic models even within the same cardinality. Though Morley's Theorem indicates that the range of possibilities us much more limited than one might have thought, there nevertheless remains considerable room for variation. For suppose a theory T has non-isomorphic models in some infinite cardinality κ. Then there is the further question how *wide* the variety of models of T of cardinality κ is. To answer this question a much finer - and much deeper - analysis of complete first order theories is needed than anything presented in these notes. Such an analysis exists. it is known as Stability Theory, a subject of considerable complexity, developed and brought to conclusion almost single-handedly by the Israeli mathematician and logician Saradon Shelah **[ref. to Shelah]**

When we move from complete to incomplete theories we find much wider ranges of possible models. Now the models of a theory T can be given a first classification in terms of the sentences they verify, in other words, in terms of those of the complete extensions of T which they verify. So the range of models of an incomplete theory T can be studied from two complementary perspectives, first the set of complete extensions of T, and second, for each of these complete extensions the range of models for that extension.

So far we have encountered examples of complete as well as of incomplete theories. But we haven't looked much at the structure of the entire field of theories in a given language L, including both its complete and its ncomplete theories. It is this issue that we will pursie in the present section.

2.2.1 Lattices of Theories.

Let L be a first order Language and let $\mathbb{T}_L$ be the set of all theories of L. The structure $\mathcal{T}_L = <\mathbb{T}_L, \subseteq>$, where $\subseteq$ is the relation of set-theoretical inclusion restricted to $\mathbb{T}_L$, is called the *Lattice of Theories of L.* We will also refer to it as the *Tarski Lattice of L* in honour of the logician A. Tarski, who was the first to study these structures.

Our first task is to show that $\mathcal{T}_L$ is a distributive lattice with 0 and 1. We already noted in the precious sections that any restriction $\subseteq_V$ of $\subseteq$ to some set V of sets is a weak partial order on V. To show that when V = $\mathbb{T}_L$ this partial ordering is a lattice, we must show that for each pair of theories T_1 and T_2 of L $\subseteq \mathbb{T}_L$ yields an infimum and a supremum with. First, note that $T_1 \cap T_2$ (where $\cap$ is set-theoretic intersection) is a theory of L. For suppose B is any sentence of L such that $T_1 \cap T_2 \models B$. Then $T_1 \models B$ and $T_2 \models B$, So, since T_1 and T_2 are theories, B ε T_1 and B ε T_2. So B ε $T_1 \cap T_2$. Since this holds for arbitrary B, $T_1 \cap T_2$ is a theory. It now follows almost directly that $T_1 \cap T_2$ is the infimum of T_1 and T_2 in $\mathcal{T}_L$. For if T is any theory of L such that T $\subseteq$ T_1 and T $\subseteq$ T_2, then T $\subseteq$ $T_1 \cap T_2$.

The case of $\cup$ is different because $T_1 \cup T_2$ is in general not a theory. (It is a theory only if $T_1 \subseteq T_2$ or $T_2 \subseteq T_1$, (See Exercise 20.ii of Ch.1) But T_1 and T_2 do have a supremum in $\mathcal{T}_L$ nevertheless, viz. the theory $Cl_L(T_1 \cup T_2)$.
To see this, observe that $T_1 \subseteq Cl_L(T_1 \cup T_2)$ and $T_2 \subseteq Cl_L(T_1 \cup T_2)$. Now suppose that T' is any theory of L such that $T_1 \subseteq$ T' and $T_2 \subseteq$ T'. Let B be any sentence from $Cl_L(T_1 \cup T_2)$. Then $T_1 \cup T_2 \models B$. So by the Completeness Theorem $T_1 \cup T_2 \vdash B$. From this is can easily be inferred that there must be a single sentence C ε T_1 and a single sentence D ε T_2, such that C & D $\vdash$ B. Since $T_1 \subseteq$ T', C ε $T_1 \cup T_2$ and thus C ε T'. Similarly D ε T'. So, T' $\models$ C & D and so since T' is a theory, C & D ε T'. So since C & D $\models$ B, also B ε T'.

Having shown that the supremum and the infimum of any two members of $\mathcal{T}_L$ exist, we facilitate further discussion by introducing the symbols $\cup_L$ and $\cap_L$ for these operations:

(1) (i) $T_1 \cup_L T_2 =_{df} Cl_L(T_1 \cup T_2)$
(ii) $T_1 \cap_L T_2 =_{df} T_1 \cap T_2$

That $\mathbb{T}_L$ has a 0 and a 1 is obvious. Its 0 is the theory $0_L = \{A \in L: \vdash A\}$ and its 1 the contradictory L-theory 1_L consisting of all sentences of L. That is distributive requires an argument. We show that the distributive law DISTR.2 holds in $\mathbb{T}_L$.[13] (The validity of the other law is shown in much the same way.)

DISTR.2 $T_1 \cup_L (T_2 \cap_L T_3) = (T_1 \cup_L T_2) \cap_L (T_1 \cup_L T_3)$

To show the inclusion of the left hand side in the right hand side is straightforward. (In fact this inclusion holds in all lattices.) To show inclusion in the opposite direction, let $B \in (T_1 \cup_L T_2) \cap_L (T_1 \cup_L T_3)$. Then $B \in (T_1 \cup_L T_2)$ and $B \in (T_1 \cup_L T_3)$. Since $B \in (T_1 \cup_L T_2)$, there are $C' \in T_1$ and $D \in T_2$ such that $C' \& D \vdash B$. Similarly, since $B \in (T_1 \cup_L T_3)$, there are $C'' \in T_1$ and $E \in T_3$ such that $C'' \& E \vdash B$. Putting $C =_{df} C' \& C''$, we have $C \& D \vdash B$ and $C \& E \vdash B$. So $C \& (D \vee E) \vdash B$. But $D \vee E \in T_2$ and $D \vee E \in T_3$. So $D \vee E \in T_2 \cap_T T_3$. So $T_1 \cup_L (T_2 \cap_L T_3) \vdash B$. So $B \in T_1 \cup_L (T_2 \cap_L T_3)$.

q.e.d.

While $\mathbb{T}_L$ is always a distributive lattice, it is never a boolean lattice. The reason is that if T is a theory of a first order language L which is not finitely axiomatisable, then there is no theory T' of L such that $T \cup_L T' = 1_L$ and $T \cap_L T' = 0_L$. And every first order language has theories that are not finitely axiomatisable. We record this fact as Theorem 4.

Thm. 4 For no first order language L is $\mathbb{T}_L$ a boolean lattice.

We postpone the proof of Thm. 4 till later in this section.

While $\mathbb{T}_L$ is never a boolean lattice, each $\mathbb{T}_L$ has a certain sublattice which invariably is boolean. This is the so-called *Lindenbaum algebra of* L.[14] It consists of all finitely axiomatisable theories of L, i.e. all

---

13 See Section 2.1.2. Note that here we have omitted the universal quantifiers binding T_1, T_2 and T_3.

14 Speaking on the one hand of 'Tarski lattices and on the other of Lindenbaum algebras will seem incoherent. The term 'Lindenbaum *algebra* has'

theories T of L such that for some finite set *A* of L-sentences $T = Cl_L(A)$. We denote the Lindenbaum Algebra *of* L as $\mathbb{L}_L$.

To show that $\mathbb{L}_L$ is a boolean lattice, we recall that a theory T is finitely axiomatisable iff there is a single sentence A such that $T = Cl_L(\{A\})$ - see Exercise 12.a of Ch. 1. (For easier reading we write 'T_A' instead of '$Cl_L(\{A\})$'.) It is straightforward to verify that if T_1 and T_2 are finitely axiomatisable theories of L and $T_1 = T_A$ and $T_2 = T_B$, then the following two conditions hold (Exercise: Show this.)

(1) (i) $T_1 \cup_L T_2 = T_{A\&B}$
(ii) $T_1 \cap_L T_2 = T_{A \vee B}$

Now let T be any finitely axiomatisable theory of L and suppose that $T = T_A$. Let $T' = T_{\neg A}$. Then according to (3.i,ii) $T_A \cup_L T_{\neg A} = T_{A\&\neg A}$ and $T_A \cap_L T_{\neg A} = T_{A \vee \neg A}$. But $T_{A\&\neg A} = Cl_L(\{A\&\neg A\}) = 1_L$ and $T_{A \vee \neg A} = Cl_L(\{A \vee \neg A\}) = 0_L$. So T' is the complement of T, in that the two satisfy the characteristic equations, repeated in (2).

(2) (i) $T \cup_L T' = 1_L$
(ii) $T \cap_L T' = 0_L$

Since for each member T of $\mathbb{L}_L$ there is a complement T' in $\mathbb{L}_L$ such that (2.i,ii) are satisfied, $\mathbb{L}_L$ is boolean. q.e.d.

As noted in the remarks leading up to Thm. 4, theories that are not finitely axiomatisable do not have boolean complements. However, it is possible to define an operation on arbitrary theories that (a) satisfies at least one of the conditions in (2), viz. (2.ii), (b) is the largest element satisfying this condition and (c) coincides with the boolean complement of any finitely axiomatisable theory. One definition of this operation is given in Def. 9.

It is possible to define a complement operation on theories of L which acts as a boolean complement when the theory in question is a theory

---

been adopted because of its general use in the literature - few people if anyone speak of the Lindenbaum *lattice* of L. Because of the equivalence between lattices and algebras nothing much hangs on this terminological issue. In fact we might just as well speak of Lindnbaum latticesa as of Lindenbaum algebras, and likewise, speaking of Tarski algebras is just as legitimate as talking about Tarski lattices.

of $\mathbb{L}_L$. The definition we will give is such that it can be applied to arbitrary theories. But only when the theory is finitely axiomatisable, will the theory and its complement stand in the relations that are distinctive of boolean algebras.

Def. 9 Let T be an element of $\mathbb{T}_L$. The *pseudocomplement of* T i*n* $\mathbb{T}_L$, $-_L T$, is defined by: $-_L T = \cup \{T' \in \mathbb{T}_L : T \cap_L T' = 0_L\}$[15]

Prop. 5 (i) $-_L T$ is the largest theory T' of L such that $T \cap_L T' = 0_L$.
(ii) Suppose that $T = T_A$. Then $-_L T = T_{\neg A}$.

Proof.

(i) Let $\underline{T} = Cl_L(-_L T)$. Suppose that $B \in T \cap \underline{T}$. Then $B \in T$ and there is a $C \in -_L T$ such that $C \models B$. But if $C \in -_L T$, then there is some theory T' such that $T \cap_L T' = 0_L$ and $C \in T'$. Since $C \in T'$ and $B \in T$, $C \vee B \in 0_L$. On the other hand, since $C \models B$ and $B \models B$. $C \vee B \models B$. So, since 0_L is a theory, $B \in 0_L$. This establishes that $\underline{T}$ is a theory T' such that $T \cap_L T' = 0_L$. Therefore $Cl_L(-_L T) = \underline{T} \subseteq -_L T$. So $-_L T = Cl_L(-_L T)$. That is, $-_L T$ is a theory It now follows directly from Def. 8 that it is the largest theory T' such that $T \cap_L T' = 0_L$.

(ii) Suppose that $T = T_A$. Then, as we have already seen, $T_{\neg A}$ is a theory T' such that $T \cap_L T' = 0_L$. So $T_{\neg A} \subseteq -_L T$. Now let T' be any theory such that $T \cap_L T' = 0_L$. Suppose that $B \in T'$. Then, since $A \in T$, $A \vee B \in 0_L$; that is, $\models A \vee B$. But $A \vee B$ is logically equivalent to $\neg A \rightarrow B$. So $\models \neg A \rightarrow B$, and therefore $\neg A \models B$. So $B \in T_{\neg A}$. This establishes that $-_L T \subseteq T_{\neg A}$. So $T_{\neg A} = -_L T$.

q.e.d.

---

15 Tarski lattices are thus structuresc which, according to a well-esrablished terminology are called *pseudo-complemented lattices*. A pseudo-complemented lattice is a lattice with an additional 1-place operation - with the properties that for all x, -x is the largest element such that $x \cap -x = 0$. Tarski-lattices have additional properties, one of which is that they are distributive. In fact, most of the well-known examples of pseudo-complemented lattices that are not Boolean algebras are distributive. However, the existence of a pseudo-complement does not entail distributivity. For instance, the 5-element lattice of Section 2.1.3 is pseudo-complemented (-1 = 0, -0 = 1, - a = b, -b = -c = a), but as we saw it is not distributive. Sometimes the pseudo-complement of x is defined as the smallest element y such that $x \cup y = 1$. From a formal poitn of view this comes in last analysis to the same ting because of the duality of $\cup$ and $\cap$.

We now proceed to the proof of Thm. 4.

Proof of Thm. 4

(a) Suppose that T is a theory of some first order language L and that $T \cup_L {}^{-L}T = 1_L$. Then there are a sentence A from T and a sentence B from ^{-L}T such that $A \& B \models \bot$. This entails that $B \models \neg A$. So we have $\neg A \in {}^{-L}T$. We show that $T = T_A$. Suppose that $C \in T$. Then, since $\neg A \in {}^{-L}T$, $C \vee \neg A \in 0_L$. So $\models C \vee \neg A$, which is equivalent to: $A \models C$. So $C \in T_A$. So we have shown that $T \subseteq T_A$. On the other hand, since $A \in T$, $T_A \subseteq T$. So $T = T_A$.

(b) We observe that the following infinite set of sentences $\{D_n\}_{n = 2,3,..}$ is strictly increasing in that for all n, $D_{n+1} \models D_n$ but not $D_n \models D_{n+1}$:

D_2: $(\exists v_1)(\exists v_2)\ v_1 \neq v_2$
D_3: $(\exists v_1)(\exists v_2)(\exists v_2)\ (v_1 \neq v_2 \& v_1 \neq v_3 \& v_2 \neq v_3)$
.
.
(D_n says that there are at least n different elements in the universe.)

Let L be any first order language and let $T_{inf,L}$ be the theory axiomatised by the sentences , i.e. $T_{inf,L} = Cl_L(\{D_n\}_{n = 2,3,..})$. (Note that the sentences only use logical vocabulary and thus belong to any first order language whatever.) Then according to Exercise 7.b of Ch. 1 $T_{inf,L}$ is not finitely axiomatisable. So $T_{inf,L}$ has no complement in L satisfying both of the two conditions (2.i,ii).
It follows that for no L is $\mathbb{T}_L$, the Tarski lattice for L, a boolean lattice.
q.e.d.

So far we have considered the Tarski lattices $\mathbb{T}_L$ of first order languages and just one type of substructure of those, the Lindenbaum algebras. But of course we could in principle study many other sublattices of the $\mathbb{T}_L$s. Of special importance among those sublattices are certain lattices whose bottom element is not 0_L, but rather some theory T of L. More particularly, it has proved useful in a variety of contexts to study (i) the lattice consisting of all extensions of T, and (ii) the lattice consisting of the *finitely axiomatisable* extensions of T (those extensions T' of T for

which there is a sentence A of L such that T' = $Cl_L(T \cup \{A\})$.[16] We call these the T*arski lattice of* L *generated by* T and the *Lindenbaum algebra of* L *generated by* T, respectively, and denote them as $\mathbb{T}_{L,T}$ and $\mathbb{L}_{L,T}$.

Def. 10 Let L be a language, T a theory of L.

a. The *Tarski lattice of* L *generated by* T is the structure $\mathbb{T}_{L,T} = <\mathbb{T}_{L,T}, \subseteq>$, where $\mathbb{T}_{L,T}$ is the set of all L-extensions of T and $\subseteq$ is the relation of set-theoretic inclusion restricted to $\mathbb{T}_{L,T}$.

b. The *Lindenbaum algebra of* L *generated by* T is the structure $\mathbb{L}_{L,T} = <\mathbb{L}_{L,T}, \subseteq>$, where $\mathbb{L}_{L,T}$ is the set of all L-extensions of T which are finitely axiomatisable over T - that is. All those L-extensions T' of T for which there is an L-sentence A such that T' = $Cl_L(T \cup \{A\})$) and $\subseteq$ is the inclusion relation on $\mathbb{L}_{L,T}$.

Like $\mathbb{T}_L$, $\mathbb{T}_{L,T}$ is always a distributive lattice with 0 and 1. This can be shown in just the same way as we did for $\mathbb{T}_L$. The argument that $\mathbb{L}_{L,T}$ is always boolean also goes as before. So far, then, there is no difference between the more general cases of $\mathbb{T}_{L,T}$ and $\mathbb{L}_{L,T}$ and the more specific cases of $\mathbb{T}_L$ and $\mathbb{L}_L$, in which the bottom element is 0_L. But there is nevertheless one difference, viz. that among the lattices $\mathbb{T}_{L,T}$ we now find many that are boolean (while, as we have seen, this is never so for the lattices $\mathbb{T}_L$). It can be inferred from what has already been established in this section that this happens only when the Tarski lattice generated by T and the Lindenbaum algebra generated by T coincide, i.e. when all extensions of T are finitely axiomatisable over T. In the next section we will see a number of comparatively simple examples of this situation.

Besides the lattices $\mathbb{T}_{L,T}$ and $\mathbb{L}_{L,T}$ other sublattices of $\mathbb{T}_L$ are worth consideration as well. Among these are in particular the lattice of all subtheories of a given theory T and the lattice consisting of all its finitely axiomatisable subtheories. (Exercise: prove that the former is again a distributive lattice with 0 and 1, where the set of tautologies of L is the 0 and T is the 1, and that the latter is a boolean lattice.) Even

---

16 Often the lattice $\mathbb{T}_{L,T}$ provides us with certain insights into the nature of T. For by telling us something aqbout the range of possible extensions of T it also tells us something about the range of its possible models, or true interpretations. and with that of the range of variability among the models of T.

more generally, we can, for any pair of L-theories T and T' such that $T \subseteq T'$, consider the Tarski lattice and Lindenbaum algebra consisting of those L-theories (or finitely axiomatisable L-theories, respectively) that lie between T and T', - in other words, at the sublattices of $\mathbb{T}_L$ whose 0 is T and whose 1 is T'. None of these, however, will be further considered in these Notes.

We have already observed that $\mathbb{T}_L$ is never boolean - not even for the simplest language {}. This is not so for the lattices $\mathbb{T}_{L,T}$. These can be boolean. Among them is the trivial lattice $\mathbb{T}_{L,\perp_L}$, whose only element is $\perp_L$, and all two element lattices $\mathbb{T}_{L,T}$, for T a consistent and complete theory of L, lattices whose only elements are $\perp_L$ and T.

In general, lattices of the form $\mathbb{T}_{L,T}$ are always both atomic and complete. More precisely, this is so for any such lattice with more than two elements. (If a lattice has ≤ 2 elements, then there are no atoms and the concept of atomicity is not applicable.) To see that $\mathbb{T}_{L,T}$ is atomic, assume that $\mathbb{T}_{L,T}$ has > 2 elements and observe that the complete consistent extension of T are the 'anti-atoms' of $\mathbb{T}_{L,T}$: they are those theories different from the inconsistent theory of L such that there is no theory between them and the inconsistent theory. It is easy to show - Exercise: do this! - that the atoms of $\mathbb{T}_{L,T}$ are precisely the theories $-_L T'$ where T' is any complete and consistent extension of T. With this in mind it is easy to see that $\mathbb{T}_{L,T}$ is atomic. For let T' be any proper extension of T (i.e. any extension of T that is different from T). Let A be any consistent sentence in T'\T - there will be such sentences if $\mathbb{T}_{L,T}$ has > 2 elements - and let T'' be any complete and consistent extension of $Cl(\{\neg A\})$. Then $-_L T''$ is an atom below T'. (Exercise: prove this!)

That $\mathbb{T}_{L,T}$ is a complete lattice is straightforward. Let $\mathcal{T}$ be any set of extensions of T. It is easy to show that $Cl_L(\cup \mathcal{T})$ is the supremum of $\mathcal{T}$.

We already know that $\mathbb{T}_{L,T}$ is not always a boolean lattice. (In particular, this is never so when T is 0_L.) For some L and T, however, $\mathbb{T}_{L,T}$ is boolean. Trivial examples are those where T is the inconsistent theory of L, in which case $\mathbb{T}_{L,T}$ is the trivial boolean algebra consisting of just one element and the case we already considered, where T is a complete consistent theory, in which case $\mathbb{T}_{L,T}$ consists of two elements, T and the inconsistent theory of L. There are also many examples of boolean $\mathbb{T}_{L,T}$ of more than 2 elements. However, *all*

boolean lattices $\mathbb{T}_{L,T}$ are finite. Note that this does not simply follow from the fact that such lattices are atomic and complete. For there exist infinite atomic and complete boolean lattices, viz. the power set inclusion structures $\langle P(X),\subseteq\rangle$ in which X is infinite.

The fact that boolean lattices of the form $\mathbb{T}_{L,T}$ are always finite thus has to do with the special properties of theory lattices. Since we have already established that $\mathbb{T}_{L,T}$ is always atomic and complete, the argument is quite simple. It goes as follows. First we observe the following general property of complete atomic boolean lattices $\mathbb{L}$:

(1) Let $\mathbb{L}$ be a complete atomic boolean lattice and let A_1 and A_2 be two distinct sets of atoms of $\mathbb{L}$. Then the suprema in $\mathbb{L}$ of these two sets, $\sup(A_1)$ and $\sup(A_2)$, are distinct.

We prove (1) by making use of (2), which we leave as an exercise:

(2) Let $\mathbb{L}$ be a boolean lattice and let a, a' be distinct atoms of $\mathbb{L}$. Then $a \leq -a'$.

Proof of (1): Let A_1 and A_2 be two distinct sets of atoms of $\mathbb{L}$. Then there is an $a \in A_1 \setminus A_2$ or there is an $a \in A_2 \setminus A_1$. Assume that $a \in A_1 \setminus A_2$. Then by (2) for each $a' \in A_2$, $a' \leq -a$. So, $\sup(A_2) \leq -a$. On the other hand $a \leq \sup(A_1)$. So it is not the case that $\sup(A_1) \leq -a$; for that would mean that $a \leq -a$, which is obviously impossible, as it would entail that $-a = 1_{\mathbb{L}}$, which evidently it isn't. (If it were, then $a = --a = 0_{\mathbb{L}}$, and thus a would not be an atom.)

We next observe (3)

(3) Any complete, atomic boolean lattice $\mathbb{L} = \langle U,\leq\rangle$ with atom set A is isomorphic to the power set inclusion lattice $\langle P(A),\subseteq\rangle$.

(3) follows from (1) and (4), the proposition that in a complete atomic boolean lattice $\mathbb{L}$ each element other than $0_{\mathbb{L}}$ is the supremum of the set of all atoms below it.

(4) Let $\mathbb{L}$ be a complete atomic boolean lattice with atom set A and let b be any element of $\mathbb{L}$ such that $b \neq 0_{\mathbb{L}}$. Let A_b be the set of atoms below b: $A_b = \{a \in A: a \leq b\}$. Then $b = \sup(A_b)$.

The proof of (4) is left as an exercise. (See **Exercise ?? at the end of this Chapter.**)

In view of (1) and (4) we can define the following map h from $\mathbb{L}$ to $P(A),\subseteq>$: for b ε U such that $b \neq 0_{\mathbb{L}}$ $h(b) = \sup(A_b)$; and $h(0_{\mathbb{L}}) = \varnothing$. It is then easy to see that h is onto an that it transfers $\leq$ into the inclusion relation on $P(A)$.

Suppose now that $\mathbb{T}_{L,T}$ is infinite. Then because of (3) its atom set A must be infinite. Now let A' be any proper infinite subset of A. Since each element a of is finitely axiomatisable we can choose for each such a a single sentence A_a which axiomatises a. Let $T(A')$ be the theory of L which is the supremum of A' in $\mathbb{T}_{L,T}$. Then, since A' is a proper subset of A, there is at least one atom a that does not belong to A'. Then, as we have seen, $T(A') \leq -a$, so $T(A')$ consistent. But then $T(A')$ is not finitely axiomatisable. The argument is like that of Exercise 12 of Ch. 1. Let a_1, a_2, ... be an enumeration of all members of A'. Note that A' is denumerable. (Why?). Furthermore, let the sentences B_n be defined as follows: (i) $B_1 = A_{a_1}$; $B_{n+1} = B_n \,\&\, A_{a_{n+1}}$. Then it is easily verified (i) that the B_n are strictly increasing in logical strength - i. e. we have for all n that $B_{n+1} \models B_n$, but not $B_n \models B_{n+1}$ - and (ii) that $T(A') = Cl_L(\{B_n\}_n = 1,2, ..)$. So we can argue as in Exercise 12 of CH.1 that $T(A')$ is not finitely axiomatisable. But then, as shown in Exercise 21 of CH. 1, $T(A') \cup -T(A') \neq 1$. So $\mathbb{T}_{L,T}$ is not boolean.

This concludes the proof of our claim that when a lattice $\mathbb{T}_{L,T}$ is boolean, it must be finite. We record this claim once more, as part of the following more elaborate Theorem 5, which gives three additional equivalent conditions.

Thm. 5 Let T be a theory in some first order language L
Then the following five statements are equivalent:

(i) $\mathbb{T}_{L,T}$ is boolean.
(ii) T has finitely many complete extensions.
(iii) T has finitely many extensions. (i.e. $\mathbb{T}_{L,T}$ is finite.)
(iv) All of T's complete extensions are finitely axiomatisable over T.
(v) All of T's extensions are finitely axiomatisable over T.

The main work of the proof of Theorem 5 has been done above. What remains is left as an exercise.

Theorem 5 entails that boolean lattices of the form $\mathbb{T}_{L,T}$ are comparatively rare. They are found only 'at the upper end' of the set of all lattices $\mathbb{T}_L$, i.e. when T is cose to being complete. (The cases we have already mentioned, i.e. the lattices $\mathbb{T}_{L,T}$ where T is itself a complete theory, are the extreme examples of this.) In the next section we will look at some simple cases of boolean lattices of the form $\mathbb{T}_{L,T}$.

To get a clear picture of the structure of the lattices $\mathbb{T}_L$ for different languages L turns out to be a far from trivial problem. Only for the very simplest languages is it possible to describe the structure of $\mathbb{T}_L$ in fairly straightforward and readily understandable terms. This is so in particular for the language without any non-logical constants, {}. Already for the language {P} whose only non-logical constant is the 1-place predicate P, ta complete description proves to be considerably more involved. But a much higher degree of complexity is reached when the language contains predicates of 2 or more places or function constants whose arity is ≥ 1. There are all sorts of questions that can be asked here, for instance:

(a) What is the full range of isomorphism types of lattices $\mathbb{T}_L$ for various first order languages?

(b) How does the structure of $\mathbb{T}_L$ depend on L?

(c) Call two languages L_1 and L_2 *isomorphic* iff they have essentially the same signature; that is, if there is a bijection h of the set NLC_1 of non-logical constants of L_1 onto the set NLC_2 of non-logical constants of L_2 which preserves signature in that for any $\alpha \in NLC_1$, $L_1(\alpha) = L_2(h(\alpha))$.

Question: Are there (finite) non-isomorphic languages for which the corresponding theory lattices are isomorphic nevertheless? And if so, for which language pairs is this so?

To none of these questions do I have answers, and I do not know whether answers to them exist.

2.2.2. Tarski Lattices of some almost complete Theories

In this section we look at two examples of Tarski lattices $\mathbb{T}_{L,T}$ which are comparatively simple and tractable.

In the first example the theory T is the theory T_{den} of arbitrary dense linear orderings. One of the extensions of this the theory T_{rat} of the ordering of the rationals (or, what comes to the same thing: the theory of all dense linear orderings without beginning or end point) which we investigated in Section 2.1.1. Of T_{rat} we showed that it is ω-categorical, und thus, since it also has the property that all its models are infinite, complete.

T_{den} is axiomatised by the following axioms $T_{den}.0$ - $T_{den}.4$. $T_{den}.1$ - $T_{den}.4$ are from our earlier axiomatisation of T_{rat}; $T_{den}.0$ has been added in order to eliminate the degenerate order which consists of just one element. (In the case of T_{rat} this possibility was excluded by the presence of axioms L5 and L6, repeated below, which assert that there is no beginning and no end point, respectively-)

$T_{den}.0$ $\quad (\exists x)(\exists y)\ (x \neq y)$

$T_{den}.1$ $\quad (\forall x)(\forall y)\ (x < y \rightarrow \neg\ (y < x))$

$T_{den}.2$ $\quad (\forall x)(\forall y)(\forall z)\ ((x < y\ \&\ y < z) \rightarrow x < z)$

$T_{den}.3$ $\quad (\forall x)(\forall y)\ (x < y\ \vee\ x = y\ \vee\ y < x)$

$T_{den}.4$ $\quad (\forall x)(\forall y)\ (x < y \rightarrow (\exists z)\ (x < z\ \&\ z < y))$

L5. $\quad (\forall x)(\exists y)\ (x < y)$

L6. $\quad (\forall x)(\exists y)\ (y < x)$

Unlike T_{rat} T_{den} is of course not complete. But it is not far removed from that. It has a total of no more than four complete extensions. One of these is T_{rat}, which we get by adding the axioms L5 and L6. The other three are obtained by adding the other boolean combinations of these two axioms: (i) {¬ L5, L6}, (ii) {L5, ¬L6}, (iii) {¬L5, ¬L6}.

We denote the four extensions of T_{den} as(i) $T_{den}(+,+)$, (ii) $T_{den}(+,-)$, (iii) $T_{den}(-,+)$and (iv) $T_{den}(-,-)$. The + and - signs indicate the presence or absence of a first or last point. For instance, if the first sign is a plus, then the models of the theory all have a beginning point, and if it is - then all models don't. In other words, $T_{den}(+,+)$ is the theory we get by adding to T_{den} the axioms ¬ L5 and ¬ L6, and so on, In particular $T_{den}(-,-) = T_{rat}$.

That each of the theories $T_{den}(+,+)$, $T_{den}(+,-)$ and $T_{den}(-,+)$ is consistent and complete can be shown in the same way as we did this for T_{rat} in Section 2.1.1. In fact, since the rational interval (0,1) is one of the models of T_{rat} (Exercise: show this!), it follows from what was shown in Section 2.1.1 that every denumerable model of T_{rat} is isomorphic to (0,1). Using the same method we can also prove that [0,1), (0,1] and [0,1] are models of $T_{den}(+,-)$, $T_{den}(-,+)$ and $T_{den}(+,+)$, respectively, and that they are the only denumerable models of these theories up to isomorphism. So since each of the theories only has infinite models (Exercise: show this!), they are all complete as well as consistent.[17]

It is also easy to show that these are all the complete and consistent extensions of T_{den}. For suppose that T is any complete extension of T_{den} and that M is a model of T. M will either have or fail to have a first point and likewise it will either have or fail to have a last point. This gives a total of four possibilities, corresponding to the four boolean combinations of L5 and L6 mentioned above. In each case T is identical with the theory we get by adding this boolean combination to T_{den}. For instance, suppose that M has both a first and a last point. Then it will verify both ¬ L5 and ¬ L6. So these sentences are consistent with T, and so, since by assumption T is complete, they must belong to T. So T is the theory $T_{den}(+,+)$. Likewise for the other three possibilities.

This shows that the lattice $\mathbb{T}_{L,T_{den}}$ has exactly four 'anti-atoms'. So it also has exactly four atoms, which means that it consists of 2^4 theories altogether. Exercise: give explicit axiomatisations for each of the theories that make up $\mathbb{T}_{L,T_{den}}$!

2.2.3 Quantifier Elimination

---

[17] The same is true for the other three complete extensions of T_{den}. Consider for instance $T_{den}(+,-)$. The only complication which we have to deal with, when constructing martching tuples $<a_1, ..., a_n>$, $<b_1, ..., b_n>$ from two models M_1, M_2 of $T_{den}(+,-)$ is that if $<a_1, ..., a_n>$ contains the first element of M_1, and more precisely, if this first element is a_i, then b_i must be the first element of M_2, and conversely. That that is the only additional precaution we need to take in constructing the finite sequences $<a_1, ..., a_n>$, $<b_1, ..., b_n>$ and the isomorphiisms between them rests on the fact that all elements of M_1 (casu quo M_2) which are distinct from its first element are "infinitely far away from it" in the sense that there are infinitely many points between any such point and the first point (just as there are infinitely many points between any two distinct points of any model of T_{den},)

Our second example concerns the theory of discrete linear orderings. We will explore the Tarski lattice $\mathbb{T}_{L,T_{dis}}$, where T_{dis} is the theory defined below.

This exploration will be more involved than that of $\mathbb{T}_{L,T_{den}}$ in the last section, and that for two distinct reasons. First, $\mathbb{T}_{L,T_{dis}}$ is a more complex lattice than $\mathbb{T}_{L,T_{den}}$, although its complexity is still quite modest when compared with most Tarski lattices. But also - and this will be the bigger hurdle we will encounter - proving that the structure of the lattice is indeed what we will claim it to be, will prove a good deal more involved than tit was in the case of $\mathbb{T}_{L,T_{den}}$ and it will require a fundamentally different method. This is the method of quantifier elimination mentioned in the title to this section.

The base theory of our lattice, T_{dis}, is once more a theory of the language L = {<}. T_{dis} is axiomatised by the axioms $T_{dis}.0$ - $T_{dis}.5$. Not surprisingly there is a considerable overlap with the axioms of T_{dis}. For after all both theories deal with linear orderings. Consequently the first four axioms are the same, and divergence from T_{den} comes only with the discreteness axioms $T_{dis}.4$ and $T_{dis}.5$.

$T_{dis}.0$ $\quad (\exists x)(\exists y)\ (x \neq y)$

$T_{dis}.1$ $\quad (\forall x)(\forall y)\ (x < y \rightarrow \neg\ (y < x))$

$T_{dis}.2$ $\quad (\forall x)(\forall y)(\forall z)\ ((x < y\ \&\ y < z) \rightarrow x < z)$

$T_{dis}.3$ $\quad (\forall x)(\forall y)\ (x < y\ \vee\ x = y\ \vee\ y < x)$

$T_{dis}.4$ $\quad (\forall x)((\exists y)\ (x < y \rightarrow ((\exists y)\ (x < y\ \&\ \neg\ (\exists z)\ (x < z\ \&\ z < y))$

$T_{dis}.5$ $\quad (\forall x)((\exists y)\ (y < x \rightarrow ((\exists y)\ (y < x\ \&\ \neg\ (\exists z)\ (y < z\ \&\ z < x))$

T_{dis} is not complete and for much the same reasons as T_{den}: Nothing is said about the existence or non-existence of beginning or end points. Using the same notation that we resorted to in our discussion of T_{den}, we define the theories $T_{dis}(+,+)$, $T_{dis}(+,-)$, $T_{dis}(-,+)$ and $T_{dis}(-,-)$ to be those which we get by adding the boolean combinations of L5 and L6 described in the last section. (Thus $T_{dis}(+,+)$ is obtained by adding $\neg$ L5 and $\neg$ L6, etc.) All of these have, like the corresponding extensions of T_{den}, infinite models. In particular, $T_{dis}(+,-)$is satisfied by the ordering of the natural numbers, $T_{dis}(-,+)$ by the order of the negative integers,$T_{den}(-,-)$ by the order of the positive and negative integers and

$T_{dis}(+,+)$ by the structure which we get when we put the negative integers "behind" the natural numbers.[18]

There is however an important difference between $T_{dis}(+,+)$ and the other three: while the latter only have infinite models, $T_{dis}(+,+)$ has finite models as well. In fact, $T_{dis}(+,+)$ has models of cardinality n for all finite $n \geq 2$: any linearly ordered set of n elements will be a model of $T_{dis}(+,+)$.[19] On the other hand it is also clear that for each finite cardinality n there is essentially just one model for T_{dis} of that cardinality: Any two linearly ordered sets of n elements are (obviously) order-isomorphic; we can define, in the obvious way, an order-preserving correspondence between them. This means that if we add to $T_{dis}(+,+)$ a sentence which states that there are exactly n elements, then the resulting theory will have for its only models the linear orders of n elements. And since any two such orders are isomorphic, it follows that all these theories are complete.

In the spirit of the notation which we have been using, let us denote as $T_{dis}(+,+,n)$ the theories obtained by adding to $T_{dis}(+,+)$ a sentence saying that there are exactly n elements; and let us denote as $T_{dis}(+,+,\infty)$ the theory obtained by adding to $T_{dis}(+,+)$ the infinitely many sentences $D\geq n$ which say that there are at least n elements.

What can we say about the theories $T_{dis}(-,-)$, $T_{dis}(+,-)$, $T_{dis}(-,+)$ and $T_{dis}(+,+,\infty)$? The first pertinent observation is that unlike what we found for the corresponding extensions of T_{den}, these theories are not ω-categorical. Let us focus on $T_{dis}(+,-)$. One of its denumerably infinite models, we noted, is the set of the natural numbers with their natural order. But there are other denumerably infinite models too, and

---

18 More presicely, we can define this structure as the ordered disjoint union of these two structures, viz as the set of all pairs <0,n>, with $n \varepsilon N$ and all pairs <1,-n> with $n \varepsilon N$, with the ordering relation < defined by:

(i) $<0,n> < <0,m>$ iff $n <_N m$
(ii) $<1,-n> < <1,-m>$ iff $m <_N n$
(iii) $<0,n> < <1,'m>$ for arbitrary n, m

19 The requirement that $n \geq 2$ comes from $T_{dis}.0$, which we have retained from our axiomatisation of T_{den}. We could have dropped this axiom without changing much to the structure of $\mathcal{T}_{L,T_{dis}}$. The only effect would have been that the degenerate, one point ordering would have been included among the possible models of T_{dis}. This would have meant that in addition to the complete extensions of T_{dis} we are in the process of describing there would have been the extension which says that there is exactly one point.

as a rule these will not be isomorphic to the natural number structure. The simplest model of $T_{dis}(+,-)$ which is not isomorphic to the natural numbers is the structure that we obtain when we put a copy of Z (the negative and positive integers) behind a copy of the natural numbers. We can make this precise in the same way as we did for the infinite model we considered for $T_{dis}(+,+)$ described in footnote 16. That is we let M be the model $< U_M, <_M >$, where

(a) $U_M = \{<0,n>: n \in N\} \cup \{<1,z>: z \in Z\}$

(b) $<_M = \{<<0,n>:, <0,m>>: n <_N m\} \cup \{<<1,z>:, <1,y>>: z < y\}$
$\cup \{<<0,n>:, <1,z>>\}$

It is obvious that M is not isomorphic to the set N of natural numbers with their standard order. Just try to construct an isomorphism between N and M, starting with the 0 of N, 0_N. Obviously there is only one element of M on which an order isomorphism h from N to M could map 0, viz. M's first point <0,0>. In other words, it is necessarily the case that $h(0_N) =<0,0>$. Likewise the number 1 of N, 1_N, which is the immediate successor of 0 in N, can only be mapped onto the immediate successor <0,1> of $<0_M$ in M. That is, we must have $h(1_N) = <0,1>$. In the same way the structure of N and M fixes the images under h of all the other elements of N. This means that, when N has been exhausted - i.e. h has been defined for all of N - only the "N-part" of M (consisting of the pairs of the form <0,n>) has been covered in the range of h.

The non-isomorphism of N and M entails that the completeness of $T_{dis}(+,-)$ cannot be established by the simple technique which we used to prove Cantor's theorem (the ω-categoricity of $T_{den}(-,-)$) in Section 2.1.1 and which would also be applied to the three other extensions of T_{den} which we considered in the last section. Nevertheless, $T_{dis}(+,-)$ is complete and the same is true of the remaining three extensions of T_{dis} which have infinite models, $T_{dis}(-,-)$, $T_{dis}(-,+)$ and $T_{dis}(+,+,\infty)$. But the proof that they are complete is harder than the Cantor-type proofs for the corresponding extensions of T_{den}. We will give the proof for the case of $T_{dis}(+,-)$. The proofs that the three other theories are complete are virtually identical.

In presenting the proof that $T_{dis}(+,-)$ is complete we will proceed as follows. We first focus on the concrete task before us. We show that any two models of $T_{dis}(+,-)$ are elementary equivalent. This argument will reveal the general features of the method used (that of quantifier

elimination). Un the next section we will then describe and discuss the method of quantifier elimination in general.

Recall the basic architecture of Cantor's proof. We considered two models $M_1 = < U_1, <_1 >$ and $M_2 = < U_2, <_2 >$ of $T_{den}(-,-)$ and constructed, by going back and forth between the universes U_1 and U_2, ever longer matching n-tuples $<a_1, ..., a_n>$ of elements from U_1 and $<b_1, ..., b_n>$ of elements from U_2, which were order-isomorphic. Because of the special properties of dense linear orderings it proved to be always possible to match a new element a_{n+1} chosen from U_1 by a new element b_{n+1} from U_2 which stood in exactly the same order relations to each of the b_i ($i = 1, ..., n$) as a_{n+1} stood to each of the a_i; and conversely. For models of theories of discrete orderings - among them the models for $T_{dis}(+,-)$ - the situation is different. Here the "distance" between two points - i.e. the number of points between them - can be either finite or infinite; and the distance could involve any finite number n of intermediate points. The model N is special among the models of $T_{den}(+,-)$ in that the distance between two of its elements is always finite. But in this respect it is unique. Any model of $T_{den}(+,-)$ which is not isomorphic to N will have points that are infinitely far from each other. (This is true in particular for the model we considered above, in which a copy of N is followed by a copy of Z. In this model there is an infinite distance between any two elements $<0,n>$ and $<1,z>$.)

A consequence of this is that when we consider a formula A of our language and two tuples $<a_1, ..., a_n>$, $<b_1, ..., b_n>$ belonging to two models M_1, M_2 and ask whether A gets the same truth value in M_2 under the assignment provided by $<b_1, ..., b_n>$ that it gets in M_2 under the assignment provided by $<a_1, ..., a_n>$, then we will have to take into account the quantifier complexity of A: It will depend on this complexity how similar $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ will have to be in order that we can be certain that they confer upon A the same truth value in their respective models M_1 and M_2. A few simple examples will illustrate this.

First consider a quantifier-free formula, e.g. $v_1 < v_2$. Let M_1, M_2 be models of $T_{dis}(+,-)$ and let $<a_1,a_2>$, $<b_1,b_2>$ be ordered pairs of elements of M_1 and M_2 which are order-isomorphic to each other, i. e. $a_1 <_{M_1} a_2$ iff $b_1 <_{M_2} b_2$. Then clearly $M_1 \models (v_1 < v_2)[a_1,a_2]$ iff

$M_2 \models (v_1 < v_2)[b_1,b_2]$. The same holds for any other quantifier-free formulas such as $v_1 < v_2 \,\&\, v_2 < v_3$, $v_1 < v_2 \,\&\, \neg(v_2 < v_3)$, etc, etc. This is just as in the case of dense orderings.

As soon as A contains quantifiers, however, the mere order isomorphism between $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ will no longer suffice. For example, let A be the formula $(\exists v_2)(v_1 < v_2 \,\&\, v_2 < v_3)$. Suppose that M_1 and M_2 are both the natural number structure N and that $<a_1, a_2> = <4,7>$ and $<b_1,b_2> = <8,9>$. Then $<a_1,a_2>$ and $<b_1,b_2>$ are order-isomorphic; yet $N \models A[a_1,a_2]$, while on the other hand not $N \models A[b_1,b_2]$. The source of the problem is obvious. A says something about the distance between the points represented by its free variables v_1 and v_3, viz. that there is at least one point between them. This is a condition which a mere order isomorphism need not preserve. And that is precisely what we see in our example: $<a_1,a_2>$ and $<b_1,b_2>$ are both order-isomorphic, but the point pair $<a_1,a_2>$ satisfies the condition that there is at least one point between them whereas the pair $<b_1,b_2>$ does not. In other words, in order to be sure that two pairs $<a_1,a_2>$ and $<b_1,b_2>$ confer upon A the same truth value, they must not just be order-isomorphic, but stand in some tighter relationship, which also involves information about how many points there are between them.

As we move to formulas A more quantifiers even stronger similarity relations must hold between $<a_1,a_2>$ and $<b_1,b_2>$ to guarantee that a_1 and a_2 satisfy A in M_1 iff b_1 and b_2 satisfy A in M_2. This is because with more quantifiers we can say more about the number of points between two given points a_1 and a_2. For instance, with two quantifiers, but not with just one, it is possible to say that there are at least two points between a_1 and a_2; and so on. And the same goes, more generally, for formulas A with free variables $x_1, ..., x_n$: Ever stronger relations must hold between an n-tuple $<a_1, ..., a_n>$ of elements from M_1 and an n-tuple $<b_1, ..., b_n>$ of elements from M_2 in order to guarantee that $<a_1, ..., a_n>$ satisfies A in M_1 iff $<b_1, ..., b_n>$ satisfies A in M_2.

It would be convenient iff we could define a relation between tuples $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ such that any two tuples standing in this relation will confer the same truth value on all formulas. But often - this is true of our present problem but also for many others - there is no direct way of defining such a single relation; all that can be done is

to define a hierarchy $\equiv_1, \equiv_2$, ... of ever tighter relations between n-tuples so that whenever $<a_1, ..., a_n> \equiv_k <b_1, ..., b_n>$, then $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ confer the same truth values on all formulas whose *quantifier depth* is $\leq$ k. By the *quantifier depth* of a formula A we understand the maximal degree of nesting of quantifiers in A. There is no particular difficulty in defining this notion for arbitrary formulas. But it is somewhat more convenient to limit our attention to prenex formulas. For a formula A in prenex form the *quantiifer depth* of A is simply the number of quantifiers in its quantifier prefix. Since every formula is logically equivalent to a formula in prenex normal form, satisfaction preservation of all prenex formulas will entail preservation of all other formulas.

For the argument below it will be also convenient to asssume a slightly different form for prenex formulas, one in which the prefix contains only existential quantifiers but no universal ones. We can obtain such a prefix from a standard prefix by replacing every universal quantifier $(\forall v_i)$ in the standard prefix by the equivalent combination $\neg(\exists v_i)\neg$. So the formulas with which we will be concerned will always begin with a (possibly empty) prefix consisting of existential quantifiers and negation signs, followed by a quantifier-free formula. The *quantifier depth* of such a formula is then the number of existential quantifiers in its prefix.

In (1) we repeat for further reference the basic requirement we have already stated on the relations $\equiv_k$.

(1) Let M_1 and M_2 be models of $T_{dis}(+,-)$. And let A be any formula of quantifier depth $\leq$ k whose free variables are among $x_1, ..., x_n$. Then for any n-tuples $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ of elements chosen from M_1 and M_2, respectively, such that $<a_1, ..., a_n> \equiv_k <b_1, ..., b_n>$, $M_1 \models A[a_1,..,a_n]$ iff $M_2 \models A[b_1,.., b_n]$.

We already know what is required of the relation $\equiv_o$, which according to (1) should guarantee that if $<a_1, ..., a_n> \equiv_o <b_1, ..., b_n>$, then $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ satisfy the same formulas of quantifier depth 0 (i.e. the same quantifier-free formulas). This requires that the function h, given by the condition: $h(a_i) = b_i$, is an order isomorphism between (the submodels of M_1 and M_2 determined by) $\{a_1, ..., a_n\}$ and $\{b_1, ..., b_n\}$, respectively. We define $\equiv_o$ accordingly:

(2) Let $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ be n-tuples of elements chosen from models M_1 and M_2, respectively. Then

$<a_1, ..., a_n> \equiv_o <b_1, ..., b_n>$ iff

the function h given by: "for i = 1,2, .., n, $h(a_i) = b_i$" is an order isomorphism between the submodels of M_1 and M_2 whose universes are $\}a_1, ..., a_n\}$ and $\{b_1, ..., b_n\}$.

A second requirement on the relations $\equiv_k$, which is imposed by the strategy we will follow to show that two models M_1 and M_2 of $T_{dis}(+,-)$ are elementarily equivalent, is that successive relations $\equiv_k$ and $\equiv_{k+1}$ stand in the following relation:

(3) Suppose that M_1 and M_2 are as under (1), that, for arbitrary number n, $<a_1,..,a_n>$, $<b_1,..,b_n>$ are n-tuples of elements of M_1 and elements of M_2, respectively and that $<a_1,..,a_n> \equiv_{k+1} <b_1,..,b_n>$. Then

i. if a is any element of M_1, then there is an element b of M_2, such that $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$.

ii. if b is any element of M_2, then there is an element a of M_1, such that $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$.

From (2) and (3) we can derive that the condition (1) holds for all formulas of the special prenex form described above, in which a quantifier-free part is preceded by a string of existential quantifiers and negations. We repeat this restricted version of (1) as (1') below. Since every formula can be transformed into a logically equivalent formula of this special form, (1') entails (1).

(1') Let M_1 and M_2 be models of $T_{dis}(+,-)$. And let A be any prenex formula with a prefix consisting of existential quantifiers and negation signs, that A has quantifier depth $\leq$ k and that its free variables are among $x_1, ..., x_n$. Then for any n-tuples $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ of elements chosen from M_1 and M_2, respectively, such that $<a_1, ..., a_n> \equiv_k <b_1, ..., b_n>$,

$$M_1 \models A[a_1,..,a_n] \text{ iff } M_2 \models A[b_1,.., b_n].$$

To derive (1') from (2) and (3) we argue by induction on the complexity of such formulas. The base case is constituted by quantifier free formulas. Strictly speaking, this reuires an inductive proof in its own right: First, when $<a_1, ..., a_n> \equiv_o <b_1, ..., b_n>$ and A is an atomic formula - that is, A is either of the form "$v_i = v_j$" or of the form "$v_i < v_j$", then obviously A is satisfied by $<a_1, ..., a_n>$ in M_1 iff it is satisfied by $<b_1, ..., b_n>$ in M_2. The inductive step then consists in showing that the condition in (1') holds for B and for C then it holds for $\neg$ B, B & C, and likewise for the other sentence connectives. But this is trivial.

The inductive step makes use of (3). Suppose that (1') holds for formulas of quantifier depth $\leq$ k and that A is a formula in our special prenex form and is of quantifier depth k + 1. If A begins with a negation sign - i.e. A is of the form $\neg$ B, where B too has our special prenex form, then the result will hold for A provided it holds for B. Let us assume therefore that A begins with an existential quantifier, i.e. A is of the form $(\exists x)B$. Suppose then that the free variables of B are among $v_1, ..., v_n$ and that $<a_1, ..., a_n> \equiv_{k+1} <b_1, ..., b_n>$. Without loss of generality we may assume x is the variable v_{n+1}. (We do not really need this assumption, but it simplifies notation.) Assume that $M_1 \models A[a_1,..,a_n]$. Then there is an element a of M_1 such that $M_1 \models B[a_1,..,a_n,a]$. Given (3) we can find an element b in M_2 such that $<a_1, ..., a_n,a> \equiv_k <b_1, ..., b_n,b>$. By induction hypothesis $M_2 \models B[b_1,..,b_n,b]$. So it follows that $M_2 \models (\exists x)B[b_1,..,b_n]$. In the same way we show that if $M_2 \models A[b_1,..,b_n]$, then $M_1 \models A[a_1,..,a_n]$.

This concludes the argument that (1) provided that we can define a sequence of relations $\equiv_o, \equiv_1, \equiv_2,..$ such that $\equiv_o$ is the relation defined in (2) and successive relations $\equiv_k, \equiv_{k+1}$ satisfy (3). In the present case . the one concerning the theory $T_{dis}(+,-)$ - the relations can be given by independent explicit definitions. (In other applications of the quantifier elimination method their definition may be more complicat4ed and require itself an induction on k.) The definitions are given in (4)

(4) Let M_1 and M_2 be models of $T_{dis}(+,-)$. Let a_{beg} be the first element of M_1in the sense of its order relation $<_{M_1}$ - there must be a unique such element since M_1 is a model of $T_{dis}(+,-)$ - and let b_{beg} be the first element of M_2. Let $<a_1, ..., a_n>$ and $<b_1, ..., b_n>$ be n-tuples from M_1 and M_2, respectively.

Then $\langle a_1, ..., a_n \rangle \equiv_k \langle b_1, ..., b_n \rangle$ iff the following conditions are fulfilled:

(i) $\langle a_1, ..., a_n \rangle$ and $\langle b_1, ..., b_n \rangle$ are order-isomorphic.
(For simplicity we assume, as we have been all along, that their elements have been arranged "in order of magnitude", i.e. $a_1 <_{M_1} a_2$, etc. and similarly for the elements of $\langle b_1, ..., b_n \rangle$)

(ii) For any pair of successive elements a_i, a_{i+1} from the first tuple and corresponding pair b_i, b_{i+1} from the second we have either (a) or (b):

(a) the number of elements) between a_i and a_{i+1} in M_1 and that between b_i and b_{i+1} in M_2 are both $< 2^k$ and they are identical;

(b) the number of elements between a_i and a_{i+1} in M_1 and that between b_i and b_{i+1} in M_2 are both $\geq 2^k$.

(iii) For the elements a_1 and b_1 we have either (c) or (d):

(c) the number of elements between a_1 and a_{beg} and that between b_1 and b_{beg} are both $< 2^k$ and they are identical;

(d) number of elements between a_1 and a_{beg} and that between b_1 and b_{beg} are both $\geq 2^k$.

N.B. For the case where $k = 0$ condition (ii) is vacuous, since the first possibility they mention - of the distances between a_i and a_{i+1} and between b_i and b_{i+1} being $< 2^o$ - cannot arise. Similarly condition (ii) is vacuous, So only (i) matters and thus the specification that (4) provides of $\equiv_o$ coincides with that given in (2).

It remains to show that the relations of (4) satisfy (3). Suppose that $\langle a_1, ..., a_n \rangle \equiv_{k+1} \langle b_1, ..., b_n \rangle$. We have to show that for any element a of M_1 there is an element b of M_2 such that $\langle a_1, ..., a_n, a \rangle \equiv_{k+1} \langle b_1, ..., b_n, b \rangle$ and conversely. we only consider the first half. Let a be any element of U_{M1}. There are three possibilities to be considered:

(i) $a <_{M_1} a_1$;
(ii) $a_i <_{M_1} a <_{M_1} a_{i+1}$ for some $i < r$

(iii) $a_n <_{M_1} a$.

Assume (i). Let $D(a_{beg},a_1)$ be the number of elements between a_{beg} and a_1. Then either (c) $D(a_{beg},a_1) < 2^{k+1}$ or (d) $D(a_{beg},a_1) \geq 2^{k+1}$. First suppose (c). Since the number of elements in M_2 between b_{beg} and b_1, $D(b_{beg},b_1)$, is the same as $D(a_{beg},a_1)$, we can pick as the b required by (3) that element of M_2 which lies just as many elements before b_1 in M_2 as a lies before a_1 in M_1. Then the distance between b and b_1 is the same as that between a and a_1 and the same is true for the distance between the b and b_{beg} and the distance between a and a_{beg}. So $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$.

Now suppose that both $D(a_{beg},a_1)$ and $D(b_{beg},b_1)$ are $\geq 2^{k+1}$. First suppose that the distance between a and a_1 is $< 2^k$. Then we pick from M_2 the element b which lies before b_1 at just the same distance that a lies before a_1 in M_1. This guarantees that there are as many elements between a and a_1 in M_1 as there are between b and in b_1 in M_2. Moreover, since by assumption the distance between a_{beg} and a_1 and that between b_{beg} and b_1 are both $\geq 2^{k+1}$, the distance between a_{beg} and a and that between b_{beg} and b will be both $\geq 2^k$. So again we have that $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$.

The second possibility to be considered is that where $D(a_{beg},a) < 2^k$. Then we pick the element b of M_2 which lies at that same distance from b_{beg}. This time $D(a_{beg},a_1)$ and $D(b_{beg},b_1)$ are both $\geq 2^k$. So again $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$.

The third possibility we must consider for the position of a before a_1 is that where both $D(a_{beg},a)$ and $D(a,a_1)$ are $\geq 2^k$. In this case the fact that $D(a_{beg},a_1)$ is $\geq 2^{k+1}$ guarantees that we can pick an element b from M_2 such that $D(b_{beg},b)$ and $D(b,b_1)$ are both $\geq 2^k$. Again $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$.

This completes case (i), in which a lies before a_1 in M_1. We leave the other two cases - that where a lies between a_i and a_{i+1} for some $i < 1$ and that where a lies beyond a - to the reader, and thus reach the end of the argument that if $<a_1, ..., a_n> \equiv_{k+1} <b_1, ..., b_n>$, then for any choice of an element a from M_1 we can make a matching choice of an

element d from M_2 such that $<a_1,..,a_n,a> \equiv_k <b_1,..,b_n,b>$[20] and therewith the proof that definition (4) entails (3).

We have now proved (1'). One step remains towards the conclusion that M_1 and M_2 are elementarily equivalent. But this is straightforward. Suppose that A is any sentence and that A' is formula in our prenex form that is logically equivalent to A. Then A' may be assumed to also be a sentence. Suppose that A' has quantifier depth k. In order that $M_1 \models A'$ iff $M_2 \models A'$ we need to show that the empty sequence $< >$ of elements of M_1 satisfies A' in M_1 iff the empty sequence $< >$ of elements of M_2 satisfies A' in M_2. According to (1') this will be the case, provided these two sequences stand in the relation $\equiv_k$. But it is obvious from def. (4) that the empty sequences of elements of M_1 and M_2 trivially satisfy this requirement.

q.e.d.

We have now proved that any two infinite models of $T_{dis}(+,-)$ are elementarily equivalent. Since $T_{dis}(+,-)$ only has infinite models, $T_{dis}(+,-)$ is complete. The argument is much like the one justifying Vaught's Test. (See Ch. 1, **Theorem ??**.) Suppose that $T_{dis}(+,-)$ were not complete. Then there would be a sentence A such that neither A nor $\neg$ A belong to $T_{dis}(+,-)$. So both $T_{dis}(+,-) \cup \{A\}$ and $T_{dis}(+,-) \cup \{\neg A\}$ are consistent. So each of them has a model. Both models must be infinite. So, because of the Downward Skolem-Löwenheim theorem, we may assume that they are both denumerably infinite. So, since they are both models of $T_{dis}(+,-)$, it follows from what we have just proved that they are elementarily equivalent. This contradicts the assumption that the first model verifies A and the second $\neg$A.

By the same method that we have used to prove that $T_{dis}(+,-)$ is complete we can also prove completeness for the three remaining theories, $T_{dis}(-,+)$, $T_{dis}(-,-)$ and $T_{dis}(+,+,\infty)$. This rounds off our survey of the complete consistent extensions of T_{dis}: There are four extensions whose models are infinite and denumerably many - the theories $T_{dis}(+,+,n)$ - whose models are of cardinality n. These latter

---

[20] N. B. the tuples $<a_1,..,a_n,a>$ and $<b_1,..,b_n,b>$ are not necessarily arranged in order of magnitude, even if this was true for the tuples $<a_1,..,a_n>$ and $<b_1,..,b_n>$, since the new elements a and b But of course we can rearrange the elements of $a_1,..,a_n,a>$ and $<b_1,..,b_n,b>$ so that their order in hte tuples reflects there order in the sense of M_1 and M_2.

theories are absolutely categorical - any two models of $T_{dis}(+,+,n)$ are isomorphic - whereas the first four are complete but not ω-categorical.

Since $\mathbb{T}_{L,T_{dis}}$ is infinite, it follows from Thm. 5 that it is not boolean. The trouble maker is $T_{dis}(+,+,\infty)$. All other complete extensions of T_{dis} are finitely axiomatisable over T_{dis} (and in fact, since T_{dis} is finitely axiomatisable itself, finitely axiomatisable simpliciter). From this and the infinity of $\mathbb{T}_{L,T_{dis}}$ we can conclude that the one remaining complete theory of $\mathbb{T}_{L,T_{dis}}$, viz. $T_{dis}(+,+,\infty)$, is not finitely axiomatisable. (This is a result that we can also easily derive directly, making use of the particular axioms - those of $T_{dis}(+,+)$ together with the difference axioms D_n - which we have given, but we get it from Thm. 5 "for free".

Exercise. Determine which extension of T_{dis} is the complement $-T_{dis}(+,+,\infty)$ of $T_{dis}(+,+,\infty)$ relative to T_{dis}. (In particular, give an explicit axiomatisation for $-T_{dis}(+,+,\infty)$.)

The purpose of this section has been two-fold. On the one hand it is meant as counterpoint to our investigation of the much simpler lattice $\mathbb{T}_{L,T_{den}}$ in Section 2.2.2. As we noted earlier, the lattice $\mathbb{T}_{L,T_{dis}}$ of this section is still of modest complexity when compared with the Tarski lattices for most languages and theories. But it is nevertheless significantly more complex than $\mathbb{T}_{L,T_{den}}$. Crucially, $\mathbb{T}_{L,T_{den}}$ is boolean while $\mathbb{T}_{L,T_{dis}}$ is not.

However, the section also has served a second, more general purpose, that of introducing the method of Quantifier Elimination. The general method is contained in the argument we have given for the inductive step in the proof of (1') from condition (3). This argument is fully general in that it makes no use of any special properties of the models for T_{dis}. To turn that arument into a proof that any two models of T_{dis} are elementarily equivalent we needed in addition (i) a definition of the relations $\equiv_k$ together with (ii) a proof that that the relations thus defined satisfy (3) and (ii) a proof that $\equiv_o$ satisfies condition (2) for quantifier free formulas. In each application of the method of Quantifier Elimination (i)-(iii) have to be dealt with anew, in a way which reflects the special properties of the problem to which it is being applied. But the general architecture is always the same. The next section contains some further general reflections about this method and some remarks about its history.

One final remark on the nature of our investigations in the last three sections (Sections 2.2.1 -2.2.3). On the one hand these invesitgations can be seen as a continuation of the exploration of first order theories of boolean and other lattices which we started in Section 2.1.2. From this pointt of view there is no fundamental difference between our exploration of Tarski and Lindenbaum lattices in the last four sections and, say, our look at the two boolean algebras of Section 2.1.4. But there is also another point of view from which what we have been doing from Section 2.2 onwards is importantly different from what precedes it. In these last sections we have been applying the formal tools of analysis - that of investigating structures as models of first order theories - to the structure of those tools themselves. In other words, here we have one example of the situation described informally in Sections 1.3.2 and 1.3.3 of Ch. 1: the possibility and potential usefulness of applying the tools of formal logic to the structures of formal logic - its expressions, languages and theories - themselves. As announced in Ch. 1 we will another instance of this in Ch. 3 when we develop set theory as a first order theory. While there are many important differences between what we will do in Ch. 3 and the explorations of the last three sectons, they nevertheless have in common that both show the methods of formal logic can be made into their own topic.

2.2.4 Why "Quantifier Elimination"?

N.B. The following section - is mostly of historical interest and can be skipped without any loss to the substance of these Notes.

The term "quantifier elimination" refers originally to a method which it describes perfectly: To show that all sentences A of a given language L have a certain semantic property which involves truth in certain Models or classes of models, show that in relation to the models M in question every sentence A is equivalent to a quantifier-free sentence A', in the sense that for each such model M we have $M \models A$ iff $M \models A'$. In the simplest cases where quantifier elimination is possible in this sense, the quantifier-free formulas A' are formulas of the very language L one starts out with. but very often the method isn't applicable in this simple form. Quantifier-free equivalents for sentences with quantifiers can be found, but only in some extension L´ of L. Typically L' is that where is a *definitional extension of* L in the following sense. Each new non-logical constant α of L' is defined by a formula ϕ_α of L, with as many free variables as α has arguments. Thus, if α is an n-place

predicate, then ϕ_α has the free variables $v_1,..., v_n$. (Function constants present an additional complication, which is not directly relevant here. So we leave them out of consideration. If necessary, n-place function constants can always be "simulated" as n+1-place predicates.) The definitions of the new constants of L' provide us with a way of expanding any model for L to a model for L': If $M = \langle U,I \rangle$ is a model for L, α a new n-place predicate of L' and ϕ_α the definition of α, then the interpretation function I' of the expansion M' of M will assign α the set of all n-tuples $\langle a_1,..., a_n \rangle$ of elements of M such that $M \models \phi_\alpha[a_1,..., a_n]$. This transforms in particular each of the models which determine the notion of equivalence relevant to the given application into corresponding L'-models.

The defining formulas ϕ_α will often contain quantifiers. When this is so, the term "quantifier elimination" for the existence, for each sentence A of L, of a quantifier-free formula in L' is easily somewhat misleading. for by permitting in the "quantifier-free" formula A' of L' that is equivalent to A predicates that are defined by quantified formulas of L we allow quantification to sneak back in as it were, and A' should be considered as "quantifier-free" only in an attenuated sense. In fact, when we translate A' back into L by replacing all occurrences in it of new predicates by their definitions in L, then we will in general get a formula A'' which does contain quantifiers. The point of the method in these cases is that while A'' does contain quantifiers, it contains them only in quite special configurations, and this is what makes it (or, equivalently, the formula A' from which A'' is obtained) behave in ways that are relevantly similar to the behaviour of the quantifier-free formulas of L. In particular - this is the crucial point here - A'' ought to behave much like a quantifier-free formula with regard to the questions of the form: "Does $M \models A''[a_1,.., a_n]$?", where M is one of the relevant L-models and $\langle a_1,..., a_n \rangle$ an n-tuple of elements from M (assuming that the free variables of A'' are among $v_1,.., v_n$). For instance, when the issue is to show that two such models M_1, M_2 are elementarily equivalent, then it should be true that $M_1 \models A''[a_1,.., a_n]$ iff $M_2 \models A''[b_1,.., b_n]$, where $a_1,.., a_n$, $b_1,.., b_n$ are from M_1, M_2, respectively, and $\langle a_1,.., a_n \rangle \equiv_0 \langle b_1,.., b_n \rangle \equiv_0$ is some relation of moderate complexity, and we should be able to prove that.

In fact, the use of quantifier elimination in this sense for the purpose of proving elementary equivalence may involve much more complicated arguments than the one that was needed in the proof above to establish the truth of condition (3).

The method of quantifier elimination in this form becomes particularly involved in those cases where it is not only necessary to extend the language L with which one starts to a larger language L', but where L must be extended with infinitely many new predicates, The definitions of these predicates will necessarily be of increasing complexity, and in particular of increasing quantifier complexity.[21]

About the simplest illustration of quantifier elimination in the literal sense of the term concerns the theory T_{rat}, to which we applied the method of Cantor's proof in Section 2.1. The simplicity of the proof that any two denumerable models of this theory are isomorphic is directly reflected in the ease with which the quantifier elimination method is applied in this instance. In particular, it is not necessary in this case to extend the language {<} of the theory to a larger language.

We begin by considering quantifier-free formulas of L in the variables $v_1, ..., v_n$. We think of these variables as designating points of some dense linear order. Among formulas of this kind there are in particular those which fully describe the order relations between these points, and also say which variables are to be seen as designating the same point. Any formula A of this kind can be written in a form which is the conjunction of three conjunctions A_1, A_2, A_3, which can be dscribed as follows.

(i) A_1 is a conjunction of equations of the form $v_i = v_j$ $(i < j \leq n)$. These give us all combinations of variables v_i, v_j which, according to the situation described by A, designate the same point.

(ii) A_2 is the conjunction of all formulas of the form $v_i \neq v_j$ $(i < j \leq n)$ such that $v_i = v_j$ is not a conjunct of A_1.

(iii) Let $x_1, ..., x_m$ $(m \leq n)$ be all those variables v_j from $\{v_1, ..., v_n\}$ such that A_1 contains no equation of the form $v_i = v_j$. Then A_1 is a conjunction of formulas $x_i < x_j$ which completely fixes a linear order between the x's.

It is easy to see (a) that any such conjunction A is consistent with

---

[21] If L is finite (i.e. has only finitely many non-logical constants), then only finitely many non-equivalent predicates of a given arity can be defined in L if we only consider defining formulas whose quantifier depth does not exceed some given fintite number k.

T_{rat} in that we can find a model M for L and objects a_1, ..., a_n of M such that $M \models T_{rat}$ and $M \models A[a_1, ..., a_n]$, and (b) A is maximal in the sense that if we take any other quantifier-free formula B of L in v_1, ..., v_n, such that B is consistent with $T_{den}(-,-)$, then either $T_{rat} \models (\forall v_1)...(\forall v_n)(A \rightarrow B)$ or $T_{rat} \models (\forall v_1)...(\forall v_n)(A \rightarrow \neg B)$; and, finally, (c) any quantifier-free formula of B L in v_1, ..., v_n that is maximal consistent in the sense above is equivalent to an A of the kind described, i.e. there is an A as described such that $T_{rat} \models (\forall v_1)...(\forall v_n)(A \leftrightarrow B)$.

(a), (b) and (c) together entail that any quantifier-free formula B of L in v_1, ..., v_n which is consistent with T_{rat} is equivalent modulo T to some disjunction $\bigvee_i A_i$ of conjunctions A_i of the described kind:

$$T_{rat} \models (\forall v_1)...(\forall v_n)(B \leftrightarrow \bigvee_i A_i).$$

We can generalise to the case of inconsistent formulae B by stipulating that they are equivalent to some fixed logical contradiction $\perp$, identifying $\perp$ with the "empty disjunction" of formulas.

Now let A be an arbitrary sentence of L in the kind of prenex form used in Section 2.2.3 - i.e. one whose prefix consists of existential quantifiers and negations - and let us assume that the matrix B of A is given as a disjunction $\bigvee_i A_i$ of maximal conjunctions A_i of the kind we have described. Without loss of generality we may assume that the matrix is immediately preceded by an existential quantifier $(\exists v_n)$. (In case the last element of the prefix is a negation sign, this negation can be moved towards the inside of the matrix formula and the resulting formula rewritten once more as a disjunction $\bigvee_i A_i$.) We first observe that $(\exists v_n)(\bigvee_i A_i)$ is logically equivalent to $\bigvee_i (\exists v_n)A_i$. Now consider any one of the disjuncts A_i. Let A'_i be the formula which we obtain from A_i by eliminating from it all conjuncts which contain v_n.
Claim: $T_{rat} \models (\exists v_n)A_i \leftrightarrow A'_i$. First the implication from left to right. This is a theorem of predicate logic. For (i) $\models A_i \rightarrow A'_i$, since in going from A_i to A'_i we have only thrown out conjuncts; (ii) since v_n does not occur in A'_i, (i) entails that A'_i also follows logically from the existential quantification $(\exists v_n)A_i$ of A_i. For the opposite direction we have to distinguish between several cases. First, suppose that v_n occurs in A_i in a conjunct of the form $v_j = v_n$. Then v_n will occur in A_i only in conjuncts that have the form of equations. So in this case, adding these

conjuncts again to A'_i and then quantifying existentially over v_n yields a formula which is entailed by A'_i, and this formula is (obviously equivalent to) $(\exists v_n)A_i$. Second suppose that v_n does not occur in A_i in a conjunct of the form $v_j = v_n$. Then v_n will occur in at least one conjunct involving $<$. There are three cases to be considered here:

(i) v_n occurs only in conjuncts of the form $v_n < v_j$. Then A_i describes v_n as the first element among its "points". In particular, v_n is described as lying before the point which is described by A'_i as the first of the points designated by $v_1, ..., v_{n-1}$. Let v_j be the variable (or one of the variables) designating this first point of the order described by A'_i. Since $T_{rat} \models (\forall v_j)(\exists v_n)\ v_n < v_j$, we also have that $T_{rat} \models A'_i \rightarrow (\exists v_n)A_i$.

(ii) The second possibility is that v_n occurs in A_i both in conjuncts of the form $v_n < v_j$ and in conjuncts of the form $v_j < v_n$. In that case there will be two variables v_j and v_k such that A_i entails that v_j, v_n and v_k are adjacent in the order it describes. This time we make use of the fact that $T_{rat} \models (\forall v_j)(\forall v_k)(v_j < v_k \rightarrow (\exists v_n)(v_j < v_n \ \& \ v_n < v_k))$ to see that $T_{rat} \models A'_i \rightarrow (\exists v_n)A_i$.

(iii) The third case is that where A_i only contains conjuncts of the form $v_j < v_n$. This case is just like case (i).

This competes the argument that

(7) $T_{rat} \models (\exists v_n)A_i \leftrightarrow A'_i$.

(7) entails that when we replace $(\exists v_n)A_i$ by A'_i in A, we obtain a sentence which is equivalent to A, but in which the quantifier $(\exists v_n)$ no longer occurs. In an analogous way we can eliminate all quantifiers of A but one. At this point we have a sentence C equivalent to A modulo T_{rat} which contains one quantifier $(\exists x)$, with or without a negation sign in front of it and some quantifier-free formula D following it in which the only variable is x. It is easy to verify by checking the small number of different forms that D can take that either $T_{rat} \models (\exists x)D$ or $T_{rat} \models \neg (\exists x)D$. Then we also have: $T_{den(-,-)} \models C$ or $T_{rat} \models \neg C$. So in particular we have $T_{rat} \models A$ or $T_{rat} \models \neg A$. This shows the completeness of $T_{den(-,-)}$ and by the same token the fact that modulo it every formula is equivalent to either a theorem of the theory or a

contradiction.

q.e.d.

Evidently this has been a rather fussy proof, with lots of little details that had to be checked along the way, and far lengthier than Cantor's proof of the same result presented in Section 2.1. For more complicated cases, where Cantor's proof can't work, the method outlined is also much fussier than the one we described in connection with the extensions of T_{dis}. Let us briefly look at the case of $T_{dis}(+,-)$ in connection with the present method. This time we must, as indicated above, extend L to a larger language L', and in fact to one with infinitely new predicates. The following 2-place predicates $D\geq_r$ for r = 1, 2, ...will fit the bill. Intuitively, $D\geq_r(x,y)$ says that x lies before y and that there are at least r points between them. It is left to the reader to define these predicates in L. (That is, to find formulas $E_r(x,y)$ of L with x and y as free variables whose extension in any model of $T_{dis}(+,-)$ consists exactly of the pairs <a,b> such that a and b stand in the relation $D\geq_r$. With the help of the predicates $D\geq_r$ we can also define predicates $D=_r$ which say that between x and y there are exactly r points. Evidently $D=_r(x,y)$ holds iff $D\geq_r(x,y)$ & $\neg$ $D\geq_{r+1}(x,y)$. For k = 1,2, ... let L_k be the extension of L with the predicates $D=_r$ for r = 1, ..., 2^k together with the predicate $D\geq_{2k+1}$. Suppose that B is a quantifier-free formula in v_1, ..., v_n of $L_{k'}$ and that $k' \leq k$. Then B is equivalent modulo $T_{dis}(+,-)$ to a disjunction of conjunctions of literals from L_k.

Now let A be a sentence of L and assume that A is in prenex form with a prefix consisting of existential quantifiers and negations. Consider the innermost quantifier $(\exists v_n)$ of A. Rewrite the matrix of A as a disjunction $\bigvee_i A_i$ of maximal consistent formulas of L. Again $(\exists v_n)\bigvee_i A_i$ is logically equivalent to $\bigvee_i (\exists v_n)A_i$. Consider $(\exists v_n)A_i$. A_i is equivalent to a disjunction $\bigvee_j A_{ij}$ of maximal consistent formulas of L_1. Let A'_{ij} be the result of eliminating all conjuncts containing v_n from A_{ij}. It is not hard to see that $T_{dis}(+,-) \models (\exists v_n)A_i \leftrightarrow \bigvee_j A'_{ij}$. So we can replace the part inside A beginning with $(\exists v_n)$ by a quantifier-free formula from L_1 in which v_n no longer occurs and which is equivalent to this part modulo $T_{dis}(+,-)$. In this way we can remove all the quantifiers from A. Note, however, that each time we remove a new quantifier the matrix formula which we remove together with it will belong to one of the languages L_k and the disjunction replacing it will then belong to the next language L_{k+1}. This recursion is the direct counterpart of the one

which in our earlier proof of this result made use of the hierarchy of relations $\{+_k\}$.

Not very nice proofs. But they do explain how our earlier, nicer, method came to its name.

2.3 More about Algebraic Theories

Our only encounter with algebraic languages and theories so far was with the languages and theories of lattice algebras and boolean algebras (L_{lata}, L_{ba}, T_{lata}, T_{ba}; see Sections 2.1.2 and 2.1.3). One of the points we stressed about those structures, all of which are lattices, was that they can be characterised alternatively as algebraic structures, involving a number of operations with certain equationally definable properties, or as structures that involve a partial ordering with special properties. As a matter of fact this kind of duality between an algebraic and a relational conception of structure is quite rare, of which the case of lattices is arguably the most striking example in mathematics and logic as they are known today. For most types of relational structures there seem to exist no algebraic alternatives that provide a significantly different perspective; and, similarly, no significantly different relational formulations seem possible for most algebraic structures that play a prominent part in mathematics.

It should be stressed that these are informal assessments, which it would be hard to turn into hard-nosed formal claims that it would be possible to prove or conclusively refute. For what is it for an alternative characterisation of a type of structure to be 'significantly' different? That seems rather a matter of taste, for which it would be difficult to find a convincing formal definition. And that significance is the crucial notion here follows from the fact that some way of redefining relational structures in algebraic terms is almost trivially possible. And the same holds for, conversely, redefining algebraic structures in relational terms. As regards the redefinition of algebraic structure in relational terms we refer to Exercise EA2 at the end of the Appendix to Ch. 1, where it was shown how each n-place function constant can be replaced by a corresponding n+1-place relation constant together with an axiom stating that the relation denoted by the new constant is functional in its last argument; and further, how each

formula couched in the original functional vocabulary is to be translated into a formula couched in the new relational one.

The converse reformulation is slightly more involved. We know from set theory that the extension of any n-place relation R - i.e. any set of n-toples of objects drawn from some domain U - can be turned into the corresponding characteristic function f_R which maps the tuples belonging to the extension to one of two special objects - the one which intuitively speaking signifies 'yes' - and maps the other n-tuples to the other special object, which intuitively means 'no'. Usually the two objects chosen for this purpose are the numbers 1 and 0, but of course that is not essential for the reduction - any two objects will do, provided that they can be kept suitably distinct from the objects in U. There are various ways in which distinctness can be secured. One of these makes use of a simple technique that has proved useful in formal logic elsewhere too is to extend the universes of the algebraic structures M that are to be redescribed in relational terms with a pair of new objects 1_M and 0_M which serve as the 'yes' and the 'no' in the context of M. Some care has to be taken to make sure that the relational translations of the sentences of the original algebraic language are true in the new extended models $M[0_M, 1_M]$ iff the original sentences were true in the non-extended models M. But these matters are essentially trivial. For details see Exercise **??** of this Chapter.

The types of algebraic structures to which we turn now, groups and semi-groups, conform to what appears to be the rule in that no significantly different relational characterisations of these types seem to exist. They are also typical of algebraic structures more generally in that they can be characterised by axioms all of which have the form of universally quantified equations, just as we found this to be possible in the case of lattices, distributive lattices and boolean algebras. In Universal Algebra - the branch of mathematics which studies algebraic structures from a general and abstract point of view - types of structure (i.e. classes of models) that are defined by sets of such equational axioms are known as *varieties*. It is important to keep the distinction between this notion and the more general one of an axiomatically definable type of algebraic structure firmly in mind. In general axiomatic characterisations of types of algebraic structures may involve axioms that can be any sentences from the first order language for which the structures are models. The equational axiomatisations that make the characterised model class into a variety constitute a comparatively small special subclass from the range of all possible first order axiomatisations. (Note in this connection that equational axioms are (i) purely universal sentences, but in addition (ii) even among the

purely universal sentences they form a specialised subclass.) It seems safe to infer that the class of varieties is a correspondingly small subclass of the class of all axiomatisable structure types.)

We start, mostly as a preamble to our discussion of the Theory of Groups, with a brief introduction to the Theory of Semi-Groups. The notion of a semi-group is simpler and more fundamental than that of a group, although, as the terminology suggests, the notion of a group came first. This is comparable to what can be observed in connection with orderings, where the notion of a linear ordering was well understood before the general notion of a partial ordering was properly articulated and made into the topic of the exploration of a theory - the Theory of Partial Orders - which subsumes the Theory of Linear Orders as one of several specialisations (Lattice Theory being another).

2.3.1 The Theory of Semi-Groups

The language of the theory of semi-groups consists of a single 2-place function constant. We follow the widely accepted convention of denoting this constant as a full stop and of writing the terms involving it in 'infix notation', just as with ordinary multiplication. So the language, L_{sg}, is {.}, and the term we get when applying. to, say, the variables x and y is written as 'x.y'.

The Theory of Semi-groups, T_{sg}, is nothing more or less than the theory of an associative operation. Thus it consists of all consequences of the single axiom ASS:

ASS $\quad (\forall x)(\forall y)(\forall z)\ x.(y.z) = (x.y).z$

Associative operations can be found in all kinds of contexts and they come in a variety of very different forms. Three salient categories are:

(i) 'arithmetical operations like addition and multiplication, as operations on a range of different domains: natural numbers, integers, rational numbers, real numbers, complex numbers.

(ii) fairly closely related to these, set-theoretical union and intersection, and more generally supremum and infimum operations in lattice-like structures.

(iii) 'function application', in the widest sense of the word. In a sense this is just one operation. But it is found in such a wide variety of

contexts that its instances provide a quite diverse spectrum of different semi-groups, both conceptually and as regards their further formal properties.

In (iii) the basic idea is that of a succession of operations which transform objects of a certain sort into other objects of that sort. the objects can be numbers, geometrical figures, linguistic expressions, computer files or documents, .. - in fact, they can be data structures of and kind. And similarly, the operations can be of any kind too, provided that they return objects of the same sort that they take as input. All that is required is that these operations can be carried out in succession, but that is in essence guaranteed by the fact that their outputs are such that they can serve again as inputs to further applications of the operations.

Under these conditions it is possible to form complex operations by combining two operations O_1 and O_2 into a complex operation $O_1.O_2$ which consists in first executing O_1 and then applying O_2 to the output that the first operation produced. That is, for any input x we have $(O_1.O_2)(x) = O_2(O_1(x))$. It should be obvious that the 'second order operation' (= operation on operations), will always be associative: First executing $O_1.O_2$ and then O_3 obviously amounts to the same thing as first executing O_1 and then $O_2.O_3$; in both cases we get a succession of first executing O_1, then executing O_2 and finally executing O_3.

More 'mathematically' the second order operator, can be identified with the operation $\circ$ of function composition: Let U be any set of 1-place functions from an 'object set' X into itself. Then for any two functions f and g from U, we can form the function $f \circ g$ which maps each object x from X to g(f(x)). Evidently this is again a function from X into X. That $\circ$ is associative follows for the obvious reasons spelled out above.

The three types of associative operations listed above are distinguished by additional formal properties. Arithmetical operations are typically commutative, ie. they satisfy the commutativity axiom COM.

COM $\quad (\forall x)(\forall y)\ x.y = y.x$

Function composition, in contrast, is in general not commutative. Consider for instance the functions $f(x) = x + 1$ and $g(x) = 2x$ on the natural numbers. Then $(f \circ g)(1) = g(f(1)) = 2(1+1) = 4$, but $(g \circ f)(1) = f(g(1)) = 2 + 1 = 3$. However, while non-commutativity is the rule for function composition, there do exist (naturally arising) function spaces

on which composition is commutative. An example is the set U of all functions (say, on the natural numbers, but other number sets will do too here) that map each number onto a certain multiple of it. That is, $U = \{\lambda x.nx: n \varepsilon \omega\}$ (where $\lambda x.nx$ is that function f which for any number y as argument returns the number n.y as value). On the other hand, in modern mathematics one studies number systems ('skew number fields') in which addition and/or multiplication are not commutative. So commutativity is a property that *tends* to hold for semi-groups of types (i) and (ii) and not to hold for semi-groups of type (iii), but this is only a matter of tendencies.

A distinction between semi-groups of types (i) and (ii) is that those of the second type typically satisfy the law of idempotency, given as IDP below, while those of type normally do not:

IDP $(\forall x)\ x.x = x$

This does not mean that in semi-groups of the first type there are no elements at all which satisfy the equation x.x = x. More often than not such semi-groups have some element that satisfies the equation. But these elements are, in case they exist at all, rare, and often they are unique. For instance, the additive semi-groups of the natural numbers, the integers and the reals (i.e. the operation of addition on the natural numbers, the integers or the reals, respectively) all have exactly one such element, viz. the number 0. In the multiplicative groups of (among other number systems) the reals and the rationals (i.e. the multiplication operation on the reals and the rationals) there are two such elements, viz. 0 and 1.

That semi-groups of the first kind contain such elements is closely connected with another property that singles out a certain subclass of semi-groups. This is the property of having an *identity*. An identity of semi-group is an element e such that for any element x of the semi-group x.e = e.x = x. In additive groups this is the unique element that satisfies the equation x.x = x, i.e. 0: for any number x, 0 + x = x + 0 = x. (That an identity satisfies x.x = x follows logically from the definition.) In the case of multiplicative semi-groups the identity is not 0 but 1.

The existence of an identity is quite common among semi-groups of each of the three types. Thus among the salient examples of semi-groups of type (ii), structures of the form $\langle U,\cup\rangle$, where U is some set of sets and $\cup$ is set-theoretic union, have an identity iff U contains a bottom element wrt. set-theoretic inclusion, i.e. an element b that is

included in all other elements of U. For then it will be the case for all x in U that $b \cup x = x \cup b = x$. (A common way for this condition to be satisfied is when u contains the empty set $\varnothing$, which will always be the bottom element so long as it is present.)

Among semi-groups of type (iii) the existence of an identity is also a common occurrence. This will be so in particular when the universe U of a given semi-group contains the identity function I_X on the associated object set X, i.e. the function whose domain is X and which maps each x in X to x. Obviously we have for any function f in U that $I_X \circ f = f \circ I_X = f$.

The existence of an identity is our first property of semi-groups that cannot be expressed by means of an equational axiom - evidently so, for we are not dealing with a general condition that all elements of the structure must satisfy, but an existence claim, to the effect that there is at least one element that satisfies a certain equational condition. As stated this has the form of an $\exists\forall$-formula; and indeed, in the language {.} there seems to be no simpler way of stating it. For the sake of explicitness we give the $\exists\forall$ formula:

IDE $\quad (\exists y)(\forall x)(y.x = x \ \& \ x.y = x)$

One might wonder if this formulation isn't somewhat redundant. Do we really need the conjunction of the two equations y.x = x and x.y = x? Wouldn't one of those be enough? The answer to this question is negative. But there are some slight subtleties to the matter, so we will dwell on it a little. Let us, just as we have called an element that satisfies the condition $(\forall x)(y.x = x \ \& \ x.y = x)$ an identity, use the terms *left identity* and *right identity* for elements that satisfy the conditions $(\forall x)y.x = x$ and $(\forall x)\ x.y = x$, respectively; and let us call the statements that a left, resp. right identity exists, IDEL and IDER:

IDEL $\quad (\exists y)(\forall x)\ y.x = x$

IDER $\quad (\exists y)(\forall x)\ x.y = x$

Evidently an identity is both a left identity and a right identity. But we will see in Section ?? that in general a left identity need not be a right identity (and thus not be an identity) and conversely. Nor does the existence of a left identity entail that there is some other element that is a right identity or vice versa. That is, in general neither of IDEL and

IDER entails the other, and so a fortiori neither entails IDE.[22] On the other hand, when a semi-group has both a left identity and a right identity, then these two elements must be identical, and this element will thus be an identity. Similarly, any two left identities and any two right identities must be identical (and so any two identities must be identical). But of course the identity of two left or two right identities doesn't entail that they will be identities.

Exercise. a. Suppose that e_l and e_r are a left an right identity of some semi-group <U,.>. Show that $e_l = e_r$.

b. Suppose that e_1 and e_2 are both left identities of <U,.>. Show that $e_1 = e_2$.

Some semi-groups with an identity are distinguished by a further property, which makes them into *groups*. A *group* is a semi-group with an identity e in which each element x as an *inverse*, i.e. an element z such that x.z = z.x = e. Expressing this property in our language of semi-groups, {.}, is cumbersome, since it must incorporate the assertion that there exists an identity within it.

INV $(\exists y)(\forall x)(y.x = x \ \& \ x.y = x \ \& \ (\forall x)(\exists z)(z.x = y \ \& \ x.z = y))$

Once again the question arises whether we need ther conjunction of the two conditions in the scope of $(\exists z)$. This time the immediate answer is negative. But here too there are subtleties that deserve to be pointed out, and which will emerge in the next section. So once again we distinguish, so that we will be in a better position to discuss those when we come to them, between a l*eft inverse* z_l *of* an element x, which has the property that $z_l.x = e$ and a *right inverse* z_r *of* x, which has the property that $x.z_r = e$.

The answer to the question above is negative in the following precise sense. Suppose that a semi-group M = <U,.> has an identity e. Then if every element of M has a left inverse it is also the case that every element has a right inverse; and conversely, if every element has a right inverse, then every element has a left inverse. Moreover, in either case the left and right inverse of any element will coincide,.

---

22 When we say that (e.g.) IDEL does not 'entail' IDER, what is meant is that IDEL doesn't entail IDER within the Theory of Semi-Groups, T_{sg}. That is, IDER does not follow logically from the conjunction of IDEL and T_{sg}'s only axiom ASS.

Consequently if every element has a left inverse, then every element x has an inverse in the sense of INV (i.e. an element z such that x.z = e & z.x = e)

The proof of these different claims is not complicated. First suppose that every element of M has a left inverse. Let x be any element of M, let z be a left inverse of x, i.e. z.x = e. We must show that x has a rght inverse. Let u be a left inverse of z, i.e. u.z = e. Then x = e.x = (u.z).x = u.(z.x) = u.e = u. But then x.z = u.z = e, so z is right inverse of x. This establishes not only that every element of M has a right inverse, but that for each x there is an element that is both left and right inverse. A parallel argument shows that this conclusion follows equally from the assumption that every element of M has a right inverse.

We can summarise the upshot of this by observing that relative to the Theory of Semi-Groups INV is equivalent to each of the two following sentences INVL and INVR.

INVL $(\exists y)(\forall x)(y.x = x \ \& \ x.y = x \ \& \ (\forall x)(\exists z)\ z.x = y\ )$

INVR $(\exists y)(\forall x)(y.x = x \ \& \ x.y = x \ \& \ (\forall x)(\exists z)\ x.z = y\ )$

In thee next section we look at the Theory of Groups. As we have seen this theory can be axiomatised in the language of semi-groups we have been using in this section (the language L_{sg}, or {.}), e.g. by the axioms ASS and INV. But the second of these is not in equational form, and it seems that it cannot be converted into such a form, or be replaced by one more others of such form that yield the same theorems in conjunction with ASS - at least not when we stick with the language L_{sg}. As we have seen this theory can be axiomatised in the language of semi-groups we have been using in this section (the language L_{sg}. (We are not giving an actual proof that such a replacement is impossible, and as far as we know such a proof this not all that easy.)

However, we will see in the next section that it does become possible to axiomatise the Theory of Groups in equational form if we extend L_{sg} with additional non-logical constants.

2.3.2 The Theory of Groups

We have already given one formulation of the first order theory of groups and thus specified what groups are like. But, as in the case of lattices, there are other ways of formalizing the notion, even if in the

present case the differences aren't quite as dramatic. As we already said, the main advantage of the alternative formulation we present below is that it enables us to state all the axioms as equations. The comparison between this new axiomatisation and the one given in the last section is interesting from a general methodological point of view in that it shows a trade-off of a kind not yet encountered: That between a parsimonious choice of primitive notions (our language {.} with its one 2-place function constant) but axioms of a more complicated structure and on the other hand a richer set of primitives with a corresponding gain in simplicity as far as teh axioms are concerned.

The section serves to focus on two other issues of general significance. The first is the question of independence as applied to axiom systems, or sets of sentences. Usually when we specify a set of axioms as a way of characterising a given formal theory, we try to avoid redundancies: none of the axioms in the set should follow logically from the rest. However, proving that this desideratum has in fact been satisfied can be very tricky. And when there are many axioms, there is a lot of work to be done, since each axiom requires its own independence proof. For the axiomatisations of group theory that are considered in this section this problem is manageable since there are few axioms to deal with. But the independence arguments we wil give for them should provide a clear impression of the general nature of independence proofs and also give a little taste of why such proofs can be difficult.

The third point of general significance that the section seeks to illustrate was already brought up in the last section, when we drew attention to the wide conceptual and formal diversity of semi-groups. This is also true of groups, and here the value of extracting what is common to a great diversity of structures by describing them as models of a single formal theory that covers them all has been of great importance in the history and current practice of pure and applied mathematics.

A fourth point concerns the special properties of 'equations', that is of those purely universal sentences in which the quantifier prefix is followed by a single equation. Equations, in this sense of the word, form a kind of closed subsystem of the set of sentences of a given language L, with their own proof theory and its own special model-theoretic properties. This subsystem is known as Equational Logic. A separate section (Section ??) will be devoted to it.

The axiomatisation of the Theory of Groups we gave in the last section had to resort to axioms that were not of equational form. These axioms

contain existential quantifiers that are needed to express that groups contain entities with special properties: (i) an identity and (ii) for each element x an inverse of x. However, we saw that if such entities exist a all, then they are unique. This means that we can also proceed as follows: We introduce constants in our language to denote these entities and then give axioms stating that the denotations of those constants have the required properties. The constants we need are (i) a o-place function constant e to denote the group identity and (ii) a 1-place function constant $^{-1}$ to denote a function that maps each element to its inverse.

Thus we are led to the language $\{\cdot, ^{-1}, e\}$, to which we will also refer as L_{G1}. $\{\cdot, ^{-1}, e\}$ is the group-theoretic vocabulary that is usually treated as basic in discussions of groups.)

In L_{G1} the Theory of Groups can be axiomatised with the axioms T_{G1}.A1-T_{G1}.A3, which we present both in the standard notation of first order predicate logic and also in the abridged notation of equational logic, in which the universal quantifiers are implicit

T_{G1}.A1	$(\forall x)\ (\forall y)(\forall z)\ \ (x\cdot y)\cdot z = x\cdot(y\cdot z)$	$(x\cdot y)\cdot z = x\cdot(y\cdot z)$
T_{G1}.A2	$(\forall x)\ x\cdot x^{-1} = e$	$x\cdot x^{-1} = e$
T_{G1}.A3	$(\forall x)\ x\cdot e = x$	$x\cdot e = x$

But whether we explicitly write the quantifiers of these axioms or not, they are there, and they are meant as axioms of a theory consisting of all sentences of L_{G1} that logically follow from them, and not just those that are universally quantified equations themselves. We will see this presently when we go through a few simple theorems of this theory and proofs of those from the axioms: some of these theorems do have the form of equations, but not all of them.

The proofs of the equational theorems that follow make use of notation that is familiar from the way arguments in universal algebra are often presented, where all mention of quantifiers is suppressed. (Where both premises and conclusions of an argument are in equational form this is very natural, and hardly needs a justification. Nevertheless, it is an interesting, and as it turns out non-trivial, logical question exactly how this form of derivation relates to standard methods of logical deduction like those discussed in Ch. 1. In Section ??, which is devoted to Equational Logic as an alternative to predicate logic, we will go into this question in detail.)

T_{G1}.T1 $x^{-1} \cdot x = e$

Proof. $x \cdot e = x \cdot (x^{-1} \cdot (x^{-1})^{-1}) = (x \cdot x^{-1}) \cdot (x^{-1})^{-1} = e \cdot (x^{-1})^{-1}$.
Therefore:

$x^{-1} \cdot x = (x^{-1} \cdot x) \cdot e = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (e \cdot (x^{-1})^{-1}) =$

$(x^{-1} \cdot e) \cdot (x^{-1})^{-1} = x^{-1} \cdot (x^{-1})^{-1} = e.$

T_{G1}.T2 $e \cdot x = x$

Proof. $e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x) =_{(TG1.T1)} x \cdot e = x$

T_{G1}.T3 $(x^{-1})^{-1} = x$

Proof. Combine T_{G1}.T2 and the first line of the proof of T_{G1}.T1.

Exercise. Turn the proofs of T_{G1}.T1 - T_{G1}.T3 into predicate logic derivations in the formal sense of the definition on p. 5.

Given what was said about groups in the last section, theorems T_{G1}.T1 and T_{G1}.T2 are a natural complement to axioms T_{G1}.A1 - T_{G1}.A3. In fact, when one looks at these axioms without the hindsight that these theorems provide, the suspicion might easily arise that the axioms are too weak. For T_{G1}.A2 only asserts that x^{-1} ia a right inverse of x, and T_{G1}.A3 only that e is a right identity. Is that enough to guarantee that e is also a left identity and x^{-1} also a left inverse? Theorems T_{G1}.T1 and T_{G1}.T2 tell us that they are. But that this is so has to do with a subtle interaction between T_{G1}.A2 and T_{G1}.A3. We will see in the next section that when one of T_{G1}.A2 and T_{G1}.A3 is changed into its opposite (i.e. T_{G1}.A2 into the axiom which says that e is a left identity), then the axiom system does become too weak.

Exercise. Prove the following theorems of G1 from its axioms:

(i) $(x.y)^{-1} = y^{-1}.x^{-1}$

(ii) $x.y = y.x \leftrightarrow y^{-1}.x.y = x \leftrightarrow y.x.y^{-1} = x \leftrightarrow x.y.x^{-1} = y \leftrightarrow x^{-1}.y.x = y$

(Here "$A \leftrightarrow B \leftrightarrow C \leftrightarrow$.." is used as shorthand for

"(A $\leftrightarrow$ B) & (B $\leftrightarrow$ C) & (C $\leftrightarrow$ .. ")

Exercise. Let "x/y" be short for "$x.y^{-1}$". Show:

(i) $e = x/x$
(ii) $x^{-1} = (x/x)/x$
(iii) $x.y = x/((y/y)/y)$

The next theorems do not have the form of equations:

T_{G1}.T4 $(\forall x)(\forall y)(\forall z)(\ x\cdot y = z \leftrightarrow z\cdot y^{-1} = x)$

Proof. First suppose that $x\cdot y = z$. Then $z\cdot y^{-1} = (x\cdot y)\cdot y^{-1} = x\cdot(y\cdot y^{-1}) = x\cdot e = x$. Conversely, if $z\cdot y^{-1} = x$, then $x\cdot y = (z\cdot y^{-1})\cdot y = z\cdot(y^{-1}\cdot y) = z\cdot e = z$.

T_{G1}.T5 $(\forall x)(\forall y)(\ x\cdot y = e \rightarrow y = x^{-1})$

Proof. Suppose $x\cdot y = e$. Then $x^{-1} = x^{-1}\cdot e = x^{-1}\cdot(x\cdot y) = (x^{-1}\cdot x)\cdot y = e\cdot y = y$.

We have now seen two formalisations of the Theory of Groups, one in the language L_{sg} and involving the axioms ASS and INV, and one in the language L_{G1} and involving the axioms T_{G1}.A1-T_{G1}.A3. The move from L_{sg} to L_{G1} was motivated by the observation that the existence statements made by INV provide to be of elements that turn out to be uniquely characterised by the conditions that IV specifies. This means that we could also have proceeded in the same way as we did when extending the theory of lattices T_{lato} in the language $\{\leq\}$ to the theory in which we have constants to refer to the operations $\cup$ and $\cap$ that T_{lato} enables us to define in terms of $\leq$. That is, we can (i) extend L_{sg} to L_{G1} (as we have done), and (ii) extend the theory $Cl_{L_{sg}}(\{ASS,INV\})$ to a theory in L_{G1} by adding the following two definitions of e and $^{-1}$ as axioms:

(Def.e) $(\forall y)(e = y \leftrightarrow (\forall z)\ z.y = z)$
(Def.$^{-1}$) $(\forall x)(\forall y)(x^{-1} = y \leftrightarrow x.y = e)$

It is not hard to show that this is the same theory as T_{G1}.

Exercise: Prove this.

The difference with the situation we found to obtain in the case of lattices is that this time the converse route is not possible: We cannot formulate the Theory of Groups in the language whose non-logical constants are just the ones that we added when passing from L_{sg} to L_{G1}; no axiomatsaton of the Theory of Groups is possible within the language $\{e,^{-1}\}$.

Exercise: Prove this. (Hint: there is no way to define the two place operation, with the help of just the 0-place function e and the 1-place function $^{-1}$.)

These formalisations of the Theory of Groups are by no means the only ones possible. As a matter of fact, in a strict formal sense the number of possible formalisations of a theory is always infinite; for any one formalisation there twill always be infinitely many alternatives, although as a rule most of these will be uninteresting variants which it is as pointless to present as they are easy to construct. But often genuinely different alternatives exist, which cast a different light on what is being formalised. The alternative formalisation of lattices as orderings and as algebras was a particularly striking example of this. Nothing quite like that compares with it in the case of groups. But there is one alternative that is worth mentioning, at least because it answers a certain formal question that naturally arises in connection with what we have said above about our two axiomatisations in the languages L_{sg} and L_{G1}. The choice between those was presented as a kind of trade-off between (i) having just the single function constant, and (ii) having only axioms in equational form. The alternative that is discussed in the following exercise can be seen as combining the advantages of both. It uses a single 2-place function constant / and it only needs equational axioms. The function / is the 'division operator' of Group Theory, which can be defined in terms of . and $^{-1}$ as: $x/y = x.(y^{-1})$.

Exercise. Give a complete axiomatisation, all axioms of which are equations, of the Theory of Groups in the language {/}, where / is the 2-place operation of group-theoretical division: More precisely, provide equational axioms A/.1,..,A/.n (for some number n) such that the theories T_1 and T_2 defined below are identical.

Definition of T_1 and T_2:

Let T/ = Cl{/}({A/.1,..,A/.n}). Let L' be the language {/, . ,$^{-1}$,e}.

(a) T_1 is the theory of L' that is obtained by adding to the axioms of T/ the following definitions of e, $^{-1}$ and . in terms of /:

(i) $(\forall x)\ e = x/x$
(ii) $(\forall x)\ x^{-1} = (x/x)/x$
(iii) $(\forall x)(\forall y)\ x.y = x/((y/y)/y)$

(b) T_2 is the theory of L' that is obtained by adding to the axioms of T_{G1} the following definition of / in terms of . and $^{-1}$:

(iv) $(\forall x)(\forall y)\ x/y = x.y^{-1}$

(Solution. One solution is the following set of axioms A/.1,.,A/.4:

A/.1	y/y = x/x	
A/.2	y/(y/y) = y	y/e = y
A/.3	(y/y)/(x/y) = y/x = x/x	e/(x/y) = y/x
A/.4	x/(y/z) = (x/((z/z)/z))/y	x/(y/z) = (x/(e/z))/y

In the formulations of A/.2-A/.4 on the right, subterms of the form α/α have been abbreviated as 'e', in accordance with A/.1.)

2.3.3 Independence

In the introduction to this section we mentioned the question of the *independence* of the members of a given axiom set. As indicated, it is generally considered a matter of logical hygiene that the sets of axioms used to formalise a given structure or concept contain no *redundant* axioms. That is, if G is any such set and A ε G, then it should not be the case that (G\{A} $\models$ A. If this is not the case, then we say that A is *independent in* G; and if all members of G are independent, G is called an *independent set* of axioms.

As a matter of fact, all axiom sets presented so far in this chapter have been independent in the sense just defined. Showing that this is so, however, is not trivial. In general, proving that an axiom set is independent tends to be not only a fair bit of work - to show that the set A1, ..., An is independent requires n separate proofs, one for each Ai - some independence questions can be a real challenge. Also independence proofs may provide real insight into what precisely is

contributed by a given axiom to the given characterisation of the intended class cof structures that is not contributed by the other axioms. More about this towards the end of this section.

Here we consider only two of the three independence questions connected with the axiom set $\{T_{G1}.A1, T_{G1}.A2, T_{G1}.A3\}$. We show the independence of $T_{G1}.A3$ from the remaining two axioms explicitly, and provide a hint for establishing the independence of $T_{G1}.A2$. As regards $T_{G1}.A1$, the reader is on his own (see Exercise **??**).

First $T_{G1}.A3$. Consider the following model $M = \langle U,F \rangle$ for L_{G1}:

(i) U = the set of all pairs $\langle i,n \rangle$, where $i \in \mathbb{Z}$ (the set of integers) and $n \in \mathbb{N}$ (the set of natural numbers).

(ii) $F(\cdot)$ = the function f such that for any $\langle i,n \rangle, \langle j,m \rangle \in U$, $f(\langle i,n \rangle, \langle j,m \rangle) = \langle i+j,m \rangle$

(iii) $F(e) = \langle 0,0 \rangle$

(iv) $F(^{-1})$ = the function g such that for any $\langle i,n \rangle \in U$, $g(\langle i,n \rangle) = \langle -i,0 \rangle$

Then it is straightforward to verify that $T_{G1}.A1$ and $T_{G1}.A2$ hold in M. But $T_{G1}.A3$ does not hold, since e.g. $\langle 1,1 \rangle \cdot e = \langle 1,1 \rangle \cdot \langle 0,0 \rangle = \langle 1,0) \neq \langle 1,1 \rangle$.

It is easy to turn this construction into a demonstration that the second axiom is independent of the other two by changing the definition of $F(\cdot)$ into

(ii') $F'(\cdot)$ = the function f' such that for any $\langle i,n \rangle, \langle j,m \rangle \in U$, $f'(\langle i,n \rangle, \langle j,m \rangle) = \langle i+j,n \rangle$

It is worth noting that while M falsifies $T_{G1}.A3$ it verifies the superficially similar sentence

$T_{G1}.A3'$ $\quad (\forall x)\ e \cdot x = x$

Recall that $T_{G1}.A3'$ is nothing other than $T_{G1}.T2$. So we have also shown that $T_{G1}.A3$ cannot be derived from $T_{G1}.A1$, $T_{G1}.A2$ and $_{G1}.A3'$. Apparently, then, this sentence is, given $T_{G1}.A1$ and $T_{G1}.A2$, genuinely weaker than $T_{G1}.A3$, and replacing $T_{G1}.A3$ by $T_{G1}.A3'$ in the axiomatisation of T_{G1} would yield a different, weaker theory. In the same vein it can be observed that the modified model $M' = \langle U,F' \rangle$ verifies the sentence

$T_{G1}.A2'$ $(\forall x)\ x^{-1} \cdot x = e$

So replacing $T_{G1}.A2$ by $T_{G1}.A2'$ while leaving $T_{G1}.A1$ and $T_{G1}.A3$ the same would also lead to a weakening of deductive power. On the other hand it is easy to verify that if we replace both $T_{G1}.A2$ and $T_{G1}.A3$ by $T_{G1}.A2'$ and $T_{G1}.A3'$ the result is a theory that is equivalent to T_{G1}.

Exercise: Show this.

Exercise: Show that the associativity axiom $T_{G1}.A1$ is independent of the axioms $T_{G1}.A2$ and $T_{G1}.A3$.

Hint: 1. Consider the model M = <U,F>, where U = the set of the rational numbers without 0 and let $F(\cdot)(r,s) = r/s$. Then $T_{G1}.A1$ evidently fails. Choose $F(^{-1})$ and F(e) so that M verifies $T_{G1}.A2$ and $T_{G1}.A3$.

Other solution. Here is another possibility. U is the set {0,1,2, ..,n-1}. F(e) = 0, $F(^{-1})(k)$ is the unique number m from U such that k + m = 0(mod n) and F(.) is defined as follows: (i) F(.)(k,k) = k; (ii) if $k \neq m$, then F(.)(k,m) = k + m (mod n). Then it is easily verified that (writing "." instead of "F(.)" and using infix notation) 0.k = k.0 = k and that $k.k^{-1} = k^{-1},k = 0$. But in general F(.) will not be associative. For instance, if n = 4, then (2.2).3 = 2.3 = 5 (mod 4) = 1, but 2.(2.3) = 2. (5 (mod 4)) = 2.1 = 3. Note that in this example F(.) is commutative and that (because of this) not only the axioms $T_{G1}.A2$ and $T_{G1}.A3$ are verified, but also the formulas which we get by switching the arguments of the left hand term around, i.e. $(\forall x)$ e.x = x and $(\forall x)\ x \cdot x^{-1}.x = e$.

[End Exercise]

The three independence arguments presented here are comparatively simple. They do give insight why each of the three axioms contributes something that the others do not, but precisely because models that satisfy all but one of the axioms are comparatively easy to find, the insight gained from any one such models (and thus from the independence proof it provides) are limited: Other models might give additional insights in the contributions of the different axioms in the set and quite possibly more important ones.

But in this regard our examples are not representative. In the history of mathematics and logic certain independence questions have had an enormous impact. Their solution have led to the discovery of structures that have proved of lasting importance and to methods of mathematical reasoning and mathematical construction that subsequently found many additional applications. Even some attempts at finding a solution to an independence question that did not answer the question that they were meant to have led to significant progress in other areas.

Perhaps the most famous example from mathematics id the parallel postulate from Euclid's axiomatisation of plane geometry, the statement that for every point p that is not on a straight line l there is exactly one straight line m that goes through p and is parallel to l. Ever since Euclid it was felt that this postulate was less self-evident than Euclid's other postulates. Since it was widely thought that Euclidean geometry described a structure that was in some sense necessary - space just couldn't have been different from what it is! - and since it was thought also that since the properties of the structure of space were necessary, they should be directly accessible to intellectual judgement, the lacking self-evidence of the parallel postulate was seen as an imperfection of Euclid's system, and an imperfection that could be removed only by either finding a more intuitive replacement for it or - even better - to derive it from Euclid's other postulates. In the course of the many centuries during which this was an open question an enormous amount of mathematical energy and ingenuity must have gone into the project of deriving the parallel postulate from the other postulates. Eventually, in the second half of the 18-th century it dawned on some mathematicians that the persistent failure to find a proof of the parallel postulate from the others might have a very simple explanation, viz that there is no such proof, in other words, that the parallel postulate was independent from the other postulates. This led to the new and contrary effort to demonstrate the independence of the parallel postulate, or, what comes to the same thing, the consistency of the other postulates with the negation of the parallel postulate. (It no longer needs to be said here that being a model in which postulates A_1, ..., A_{n-1} hold and A_n doesn't is the same as being a model in which A_1, ..., A_{n-1} and $\neg A_n$ hold together.) The models of the negation of the parallel postulate jointly with the other Euclidean postulates - as described in the work of the Hungarian mathematician Janos Bolyai (1802-1860), the Russian mathematician Lobachewski (1792-1856) and the German mathematicians Gauss (1777-1855) and Riemann (1826-1866) - have done more than anything else to revolutionarise geometry as mathematical discipline as it in the course of the 19-th century. And it has also deeply affected our understanding of the distinction between

necessary and contingent truth as well as the distinction between geometry as a conceptual structure (along the lines it was seen by, for instance, Kant) and geometry as part of the structure of the physical world.[23]

A second independence problem, which was specific to the development of mathematical logic in the 20-th century, concerns the Continuum Hypothesis in Set Theory, the Hypothesis that there are no sets whose cardinality is intermediate between that of the natural numbers (the smallest infinite cardinality) and that of the set of real numbers, which is the same as that of the power set of the set of natural numbers). As we noted earlier, the Continuum Hypothesis was formulated by Cantor, the founder of set theory. Cantor is said to have worked desperately on a proof of the Continuum Hypothesis from other set-theoretical principles, whose validity he did not consider in doubt, and the effort is supposed to have seriously affected his health. His unsuccessful efforts were followed by those of many others, and among these efforts were in particular those to derive the Continuum Hypothesis from the other established set-theoretical axioms e.g. those of Zermelo-Fraenkel (see Ch. 3). But in this case too eventually the suspicion arose that no such derivation could be given, since the Continuum Hypothesis was in fact independent from the other, uncontroversial, axioms of set theory. And independence was finally proved in 1963 by the American mathematician Paul Cohen. In this case too the method used to establish independence has proved immensely fruitful, leading in particular to a series of further independence results within the realm of set theory.

There is an interesting similarity between these tow cases - the parallel postulate in geometry and the Continuum Hypothesis in set theory - in that in both cases a conception of the subject matter as involving necessary and therefore presumably ultimately self-evident truths drove scholars to persistent efforts to decide what seemed not self-

---

[23] The first to have clearly understood this second disrtinction appears to have been Gauss , who engaged as early as the first half of the nineteenth century in a large scale prohect of geodetical measurments in order to determine whether the physical geometry whose straight lines are the paths of light rays is in fact Euclidean or not. (i.e. if light rays conform to the parallel postulate.) Gauss suspician of non-Euclidenan character of the genoatryx of light rays was confirmed only when in the forst quarter of the 20-th century physisicts and astronomers looked for a experimental confirmation of one of the implications of Einstein's gedneral Theory of Relativity, which is that gravitation 'bends' the paths of light rays, so that the geometry they define is - in the presence of graviational fields, which is always the case in our actual cosmos - non-Eucildean. Einstein's Theory of General Relativity, it has been sad would not have been possible without the work of Riemann.

evident on the bassi of those principles that were considered self-evident. One of the general lessons that has been learned from both efforts is that the line between necessity and contingency is much more difficult to draw than people seem to have realised through most of the history of philosophy (and than, in the depth of our hearts, many of us would still like to believe today); and, connected with that, that we should not set too much store by our intuitions on what is 'self-evident' and what is not.

2.3.4 The Theory of Groups and Group Theory

1. What has been called the (first order) Theory of Groups here should not be confused with what is normally understood by 'Group Theory'. First, the 'mini-theorems' of the Theory of Groups of which we have given a few examples here bear no comparison with the theorems about groups that mathematicians find interesting. But more fundamentally, those results can as a rule not even be stated within the first order languages we have been using. For instance, many results in Group Theory have to do with characterisations of groups in terms of the kinds of subgroups they have - that is, in our terminology, in terms of their submodels. (Note that a submodel of a structure that satisfies the axioms T_{G1}.A1-T_{G1}.A3 will automatically be itself a model of these axioms and thus again a group. (Exercise: Prove this and/or Section ?? below.) To state such a characterisations of a group we need to quantify over its subgroups and thus over subsets of its universe, and to do that we need second, not first order logic. So at a minimum we will need the second order extension of one of our first order languages {.} or {,.,$^{-1}$,e}. Also, there are many theorems of Group Theory which involve reference to natural numbers (e.g. to describe the possible size(s) of finite groups with certain properties, and/or the sizes of certain parts of them. The proofs of such theorems often make use of quite complicated facts of combinatorial number theory,. In these cases formalisation requires a logical vocabulary that includes number-theoretic notions as well as the group-theoretic ones that are the only non-logical constants of the language we have used here, and for a formalisation of the proofs of these statements we will need an axiomatisation of number theory as well.

All this goes to say that Group Theory as it is practiced by algebraists involves far more than our 'bare bones' languages provide. Even if such a language suffices to characterise the general notion of a group, it falls far short of what is needed to state and proof what a mathematician wants to know. This is a somewhat sobering comment on the power of

first order formalisations, not only of the structures that are the subject of group Theory, but of most kinds of mathematically interesting structures generally.

2. It was pointed out more than once in this Chapter that the point of many algebraic theories is that their models cover a wide range of different structures. This is true in particular of the theory of groups. The class of all groups shows a great deal of diversity, in the sense that it contains structures which vary substantially either in their conception or in their formal properties or both.

The value of an algebraic theory with such coverage is, we have noted, that the theorems that can be derived from the general theory are applicable to all the different structures that are among its models. This is as true of the Theory of Groups as it is of other theories with wide structure coverage. But on the other hand the diversity among the different types of groups is such, and certain types of groups are so important, that these types have become the subject of a separate branch of mathematical investigation. A prominent example of this is the class of *Abelian* or *commutative* groups, in which the group operation $\cdot$ is commutative (i.e. $x \cdot y = y \cdot x$ holds for all elements x,y of the group).

In this particular case the additional property that singles out the given class of groups, viz. commutativity of $\cdot$, can be expressed by a first order axiom. But for many other properties that define important subtypes of groups this is not so. An example is the notion of a *simple group*, i.e. group that doesn't contain any proper subgroups (i.e. for which there are no properly included submodels which consist of more than one element); the notion of a finite group - finiteness cannot be expressed by a first order axiom -; or the class of all *permutation* groups, a notion which will be explained below.

To give an impression of how different certain models of the Theory of Groups can be from each other in origin and/or appearance we remind the reader of the two types of examples that were mentioned briefly in the introduction to Section 2.2.1. The first type, it may be recalled, consists of structures in which the group operation $\cdot$ is one of the familiar arithmetical operations of addition or multiplication, or some variant thereof. One example of this type of groups are: the integers with the binary operation of addition, the 1-place operation of sign inversion (i.e. n^{-1} is the number -n) and the number 0 as e constitute a group. Similar examples are provided by the rational numbers and the

real numbers, each with the same operations of addition, sign inversion and 0. Closely related examples are the *additive groups modulo n*, consisting of the numbers {0,1,..., n} with "+ mod(ulo) n" for the operation $\cdot$ (where i+j (mod n) is the remainder of i+j after division by n), "sign inversion modulo n" for the operation $^{-1}$ (i.e. $i^{-1} = n - i$) and again 0 as e. Besides these additive groups there are also multiplicative groups, in which $\cdot$ is multiplication. One example we have already encountered: the rational numbers without 0, with multiplication for $\cdot$, $1/r$ for r^{-1}, and 1 for e. Yet another example is provided by the real numbers (also without 0) with the usual operations of times, multiplicative inverse and 1. There are many more examples of this general sort, involving either some variant of addition or multiplication and/or the use of some alternative notion of "number" (complex numbers, quaternions, etc.).

As a rule groups of this type are commutative, since operations of addition and multiplication tend to be commutative (though there are exceptions).

The second type of group to be mentioned here is that where the elements of the group are functions, $\cdot$ is the operation of function composition, $^{-1}$ is function inverse and e is the identity map. In order that these notions are defined for all elements of the structure it is necessary that all elements (i.e. all functions) have one and the same domain and range. Moreover, the requirement that the inverse operation be everywhere defined entails that all functions are injections. Thus a group of this kind will consist of a set of bijections from some given set X to itself. Such bijections from X to X are also known as *permutations of* X.

It is easy to verify that any set of permutations from X to X which includes the identity map on X and is closed under inverses and function composition forms a group. (Exercise: Show this.) Such groups are called *permutation groups*. Within the class of permutation groups we still find a remarkable spectrum of variety. Among the simplest examples are those groups which consist of all permutations of some finite set $\{a_1,..., a_n\}$. Evidently, the properties of any such group are determined entirely by the cardinality of the set - the group of all permutations of $\{a_1,..., a_n\}$ and the group of all permutations of $\{b_1,..., b_m\}$ are isomorphic iff n = m. So it is possible to confine attention to the full permutation groups of {1,..., n} for the different natural numbers n.

Function composition is usually not a commutative operation. So, contrary to the groups based on arithmetical operations permutation groups are hardly ever commutative.

Exercise. i. Show this, by defining a permutation group in which the commutativity law $x \cdot y = y \cdot x$ is invalid.

ii. What is the smallest number n such that the full permutation group on {1,..., n} is not commutative?

[To be added to the list of exercises at the end pf Ch. 2]

Exercise: In Section 2.2.1.1 it was shown that the axiom $T_{G1}.A1$ is independent of the axioms $T_{G1}.A2$ and $T_{G1}.A3$. The model discussed in that exercise did not establish the following stronger independence result, according to which $T_{G1}.A1$ is not entailed by the set consisting of $T_{G1}.A2$ and $T_{G1}.A3$ and their "converses" $T_{G1}.A2'$ and $T_{G1}.A3'$:

$T_{G1}.A2'$ $\quad x^{-1} \cdot x = e$

$T_{G1}.A3'$ $\quad e \cdot x = x$

One way to get this stronger result is to make use of permutation models. Let M = <U,F>, where U is the set of permutations of the set {1,2, ..., n}, for some n > 2. $F(^{-1})$ and F(e) are defined for permutation groups, i.e. $F(^{-1})(f)$ is the inverse f^{-1} of f and F(e) is the identity map.

But we now define $F(\cdot)$ by: $F(\cdot)(f,g) = g^{-1} \circ f$. Show that in this model $T_{G1}.A2$, $T_{G1}.A3$, $T_{G1}.A2'$ and $T_{G1}.A3'$ all hold, but that $T_{G1}.A1$ fails.

Exercise: Missing from the independence proof for the axiom set {$T_{G1}.A1$, $T_{G1}.A2$, $T_{G1}.A3$} in Section 2.2.1.1 was the independence of $T_{G1}.A2$.

To show independence of this axiom from the other two is very easy, because it is the only axiom that contains the opersation $^{-1}$.

a. Why? Prove the independence of $T_{G1}.A2$.

More interesting is the independence from $T_{G1}.A1$ and $T_{G1}.A3$ of the weaker principles (i) that there is for each x an element y such that

$x \cdot y = e$ and (ii) that, for ay $x_,$, any two elements y and y' such that $x \cdot y = e$ and $x \cdot y = e$ are identical:

(i) $(\forall x)(\exists y)\ x \cdot y = e$

(ii) $(\forall x)(\forall y)(\forall y')(x \cdot y = e\ \&\ x \cdot y' = e \rightarrow y = y')$

b. Prove the independence of (i) and of (ii) from T_{G1}.A1 and T_{G1}.A3.

2.4 Equational Logic.

Equations - purely universal sentences whose matrices are of the form $\sigma = \tau$, where σ and τ are terms - have special properties. First, they allow for a special method of deduction: if an equation B follows from equations A_1, ..., A_n, then this can be shown by deriving B from A_1, ..., A_n via special rules, which are designed to fit the special form that equations have.

Secondly, equations are characterised by special model-theoretic properties. These of course include the properties that are shared by all purely universal sentences (see Ch. 1, Sn 1.5.2). But equations are distinguished from universal sentences in general by some additional properties. As for purely universal sentences in general this fact can be cast in the mould of a preservation theorem, a theorem first stated and proved by the American algebraist G. Birkhoff.

These then are the topics of this section. We will first present the special deduction system for equations and prove its soundness and completeness, and then present and prove Birkhoff's Theorem.

Let $L = \{f_1, ..., f_k\}$ be an algebraic language, where, for i = 1, ..., k, f_i is an n(i)-ary function constant. By an *identity of* L we understand any purely universal sentence of the form $(\forall x_1)...(\forall x_m)\ s = t$, where s and t are terms of L and $x_1,...,x_m$ are the variables that have occurrences in at least one of s and t. We denote the identity $(\forall x_1)...(\forall x_m)\ s = t$ also as "$s \equiv t$".

There is a sense in which the identities of L form a "self-contained" subsystem of the set of all formulae of L:

Suppose $\Gamma \models E$, where E is an identity and Γ is set of identities. Then it is possible to derive E from Γ by means of a set of five inference rules $RE_{ref.}$,..., $RE_{repl.}$, each of which only involves identities. That is, there always exists in such a case a derivation of E from Γ which consists of identities only (and in which each line is either a premise from Γ or comes from earlier lines by application of one of the rules).

Here are the rules:

$RE_{refl.}$ $t = t$ (that is: each identity of the form "t = t", where t is any term, may be written down as a new line; thus this rule functions as an axiom.)

$RE_{sym.}$
$$\frac{s = t}{t = s}$$

$RE_{trans.}$
$$\frac{r = s,\ s = t}{r = t}$$

$RE_{subst.}$ Suppose that x_1 ,..., x_m are the free variables occurring in the identity s = t and that r_1,..., r_m are terms. Let s' be the result of simultaneously substituting the terms r_1,..., r_m for the variables x_1 ,..., x_m in s; and likewise for t' and t. Then

$$\frac{s = t}{s' = t'}$$

N.B. This rule also covers the case where we substitute terms for only some of the free variables in s = t (and in particular the case where we do this for only one variable). In such cases we choose for each variable x_i that we want to "leave alone" that variable itself as term r_i.

$RE_{repl.}$ Suppose that s has an occurrence as a subterm in t and that t' results from t by replacing this occurrence of s in t by the term s'. Then

$$\frac{s = s'}{t = t'}$$

A *EL Derivation (Equational Logic derivation) from* a set of equations Γ *in* an algebraic language L is a sequence $<E_1, ..., E_p>$ of identities of L in which each line E_i either (i) is a member of Γ, or (ii) results from an application of $RE_{refl.}$, or (iii) comes from one or more earlier lines by an application of one of the rules RE_{sym} - $RE_{repl.}$.

Exercise. Show that the proofs of $T_{G1}.T1$ - $T_{G1}.T3$ from T_{G1} can be turned into derivations of Equational Logic.

Theorem 12 (Completeness Theorem for Equational Logic).

Suppose that L is an algebraic language and that $\Gamma \models E$, where E is an identity of L and Γ is set of identities of L. Then $\Gamma \vdash_{eq} E$
(That is, there is a derivation in Equational Logic of E from Γ in L.)

Proof: As in the completeness proof for the first order predicate logic we proceed by contraposition. Suppose that it is not the case that $\Gamma \vdash_{eq} s_o = t_o$. We construct a model M such that $M \models \Gamma$ but not $M \models s_o = t_o$. (Recall in this connection that the identities are really universally quantified formulas. Thus $M \models \gamma$ means that for all possible value assignments **a** to the variables of γ $[[\gamma]]_{M,\mathbf{a}} = 1$. On the other hand, in order to show that not $M \models s_o = t_o$ it suffices to find one assignment **b** such that $[[s_o = t_o]]_{M,\mathbf{b}} \neq 1$.)

Informally, we proceed as follows: We identify all terms s, t for which the identity s = t is derivable from Γ. The (equivalence) classes [s], [t], ... obtained in this way will be the elements of the universe of M. We can then define on this universe the interpretations of the function constants of L so that the identities in Γ are all universally satisfied in M. Since it is not the case that $M \models s_o = t_o$, s_o and t_o will not belong to the same equivalence class; hence if **b** assigns to each of the free variables of $s_o = t_o$ its own equivalence class, then $[[s_o]]_{M,\mathbf{b}} \neq [[t_o]]_{M,\mathbf{b}}$.

Formally: Let the relation $\sim_\Gamma$ on the terms of L be defined by:

(1) $s \sim_\Gamma t$ iff $\Gamma \vdash_{eq} s = t$.

Because of the rules $RE_{refl.}$, $RE_{sym.}$ and $RE_{trans.}$ $\sim_\Gamma$ is an equivalence relation. So we can form the corresponding equivalence classes $[t]_{\sim_\Gamma}$. Let $U_M = \{[s]_{\sim_\Gamma}: s \text{ a term of L}\}$. Furthermore, in virtue of $RE_{repl.}$, $\sim_\Gamma$ is a *congruence relation with respect to* each function constant f^n of L, that is:

(2) when for $i = 1, .., n$, $s_i \sim_\Gamma t_i$, then $f(s_1, ..., s_n) \sim_\Gamma f(t_1, ..., t_n)$.

This means that the following definition of the interpretation f_M of f^n in M is coherent and defines a total function on U_M:

(3) $\quad <[t_1]_{\sim_\Gamma}, .., [t_n]_{\sim_\Gamma}, [t]_{\sim_\Gamma}> \varepsilon f_M$ iff $\Gamma \vdash_{eq} f(_1, .., t_n) = t$

(As regards totality of f_M: EQ1 guarantees that there is at least one term t such that "$f(t_1, .., t_n) = t$" is derivable, viz. $f(_1, .., t_n)$.)

This completes the definition of M. To show that M is a countermodel to the claim that $\Gamma \models s = t$ we first establish the following:

(4) Let r be any term of L with variables $x_1,.., x_n$ and let **a** be an assignment in M such that for $j = 1, ..., n$, $\mathbf{a}(x_j) = [x_j]_{\sim_\Gamma}$. Then $[[r]]_{M,\mathbf{a}} = [r]_{\sim_\Gamma}$.

(4) is proved by a simple induction on the complexity of r.

We now show that for each $E_i \varepsilon \Gamma$, E_i is true in M. Suppose that E_i is the equation $s_i = t_i$. Recall that "equations" are really *sentences*, which are obtained from the bare equations by universally quantifying over all the variables occurring in them. So in order that the equation $s_i = t_i$ is true in M it is necessary and sufficient to show that for arbitrary assignments **a** in M, $[s_i = t_i]_{M,\mathbf{a}} = 1$.

Assume that $x_1,.., x_n$ are the variables occurring in $s_i = t_i$. Let **a** be any assignment in M. Suppose that for $j = 1, ..., r$, $\mathbf{a}(x_j) = [r_j]_{\sim_\Gamma}$. Let s_i' be the term $s_i[r_1/x_1,..., r_n/x_n]$ - i.e. s' is the result of simultaneously substituting the terms r_j for the variables x_j in s - and similarly for t_i and t_i'.

Since $s_i = t_i \varepsilon \Gamma$, we have, trivially, $\Gamma \vdash_{eq} s_i = t_i$. So, by the rule $RE_{subst.}$ it follows that we also have $\Gamma \vdash_{eq} s_i' = t_i'$. So

(5) $[s_i']_{\sim\Gamma} = [t_i']_{\sim\Gamma}$.

Let $y_1,.., y_m$ be all the variables occurring in $s_i' = t_i'$ and let $\mathbf{a}'$ be an assignment such that for h = 1, ..., m, $\mathbf{a}'(y_h]) = [y_h]_{\sim\Gamma}$. From Lemma 3, established in connection with the Completeness Proof for Predicate Logic in Ch. I, we know that:

(6) $[[s_i']]_{M,a'} = [[s_i]]_{M,\mathbf{a}''}$,

where $\mathbf{a}'' = \mathbf{a}'[\ [[r_1]]_{M,\mathbf{a}'}/x_1,..., [[r_n]]_{M,\mathbf{a}'}/x_n\ ]$.

By (4) we get (i)

(7) $[[s_i']]_{M,\mathbf{a}'} = [s_i']_{\sim\Gamma}$ and $[[t_i]]_{M,\mathbf{a}'} = [t_i']_{\sim\Gamma}$.

and (ii)

(8) $[[r_j]]_{M,\mathbf{a}'} = [r_j]_{\sim\Gamma}$, for j = 1,..., n.

From (8) it follows that $\mathbf{a}'' = \mathbf{a}'[\ [r_1]_{\sim\Gamma}/x_1,..., [r_n]_{\sim\Gamma}/x_n\ ]$. Thus $\mathbf{a}''$ and $\mathbf{a}$ coincide on the variables $x_1,.., x_n$. Therefore, since $x_1,.., x_n$ are all the (free) variables of $s_i = t_i$, it follows by Lemma 1 from Part I that

(9) $[[s_i]]_{M,\mathbf{a}} = [s_i]]_{M,\mathbf{a}''}$ and $[[t_i]]_{M,\mathbf{a}} = [t_i]]_{M,\mathbf{a}''}$.

From (5), (6) and (9) we get:

$[s_i]]_{M,\mathbf{a}} = [s_i]]_{M,\mathbf{a}''} = [[s_i']]_{M,\mathbf{a}'} = [s_i']_{\sim\Gamma} = = [t_i']_{\sim\Gamma} = [[t_i']]_{M,\mathbf{a}'} = [t_i]]_{M,\mathbf{a}''} = [t_i]]_{M,\mathbf{a}}$.

This establishes that $M \models \Gamma$.

To see that not $M \models s = t$, it suffices to note that it follows from (4) above that $[[s]]_{M,\mathbf{b}} \neq [[t]]_{M,\mathbf{b}}$, where $\mathbf{b}$ is an assignment such that $\mathbf{b}(w_i) = [w_i]_{\sim\Gamma}$, for i = 1,..., h, where w_1, ..., w_h are all the variables occurring in s = t. The existence of such assignments entails that the equational sentence s = t is false in M.

q.e.d.

It is striking how much simpler this proof is than the Completeness Proof we gave in Part I. In a way this should not come as a surprise since we are dealing with formulas of a comparatively simple logical structure. Still, it is to be notes that while the present result is weaker than the full completeness proof precisely in that it deals with a small subclass of formulas, it is stronger in that it shows that when G and E stand in the consequence relation then a proof can be found of a very special and simple form. The following Corollary makes this a little more explicit.

Corollary. If L is an algebraic language and $\Gamma \vdash E$, where, as above, E is an identity of L and Γ is set of equations of L and $\vdash$ is the proof relation of full first order logic, then $\Gamma \vdash_{eq} E$.

This Corollary follows immediately from the Theorem and the soundness of the proof relation $\vdash$. The result is interesting in its own right insofar as it gives a certain normal form for proofs whose premises and conclusion all have the simple form of a universally quantified equation.

(To turn a derivation within Equational Logic into a "simple" proof of the universal generalisation of the conclusion from the universal generalisations of the premises is not completely trivial but very nearly so. In particular, a little reflection makes clear that one can turn the proof into (i) a series of applications of UI to the needed premises and to the identiy axioms; (ii) a series of steps involving MP corresponding to the successive steps of the given Equational Logic proof; and (iii) UG on the variables of the conclusion.)

Note also that the present proof yields like the completeness proof we presented for the full predicate calculus the additional information that a countermodel never need be more than denumerable in size. From the proof we have just gone through this follows from the fact that for any of the languages L we consider in this script the set of terms is denumerable. So a model whose universe consists of equivalence classes of such terms can be at most denumerable.

It should be noted, though, that in the case of equational logic the counter models constructed in the completeness proof as we have presented it here are almost always denumerably infinite. The reason is simple and relates to equations of the form $v_i = v_j$, with variables on both sides of =. If any such equation is entailed by a given set of equations Γ, then this will be true for all of them. For it is easy to see

that any one entails any other. So we have only two possibilities as regards such equations: (i) for all i,j such that $i \neq j$, $[v_i]_{\sim_\Gamma} \neq [v_j]_{\sim_\Gamma}$, in which case the model $M_{\sim_\Gamma}$ will be infinite; or (ii) for some i,j such that $i \neq j$, $[v_i]_{\sim_\Gamma} = [v_j]_{\sim_\Gamma}$, in which case we have $[s]_{\sim_\Gamma} = [t]_{\sim_\Gamma}$ for all terms s, t. In this second case the model $M_{\sim_\Gamma}$ will have a universe consisting of only one element, viz. the set of all terms of L.

Identities (i.e. equational sentences) differ from purely universal sentences in general in that they have special preservation properties. More precisely, we have a preservation theorem for conjunctions of identities: A sentence of L is logically equivalent to a conjunction of identities iff it is preserved under (i) submodels; (ii) homomorphic images; and (iii) direct products.

Of the three model-theoretic relations that are involved in these preservation properties the first two -that of a model M being a submodel oof some other model M' and that of h being a homomorphism of a model M into a model M' have laready been defined (the first in Ch. 1 Sn. 1.5.2, Def. 20 and the second in this Chapter, Sn. 2.1.6, Def. 8).

The *direct product* $M_1 \otimes M_2$ of two models $M_1 = \langle U_1,F_1\rangle$ and $M_2 = \langle U_2,F_2\rangle$ of L is defined as follows: The universe U of the product is the set of all ordered pairs $\langle a,b\rangle$ with $a \in U_1$ and $b \in U_2$; and for any n-place function constant f, the interpretation F of f is the function defined as follows:

$$F(f)(\langle a_1,b_1\rangle,..,\ \langle a_n,b_n\rangle) = \langle F_1(f)(a_1,..,\ a_n),\ F_2(f)(b_1,..,\ b_n)\rangle.$$

Def. 11 Let $M_1 = \langle U_1,F_1\rangle$ and $M_2 = \langle U_2,F_2\rangle$ be models for the algebraic language L. The *direct product of* M_1 and M_2 is the model $M = \langle U,F\rangle$, where:

(i) $U = \{\langle a,b\rangle: a \in U_1 \ \&\ b \in U_2\}$
(ii) $F(f) = \{\langle\langle a_1,b_1\rangle,..,\langle a_n,b_n\rangle,\langle F_1(f)(a_1,..,\ a_n),\ F_2(f)(b_1,..,\ b_n)\rangle:$
$a_1,..,\ a_n \in U_1 \ \&\ b_1,..,\ b_n \in U_2\}$

The direct product of M_1 and M_2 is denoted as $M_1 \otimes M_2$.

Exercise: Show that if E is an equation of L, M is the direct product $M_1 \otimes M_2$ of two models M_1 and M_2 for L, $M_1 \models E$ and $M_2 \models E$, then

$M \models E$.

Hint. First show, by induction on the complexity of terms t of L, that for any assignments **a** in M_1 and **b** in M_2, the product assignment **a** $\otimes$ **b** in $M_1 \otimes M_2$ assigns to t in $M_1 \otimes M_2$ the value $<[[t]]_{M_1,\mathbf{a}}, [[t]]_{M_2,\mathbf{b}}>$. Here **a** $\otimes$ **b** is the assignment which assigns to each variable v_i the element $<\mathbf{a}(v_i),\mathbf{b}(v_i)>$ of $M_1 \otimes M_2$.

Before we turn to the exact formulation and proof of the preservation result for conjunctions of identities, it will be useful to first make a general observation about a special type of model for algebraic languages. These are the so-called *term models*. We encountered an example of such a model in the Completeness Proof for Equational Logic just given, where we constructed a counter example to the consequence claim $\Gamma \models E$ in the form of a model M whose elements were equivalence classes of terms. In general, a term model for an algebraic language L is a model whose universe consists of equivalence classes of the terms of L, where these equivalence classes are generated by equivalence relations which are also congruence relations with respect to all the function constants of L.

More specifically, given a congruence relation $\sim$ of the set Te_L of all terms of L, the corresponding model $M_\sim$ will have for its universe the set $\{[t]: t \varepsilon Te_L\}$, where Te_L, and as interpretation for any n-place function constant f of L the function defined by:

$$f_{M\sim}([t_1]_\sim, ..., t[t_n]_\sim) = [f(t_1, ..., t_n)]_\sim$$

The term models for a given algebraic language L are situated between two extremes. At the one and of the spectrum we find the so-called *free algebra for* the language L. This is the model generated by the identity relation on the set Te_L. Obviously this is an equivalence relation and congruence relation wrt to all function constants of L. Its equivalence classes are all the singleton sets {t}, where $t \varepsilon Te_L$. We denote this model as $M_{fr(L)}$. Clearly any other term model $M_\sim$ for L, generated by some congruence relation $\sim$, is a homomorphic image of $M_{fr(L)}$. For it easy to see that the map $\{t\} \Rightarrow [t]_\sim$ is is homomorpism from $M_{fr(L)}$ onto $M_\sim$. At the other end of the spectrum we find the model generated by the universal relation U_{Te_L} on Te_L. Again, this

relation is an equivalence relation and congruence relation wrt. the function constants of L. The model generated by this relation has for its universe the singleton set $\{Te_L\}$, and the interpretation of the function constants are, of necessity functions which map the one tuple all of whose members are the one element of this universe to this element. Since every congruence relation $\sim$ is a refinement of U_{Te_L}, the model just described is a homomorphic image of the model $M_\sim$.

More generally, if $\sim_1$ and $\sim_2$ are equivalence and congruence relations on Te_L and $\sim_1 \subseteq \sim_2$, then $M_{\sim_2}$ is a homomorphic image of $M_{\sim_1}$. For the map h which maps each element $[t]_{\sim_1}$ of the universe of $M_{\sim_1}$ onto $[t]_{\sim_2}$ is a homomorphism from $M_{\sim_1}$ onto $M_{\sim_2}$. At the opposite and from we find the one element term algebra $<Te_L, F>$, where for any $f^n \varepsilon$ L, $F(f) = \{<Te_L, .., Te_L, Te_L>\}$ (with $<Te_L, .., Te_L, Te_L>$ the n+1-tuple all of whose members are Te_L).

Given any model M for L we can associate a term model with M in several ways. First, we can form the equivalence relation $\sim_M$ on the set of terms of L defined by: $s \sim_M t$ iff $M \models s \equiv t$. Evidently, the resulting term model $M_{\sim_M}$ will verify exactly the same equations as M. But beyond that it is not so easy to say how M and $M_{\sim_M}$ are related. A second method goes as follows. We extend L to a language L^+ with names for each of the objects in U_M. (i.e. $L^+ = L \cup \{c_a: a \varepsilon U_M\}$; cf. the definition of the diagram of M in Ch. 1.) Let M^+ be the expansion of M in L^+, i.e. $c_{a_{M+}} = a$ for $c_a \varepsilon L^+ \backslash L$ and otherwise M^+ is like M. Now let $\sim_{M+}$ be the relation between terms of L^+ defined by

$$s \sim_{M+} t \text{ iff } M^+ \models s \equiv t$$

$\sim_{M+}$ is an equivalence relation on Te_{L+} and a congruence relation wrt all function constants of L^+. Thus $M_{\sim_{M+}}$ is a well-defined model for L^+. In this case too an equation of L will be true in the derived term model iff it is true in the original model M. Moreover, since for distinct objects a and b in U_M, $M^+ \models c_a \neq c_b$, $[c_a]_{\sim_{M+}} \neq [c_b]_{\sim_{M+}}$. So the map $a \Rightarrow [c_a]_{\sim_{M+}}$ is a 1-1 map into the universe of $M_{\sim_{M+}}$. It is easy to verify that this map is an isomorphism between M and a submodel of $M_{\sim_{M+}}$, but in general this will be a proper submodel of $M_{\sim_{M+}}$. A third possibility is to form a model $M'_{\sim_{M+}}$, whose universe consists of the

equivalence classes under $\sim_{M+}$ of all the *closed* terms of L^+. Here, the map $a \Rightarrow [c_a]_{\sim_{M+}}$ is a 1-1 map *onto* the universe of $M'_{\sim_{M+}}$ and thus an isomorphism from M to $M'_{\sim_{M+}}$.

To conclude these remarks on term models, we recall an important property concerning the values of terms in term models which we established and made use of in the Completeness Proof above:

(*) Let $M_\sim$ be a term model for the language L based on the congruence relation $\sim$, let t be a term of L, let $x_1, ..., x_n$ be the variables occurring in t and let a be an assignment in $M_\sim$ such that for i = 1,.., n, $a(x_i) = [x_i]_\sim$. Then $[t]_{M_\sim, a} = [t]_\sim$.

As we have seen, (6) can be proved by a simple induction on the complexity of terms.

We are now ready to prove the mentioned preservation theorem for equations:

Theorem 13 (Birkhoff)

Let L be an algebraic language. A sentence A of L is logically equivalent to a conjunction of identities of L iff (a) A is satisfiable and (b) A is preserved under (i) submodels; (ii) homomorphic images; and (iii) direct products.

Proof

$\Rightarrow$ The direction from left to right is straightforward. Clearly each identity is preserved by taking submodels (since identities are purely universal sentences), direct products (since the matrix of an identity is an atomic formula); and homomorphic images (since the matrix has the form of an equation "s = t"). And since the individual identities satisfy these conditions, the same is obviously true of their conjunctions. Finally, if the truth of each such conjunction is preserved under the model relations in question, then the same will be true for any sentence that is logically equivalent to such a conjunction.

$\Leftarrow$ The hard part is (as always with preservation theorems) the direction from right to left. Suppose that A is a sentence that is

preserved under taking submodels, direct products and homomorphisms. Let $\Gamma = \{E: E \text{ is an identity such that } A \vdash E\}$. First we show that if $M{\sim}_\Gamma \models A$, then $\Gamma \models A$.

To show that $\Gamma \models A$, we have to show that if M is any model of Γ, then $M \models A$. In view of the Completeness Proof we know that it suffices to show this for denumerable models. So let M be a denumerable model of Γ. Let g be an assignment in M which maps the set of variables onto U_M. We extend g to the set of all terms of L by letting $g(t) = [[t]]_{M,g}$. Suppose that s and t are two terms of L such that $s \sim_\Gamma t$. Then $\Gamma \models s = t$. So, since $M \models \Gamma$, $M \models s = t$. So $[[s = t]]_{M,g} = 1$. So $[[s]]_{M,g} = [[t]]_{M,g}$. So the map g from terms t to elements $[[t]]_{M,g}$ induces a map from the equivalence classes $[t]_{\sim_\Gamma}$ onto the elements of M. It is also easily verified that this map is a homorphism. So, since A is preserved by homomorphisms and by assumption $M{\sim}_\Gamma \models A$, it follows that $M \models A$.

So we conclude that $\Gamma \models A$. But then there is a finite set of $E_1, .., E_n$ in Γ such that $E_1 \& .. \& E_n \vdash A$. So, since on the other hand $A \vdash E_i$ for all i ($1 \leq i \leq n$), $\vdash A \leftrightarrow (E_1 \& .. \& E_n)$.

It remains to show that $M{\sim}_\Gamma \models A$. Suppose not. Then $M{\sim}_\Gamma \models \neg A$. Let $(M{\sim}_\Gamma)^+$ be the expansion of $M{\sim}_\Gamma$ in some language $L^+ = L \cup \{c_a: a \in U_{M{\sim}_\Gamma}\}$ and let $D((M{\sim}_\Gamma)^+)$ be the set of (a) all equations s = t with s, t constant terms of L^+ and (b) all negations of such sentences. Then $D((M{\sim}_\Gamma)^+) \cup \{A\}$ is inconsistent. For if not, then $D((M{\sim}_\Gamma)^+) \cup \{A\}$ has a model. But this model will be (isomorphic to) an extension of $(M{\sim}_\Gamma)^+$. So $(M{\sim}_\Gamma)^+$ will be a submodel of this model and consequently, because A is preserved by taking submodels, $(M{\sim}_\Gamma)^+ \models A$. This contradicts the assumption that $M{\sim}_\Gamma \models \neg A$. Since $D((M{\sim}_\Gamma)^+) \cup \{A\}$ is inconsistent, there are $E_1, .., E_k, D_1, .., D_n$ in $D((M{\sim}_\Gamma)^+)$, where the E_i are of type (a) and the D_j of type (b) (see the def. of $D((M{\sim}_\Gamma)^+)$) and

(1) $A \vdash \neg (E_1 \& .. \& E_k \& D_1 \& . \& D_m)$

Since A does not contain any of the constants $\{c_a: a \in U_{M\sim\Gamma}\}$,

(2) $A \vdash (\forall x_1)..(\forall x_r) \neg (E'_1 \& .. \& E'_k \& D'_1 \& . \& D'_m)$

where (i) $c_{a_1}, ..., c_{a_r}$ are all the new constants occurring in $E_1, .. , E_k, D_1, .., D_m$, (ii) $x_1, ..., x_r$ are r new variables (i.e. variables not occurring in A or $E_1, .. , E_k, D_1, .., D_m$) and (iii) the E'_i and D'_j are the result of replacing in the E_i and D_j the constants c_{a_h} by the variables x_h.

First assume that k = 0 (i.e. all the conjuncts on the right hand side in (1) are of type (b)):

(3) $A \vdash (\forall x_1)..(\forall x_r) \neg (D'_1 \& . \& D'_m)$

Consider D'_1. Suppose D'_1 is the inequality $s_1 \neq t_1$. We know that the elements $a_1, ..., a_r$ of $U_{M\sim\Gamma}$ satisfy $s_1 \neq t_1$ in $M\sim_\Gamma$. This means that the identity $s_1 = t_1$ does not belong to Γ, for if it did it would be satisfied in $M\sim_\Gamma$ by all possible combinations of elements of $U_{M\sim\Gamma}$. So it is not the case that $A \vdash s_1 = t_1$. That is, A is consistent with $(\exists x_1)..(\exists x_r)\ s_1 \neq t_1$. So there is a model M_1 of $\{A\} \cup \{(\exists x_1)..(\exists x_r)\ s_1 \neq t_1\}$. So there are objects $a_{11}, ..., a_{1r}$ in U_{M_1} which satisfy $s_1 \neq t_1$ in M_1. In the same way we can find models M_j of $\{A\} \cup \{(\exists x_1)..(\exists x_r)\ s_j \neq t_j\}$ and sequences of objects $a_{j1}, ..., a_{jr}$ in their universes which satisfy $s_j \neq t_j$, for each of the remaining disjuncts D'_j. Let M be the direct product of the models M_j and let for i = 1, ..., r $b_i = \langle a_{1,i}, ..., a_{m,i}\rangle$. Then (i) since A is preserved by direct products, A holds in M and (ii) the sequence $\langle b_1, ..., b_r\rangle$ simultaneously satisfies all inequalities $s_1 \neq t_1, .., s_m \neq t_m$ in M. But the existence of such a model contradicts (3).

Now assume that k > 0. Consider E'_1. E_1 is of the form $s_1 = t_1$. Since $(M\sim_\Gamma)^+ \models E_1$, the elements $a_1, ..., a_r$ of $M\sim_\Gamma$ satisfy the equation $s'_1 = t'_1$ in $M\sim_\Gamma$. Now let $q_1, ..., q_r$, be terms of L such that for i = 1, .., r, $q_i \in a_i$. Then we have, for i = 1, .., r, $a_i = [q_i]\sim_\Gamma$. Let $z_1, ..., z_s$ be all the variables occurring in $q_1, ..., q_r$, and let b be an assignment such that for h = 1, ..., s, $b(z_h) = [z_h]\sim_\Gamma$. Then according to (*), we have for i = 1, ..., r that

(4) $[[q_i]]_{M\sim\Gamma},b = [q_i]_{\sim\Gamma}$.

Let s''_1 be the result of substituting the terms q_i for the variables x_i in s'_1; in the same way we obtain t''_1 from t'_1. By Lemma 3 of Ch. 1,

(5) $[[s''1]]_{M\sim\Gamma},b = [[s'1]]_{M\sim\Gamma},b'$,

where b' is the assignment which is like b except that for i = 1, ..., r, $b'(x_i) = [[q_i]]_{M\sim\Gamma},b$; and similarly for t''_1 and t'_1. But according to (4), $[[q_i]]_{M\sim\Gamma},b = [q_i]_{\sim\Gamma} = a_i$. So $[[s'1]]_{M\sim\Gamma},b'$ is the value of s'_1 in $M\sim\Gamma$ under any assignment which assigns the a_i to the x_i, and the same is true for $[[t'1]]_{M\sim\Gamma},b'$. Since $M\sim\Gamma \models s'_1 = t'_1\ [a_1, ..., a_r]$, it thus follows that

(6) $[[s''1]]_{M\sim\Gamma},b = [[t''1]]_{M\sim\Gamma},b$.

Now note that the variables in s''_1 and t''_1 are $z_1, ..., z_s$. So we can apply (*) once more, obtaining that $[[s''1]]_{M\sim\Gamma},b = [s''1]_{\sim\Gamma}$ and similarly for t''_1. So from (6) we conclude that $[s''1]_{\sim\Gamma} = [t''1]_{\sim\Gamma}$, that is:

(7) $s''_1 \sim\Gamma t''_1$.

But this means that

(8) $\Gamma \vdash s''_1 = t''_1$.

Since $A \vdash \Gamma$, $A \vdash s''_1 = t''_1$, that is

(9) $A \vdash (\forall z_1)..(\forall z_s)(s'[qi/xi] = t'[qi/xi])$

Now substitute the terms $q_1,..,q_r$ for the corresponding variables $x_1,..,x_r$ throughout the matrix of the formula on the right of $\vdash$ in (2). This will turn the conjuncts E'_i, D'_j into new conjuncts E''_i, D''_j which are substitution instances of the E'_i and D'_j. From (2) we infer that

(10) $A \vdash (\forall z_1)..(\forall z_s) \neg (E''_1 \& .. \& E''_k \& D''_1 \&. \& D''_m)$

Note further that the argument we have just given for E'_1 applies equally to each of the other E'_j (if any) and that the choice of the terms q_i can be the same in each case (i.e. irrespective of which E'_j we consider. In other words we have:

(11) $A \vdash s''_j \equiv t''_j$, for j = 1, .., k.

Because of (9) we can eliminate the disjunct E''_1 from the negated conjunction. This reduces (10) to (12)

(12) $A \vdash (\forall z_1)..(\forall z_s) \neg (E''_2 \,\&\, .. \,\&\, E''_k \,\&\, D''_1 \,\&. \,\&\, D''_m)$

But because of (11), the same argument applies to each of the other E''_i (i = 2,.., k). So each of these conjuncts can be removed from (12) and we end up with a formula of the form (4) with each of the conjuncts satisfiable in $M_{\sim\Gamma}$. We have already seen that this leads to a contradiction.

q.e.d

Exercise. Let L be the algebraic language consisting of two 1-place function constants f and g. Let Γ be the pair of equations {f(x) = x, g(x) = x}. Show: there is no single equation E of L which is logically equivalent to the conjunction $(\forall x)(f(x) = x) \,\&\, (\forall x)(g(x) = x)$.

We conclude this section with the comment which we promised in the introduction to Section 2.2. There we noted that formulas that contain function constants may seem to carry, because of those function constants, additional quantificational information other than what is directly visible from the quantifiers that are overtly displayed. This extra information becomes explicit, when the formula is translated into one in which the function constants are replaced by predicates. In particular, this translation will normally convert a purely universal formula into one that is AE. In the light of this observation it might seem surprising that the preservation theorem for purely universal formulas which we proved towards the end of Ch. 1 applies not only to languages that only have predicates, but also to those some or all of whose non-logical constants are function constants. If it is true, one might ask, that in general a purely universal sentence with function constants has the force of an AE sentence, how then can it be that such formulas obey the same model-theoretic restrictions as the "genuinely purely universal" sentences which consist of a purely universal prefix

followed by a quantifier-free matrix in which there are no function symbols?

The explanation of this apparent paradox is that when we are dealing with a language L which has function constants, the submodel relation between models for L is subject to restrictions which do not play a role when we deal with models for languages which only have predicates. Whenever M = <U, F> is any model for a language without function constants and U' is a subset of U, then there is always a unique submodel M' = <U',F'> of M, in which F' assigns to each predicate P of the language the restriction to U' of the interpretation F(P) assigned to P in M. When the language L contains function constants, this no longer holds in general. Suppose for instance that L contains the 1-place function constant f and let M be any model for <U, F> and U' a subset of U. In order that there be a submodel M' = <U',F'> of M whose universe is U' it should be the case that the restriction of F(f) to U' satisfies the requirements for interpretations of 1-place function constants, viz that the interpretation is a function from the universe into itself. In general this won't be the case, for there may well be elements a ε U' such that F(f)(a) belongs to U \U'. In that case the pair <a,F(f)(a)> will not belong to F'(f), F'(f) will thus only be a partial but not a total function form U' into U' and thus unsuitable as interpretation for f.

The upshot of this is that when L contains function constants, then the submodel relation is much harder to satisfy than it is for pure predicate languages. Consequently truth preservation under arbitrary submodels is a condition that is easier to satisfy for such languages than for pure predicate languages - since there are fewer submodels, it is easier for a sentence to have the property that whenever it is true in a given model it is also true in all its submodels. In fact, the general validity of preservation theorem of Ch. shows that the extra quantificational complexity that formulas may seem to have because of containing complex terms is "matched" by the special constraints which function constants impose on the submodel relation.

Arguably this comment would have been more appropriate after the proof of the preservation theorem in Ch. 1. But since the general issue that prompted it was raised only in this chapter, this seemed the next best place to make the comment. For the preservation properties of universally quantified equations are, as Birkhoff's Theorem asserts, even stricter than those for purely universal formulas - preservation under formation of submodels being one (but only one) of the properties that distinguish sentences that are equivalent to a universally

quantified equation. Since universally quantified equations are preserved under submodel formation and since they too will usually produce additional existential quantifiers when translated into formulas with predicates, they too give rise to the apparent paradox of which we have spoken.

2.4.1 Unification

A very different conception and use of equations is found in connection with *unification*. Here equations are understood as constraints on a structure consisting of (presumably) connected objects which are represented by the variables of a given set $\mathbb{E}$ of equations. Thus the equations in the set $\mathbb{E}$ are not understood as universally true - i.e. as universally quantified sentences - but as "locally true" - i.e. as true of the particular objects which the variables occurring in $\mathbb{E}$ represent. What one is after is a particular set of values for the variables for which all the equations are satisfied.

In certain situations one moreover wants the simultaneous solution to $\mathbb{E}$ to be "provably correct". More specifically, what one is looking for is a way of specifying the values so that the fact that they form a solution to the equations becomes a fact of pure logic. There is one salient and natural way in which this may be accomplished, and it is this: Let L be the language of the equations in $\mathbb{E}$ and let $M_{fr(L)}$ be the free algebra for L. Suppose that $x_1, ..., x_n$ are the variables occurring in $\mathbb{E}$ and that a is an assignment in $M_{fr(L)}$ such that $[[E]]_{M_{fr(L)},a} = 1$ for all $E \varepsilon \mathbb{E}$. Suppose that for $i = 1, ..., n$, $a(x_i) = [r_i] = \{r_i\}$. It is easy to see that, supposing that E is the equation $s = t$, s' is the result of replacing $x_1, ..., x_n$ in s by $r_1, ..., r_n$ and likewise for t' and t, $[[E]]_{M_{fr(L)},a} = 1$ implies that the equation $s' = t'$ is a tautology, i.e. s' is the very same term as t'; and thus that $s' = t'$ is a (trivial) theorem of pure logic.

The problem of finding such a "logically valid" simultaneous solution to the equations in a given set $\mathbb{E}$ in the free algebra for L is known as the problem of (*term*)*unification*.

The problem of unification is usually stated as the question whether a set of equations has a *unifier* (or *unifying substitution*). Let us begin by introducing the relevant notions.

Def. 12 Let L be an algebraic language, X a set of variables.

i. A *substitution on* X *in* L is a function σ with domain X, which assigns each variable x_i in X a term r_i of L.

ii. Suppose σ is a substitution on X. There is a standard extension σ' of σ to the set of all variables, defined by

$\sigma'(v_j) = \sigma(v_j)$, if $v_j \varepsilon X$
$\sigma'(v_j) = v_j$ otherwise

Since there is an obvious 1-1 correspondence between substitutions on subsets X of the set of all variables and their extensions as just defined, we won't distinguish between them, using "σ" both to refer to the substitution σ on X itself and to its extension σ'.

iii. Let σ, τ be two substitutions. By $\sigma \circ \tau$, the *composition of* σ and τ, we understand the substitution ρ which assigns to each variable v_j the term $\rho(v_j)$ which we obtain by simultaneously substituting for the variables v_k occurring in $\sigma(v_j)$ the terms $\tau(v_k)$.

iv. Suppose that $\mathbb{E}$ is a set of equations of L and that σ is a substitution in L. Then σ is called a *unifier of* $\mathbb{E}$ iff for each $E \varepsilon \mathbb{E}$, $\models E[\sigma]$, where $E[\sigma]$ is the result of simultaneously substituting the terms $\sigma(v_j)$ for the variables v_j which have free occurrences in E.

The main result about unification is that for finite sets of equations the problem whether a unifier exists is decidable: There exists an algorithm (due to Martelli & Montanari), which will find a unifier in a finite number of steps if one exists, and will return a negative answer to the question, when there is no simultaneous solution. Moreover, the algorithm returns, in those cases where there is a solution, a so-called "most general unifier" for the given equation set.

Def. 13 Let $\mathbb{E}$ be a set of equations of L and σ a substitution in L. Then σ is called a *most general unifier of* $\mathbb{E}$ iff (i) σ is a unifier of $\mathbb{E}$; and (ii) for any unifier ρ of $\mathbb{E}$ there is a substitution τ such that $\rho = \sigma \circ \tau$.

Thm. 14

i. There exists an algorithm which (i) returns for any finite set of equations $\mathbb{E}$ of any algebraic language L in finitely many steps either a unifier σ for $\mathbb{E}$ or else the answer that no unifier of $\mathbb{E}$ exists.

ii. The unifier σ which the algorithm returns when $\mathbb{E}$ does admit of a simultaneous solution is a most general unifier for $\mathbb{E}$.

[Ref. ??]

N.B. 1. Note that when $\mathbb{E} = \{E_1, ..., E_n\}$, then, if σ is unifier of ,

$$\models (\forall x_1)...(\forall x_k)(E_1[\sigma] \,\&\, ... \,\&\, E_n[\sigma]),$$

where $x_1, ..., x_k$ are all the variables occurring in $(E_1[\sigma], ..., E_n[\sigma]$. This formulation is especially apt to show how strong a claim unifiability really is.

2. The unification problem is special in that it asks for a substitution which turns all equations in the set into tautologies. There are many situations where such a result is stronger than one really needs. Rather, what is wanted is a substitution which turns all equations into theorems of a given theory T:

For all $E \varepsilon \mathbb{E}$, $T \models E[\sigma]$

It should be stressed that with each different T the corresponding unification problem one is dealing with is a different one; and as a rule the problems are very different indeed, involving very different combinatorics, as a function of the axiomatic principles that T includes. This is so in particular in certain cases where T is itself an equational theory. For a few simple examples of such equational theories T the unification problem has been showed to be undecidable - which is one indication of how different the problem may become when a non-tautological theory T is brought into play.

(Two references on Unification: (i) Martelli, A. & U. Montanari. An Efficient Unification Algorithm. ACM Transactrions on Programming Languages and Systems, April 1982. (ii) Lloyd, J.W., *Foundations of Logic Programming*. Springer, 1984)

2.5 Definitions.

It is common practice to extend given scientific theories by adding new notions via definitions. Sometimes the point of a definition is strictly one of notational convenience: the defined concept abbreviates a complicated expression in the "primitive" vocabulary of the theory (that is, of the vocabulary in which the theory is given initially) and thus allows simplification of statements which contain this expression as a part. In other cases the defined notion has a conceptual significance of its own, which will make it easier to understand and handle statements in which it is represented as a unit - i.e. by a single symbol or term - than they would be if the concept were circumscribed in the theory's primitive vocabulary. And in yet other cases the defined concept may be one that is directly accessible to empirical observation, and deserve to be made explicit by a separate definition for that reason. In fact, the method of introducing concepts by definition is so general and of such methodological importance that most textbooks on logic and/or scientific methodology devote a separate chapter to it.

Here we will look at issues connected with definitions within the specific context of theories formalised within first order logic. That somewhat limits the range of issues that the theory and pactice of definition give rise to in general. Nevertheless, there remain a number of useful things to be said and these we will address. (Something that does not fit within the setting we adopt here is the conceptually important question of (non-)circularity of definitions. We will have a few observations about this notion towards the end of the section.)

In relation to first order theories questions of definition arise in two different settings. The first is that implicit in what was said in the opening paragraph: We have a theory T of some first order language L and want to extend T by adding some notion by definition. Formally this will consist in (i) choosing a new symbol α for the notion that is to be added to it, (ii) extending the language L to the language $L' = L \cup \{\alpha\}$ and then (iii) extending T to the theory T' of L' which is obtained by adding the definition of α to T and then closing under logical consequence in L'. This is what might be called the *external perspective* on definition.

But questions of definition can also be raised from a theory-i*nternal* prespective. Suppose again that T is a theory of L but now α is a non-logical constant of L. We can then ask the question whether α could not be defined within T in terms of its remaining vocabulary: Is there a definition D of α in terms of the remaining vocabulary which (i) is a

theorem of T and (ii) will give us back all of T when combined with the reduction T' of T to the language L' = L\{α} (i.e. the theory which consists of all theorems of T that belong to L')? Or - to put the question a little more informally - could we not eliminate all statements involving α from T and then restore them again to T by adding D?

In order to state this second question with the necessary precision we need to first have a clearer notion of what a "definition" is. We just spoke of "adding a definition of α" to some first order theory. That implies that the defintion in question must be a first order sentence, which we can add to a theory as an additional axiom. But which sentences should qualify as possible definitions of some non-logical constant α? What do we, or should we, expect of a sentence that is to serve as a definition? There are two criteria that, as the result of discussions of the purpose and form of definitions that stretched over centuries, have emerged as the central functional requirements. These are:

(i) *conservativity*

and

(ii) *determination*.

(i) Conservativity is a notion that does not only arise in connection with definitions. Its general context is that of a theory T of some language L and an extension T' of T whose language is some extension L' of L. T' is called a *conservative extension of* T iff T' coincides with T as far as L is concerned: if A is a sentence of L, then A is a theorem of T' iff it is a theorem of T.

The notion of conservativity as definability constraint involves a straightforward application of the "conservative extension" relation. Intuitively, the constraint is that adding a definition D of a new notion α to a theory T should not introduce new information that is expressible in the primitive vocabulary L of T. The formal expression of this requirement is as follows: every sentence A of L that is a theorem of the theory T' = $Cl_{L'}(T \cup \{D\})$ (where as before L' is the language L ∪ {α}) is already a theorem of L; or, put in terms of the notion just introduced: T' is a conservative extension of T.

(ii) Determination is the principle that a definition D of α should fully determine the extension of α when the extensions of the notions in

terms of which D defines α are given. The formal characterisation of this condition is model-theoretic: Let T and T' be as under (i) and let M = <U,F> be a model for L that is a model of T. Then there should be one and only one way to expand M to a model M' = <U,F'> for the language L' = L $\cup$ {α} that is a model of M'. That is, there ought to be only one way of extending F to an interpretation function F' of the non-logical constants of L', i.e. only one way of adding an interpretation F'(α) for α which verifies all the additional theorems of T' (including, in particular, the new "axiom" D)

Of these two criteria determination is the stronger one; it entails conservativity. For supppose that T, T', L and L' are as above and that D satisfies determination of α in relation to T. That is:

(3) For every model M of T there is one and only one expansion M' of M to L' which is a model of T'.

To show that T' = $Cl_{L'}$(T $\cup$ {D}) is a conservative extension of T assume that (3) holds and that A is a sentence of L such that T' $\models$ A. We must show that T $\models$ A. Suppose that it is not the case that T $\models$ A, Then T $\cup$ {$\neg$ A} consistent. Let M be a model of T $\cup$ {$\neg$ A}. Then there will be no expansion M' of M that is a model of T'. For every such expansion will verify $\neg$ A, while A is a theorem of T'.[24]

With this we are now in a position to address the question what form a definition should have in order that the mentioned criteria are satisfied. Since determination entails conservativity, it suffices to consider just determination.

Within formal logic we find two different forms of definitions which both satisfy determination. For the first of these, known as *explicit definition*, this is almost trivial. For the second, *definition by recursion*, - also called "definition by induction", or "recursive definition" or "inductive definition" - determination isn't quite as obvious, but even for this type of definition it is relatively easy to see that all the familiar

---

24 It is natural to ask whether conservativity in its turn entails determination. As it stands, I so not know the answer to this question. (I suspect the answer must be known but i haven't done the extensive literatire check need to find out whether this is so.) my hunch is that the entailment in this direction does not hold. it may fold under certain restrictions, but I have no clear idea what these might be either.

instances do satisfy determination. In this section we will only consider explicit definitions.[25]

Explicit definitions are universally quantified biconditionals in which an atomic formula involving the symbol that is being defined stands to the left of $\leftrightarrow$ and its definition - some formula A of the language in which the new symbol is being defined - to its right. (The left hand side and the right hand side are often referred to as the *definiendum* and the *definiens* of the given definition.) Exactly what this comes to still depends on what type of symbol α - or, more accurately: what type of non-logical constant α - is being defined. If α is an n-place predicate P, then an *explicit definition for* α *in* a language L has the form specified in (4)

(4) $(\forall x_1)...(\forall x_n)(P(x_1,.., x_n) \leftrightarrow A(x_1,.., x_n))$,

where $x_1,.., x_n$ are n distinct variables and A is a formula of L not containing P whose only free variabes are $x_1,.., x_n$.

Explicit definitions of function constants are essentially of the same form, except that the atomic formula on the left reflects the fact that we are dealing with a function constant rather than a predicate constant. The form of an explicit definition for an n-place function constant is given in (5).

(5) $(\forall x_1)...(\forall x_n)(\forall x_{n+1})(f(x,.,x_n) = x_{n+1} \leftrightarrow A(x_1,.., x_n,x_{n+1}))$,

where $x_1,.,x_n, x_{n+1}$ are n + 1 distinct variables and A is a formula of L not containing f whose only free variabes are $x_1,.., x_{n+1}$.

It is easy to see that sentences of the form (4) sastisfy determination. Suppose again that T is a theory of L, that P does not belong to L and that we form the theory $T' = Cl_{L'}(T \cup \{D\})$ of the language $L' = L \cup \{P\}$, where D has the form given in (4). Let $M = \langle U,F \rangle$ be a model of T. The right hand side A of D has for its extension the set $[[A]]^M$ in M, where $[[A]]^M = \{\langle u_1,.., u_n\rangle$: for i = 1,..,n, $u_i \in U$ & $[[A]]^M[u_1,.,u_n] = 1\}$.
Let $M' = \langle U,F' \rangle$ be the expansion of M to L' defined by F' =

---

[25] Examples of recursive definitions will be encountered in the next section, where we deal with the axiomatisation of natural number arithmetic. In chapter 3 recusive definitions will be discussed in greater depth; there we will in particular look at the systematic connections that exist between recursive and explicit definitions.

$F \cup \{<P,[[A]]^M>\}$. It is easily verified that $M' \models T'$. (This follows from the fact that on the one hand $M' \models T$, while on the other the choice of $F'(P)$ guarantees that $M' \models D$.) This establishes that there is at least one expansion of M which verifies T'. Secondly, suppose that $M'' = <U,F''>$ is another expansion of M such that $M'' \models T'$. Then in particular $M'' \models D$. This means that for every n-tuple $<u_1,..,u_n>$ of elements of U, $[[P(x_1,.,x_n)]]^{M''}[u_1,..,u_n] = 1$ iff $[[A]]^{M''}[u_1,..,u_n] = 1$.

In other words, $[[P(x_1,.,x_n)]]^{M''} = [[A]]^{M''}$, where

$[[P(x_1,.,x_n)]]^{M''} = \{<u_1,..,u_n>: u_i..u_n \in U \ \& \ [[P(x_1,.,x_n)]]^M[u_1,..,u_n] = 1\}$

and

$[[A]]^{M''} = \{<u_1,..,u_n>: u_i \in U \text{ for } i = 1,..,n \ \& \ [[A]]^M[u_1,..,u_n] = 1\}$.

But the first of these two sets is nothing other than $F''(P)$ and the second set equals $[[A]]^M$. This entails that $[[P(x_1,.,x_n)]]^{M''} = [[A]]^M = [[P(x_1,.,x_n)]]^{M'}$ and thus that $M'' = M'$.

The case of (5) is a little more complicated. A definition D of the form (5) does not automatically guarantee determination, because the form of D imposes certain constraints on the semantics of its definiens A. D says that $A(x_1,..,x_n,x_{n+1})$ is equivalent to a statement of the form "$f(x_1,..,x_n) = x_{n+1}$". This means that in any model M' of D there will have to be for any n-tuple $<u_1,.,u_n>$ of elements of the unviverse exactly one u_{n+1} such that $[[A]]^{M'}[u_1,..,u_n,u_{n+1}] = 1$. This means that the correspnding "unique value" condition (6) for A will be a theorem of T', whether or not it is a theorem of T.

(6) $(\forall x_1)...(\forall x_n)(\exists y)(A(x_1,..,x_n,y) \ \& \ (\forall y')A(x_1,..,x_n,y) \rightarrow y' = y)$,

So if T' is to be a conservative extension of T, then (6) should be a theorem of T to begin with.

The upshot of this is that an explicit definition D of a function constant is acceptable as an addition to a theory T only if T already entails the corresponding unique value condition (6) for its definiens A. For only then will the addition of D be conservative. However, when this condition is fulfilled, then the addition of D will not only satisfy conservativity but also determination. (The argument is the same as for explicit definitions of predicates.)

The general moral of this discussion is that sentences of the form (4) and, with the qualifications just noted, also those of the form (5) satisfy the requirements we laid down for good definitions. This is consistent with the almost universal practice to cast definitions of new symbols in these particular forms.[26]

This concludes our discussion of the external perspective on the question what constitutes a proper definition, and we now turn to the internal perspective. In discussing the questions that this perspective gives rise to we follow the tradition in that we assume the notion of an explicit definition, as specified in (4) and (5), as our syntactic characterisation of proper definitions.

Suppose that T is a theory of the language L and that α is a non-logical constant of L. We already stated what it means for α to count as definable within T: there has to be some definition D of α in the language $L' = L\backslash\{\alpha\}$) such that $T = Cl_L(T' \cup \{D\})$, where $T' = T \cap \{A: A \in L'\}$. Now that we have adopted a specific syntactic characterisation of definitions we can turn this notion of definability into a strictly formal characterisation:

(7) Let T be a theory of a first order language L and α a non-logical constant of L. Let $L' = L\backslash\{\alpha\}$ and $T' = T \cap \{A: A \text{ is a sentence of } L'\}$. Then α is *explicitly definable in* T iff there exists an explicit definition D of α in L' such that $T = Cl_L(T' \cup \{D\})$.

We have already seen that when α is explicitly definable in T, then α is also *implicitly definable in* T, where implicit definability is characterised model-theoretically as in (8).

(8) Let T, L, α, L' and T' as in (7). Then α is *implicitly definable in* T iff the following condition holds:

Every model M' of T' can be expanded in one and only one way to a model M of T

It is an interesting fact that the converse of this implication - that implicit definability entails explicit definability - also holds. This result

---

[26] Recursive definitions are found almost exclusively within mathematics, something that has to do with the circumstance that they are suitable for domains that have the special "recursive" structure that such definitions presuppose.

differs from the statement that explicit definability entails implicit definability in that it depends on specific properties of first order predicate logic and is not generalisable to other logical formalisms (such as, for instance, higher order predicate logic). The result is known as *Beth's Definability Theorem*, after the Dutch logician E.W. Beth (1908-1964) who formulated and proved the theorem. To do justice to its importance we state Beth' Theorem once more, as a separate theorem with its own number.

Theorem 15 (Beth's Definability Theorem)

> Let L be a language of first order logic, α a non-logical constant of L and T a theory of L. If α is implicitly definable in T, then α is explicitly definable in T.

The proof of Beth's Theorem that we will present here is not the proof which Beth gave himself. But it is, I believe, the most popular proof of the theorem today. It makes use of another important theorem abolut first order logic, the so-called "Craig Interpolation Lemma". Craig proved this theorem on the way towards some other result in proof theory in which he was interested ar that point, hence the name "Interpolation *Lemma*". But it states a proposition which has come to be recognised as a salient fact about first order predicate logic in its own right. As in the case of Beth's Definability Theorem, there are other logical formalisms than first order logic for which the Interpolation Lemma does not hold, and in fact, validity of the Lemma has become (like the validity of Beth's Theorem) an important property in terms of which logical formalisms are classified. (Satisfying Craig's Lemma can be seen as a certain kind of well-behavedness for formal systems.)

The Interpolation Lemma says that if A and B are sentences of first order logic and $A \vdash B$, then there is a sentence C in the comon vocabulary of A and B such that $A \vdash C$ and $C \vdash B$. We can roughly paraphrase this as: That which is responsible for the fact that A is logically at least as strong as B can be articulated in just the terminology that is common to them both. A formal statement of the Interpolation Lemma is given as Theorem 16.

Theorem 16 (Craig's Interpolation Lemma).

Suppose that A is a sentence belonging to some first order language L_1, B a sentence belonging to some first order language L_2 and that L_1 and L_2 are compatible in that L_1 and L_2 assign the same signature to the

symbols they have in common. We denote the language whose non-logical constants are those common to L_1 and L_2 as L. Suppose that $A \vdash B$. Then there is a sentence C belonging to L such that $A \vdash C$ and $C \vdash B$.

The Interpolation Lemma can be proved quite easily on the basis of the completeness proof for first order logic that is given in the Appendix to Ch. 1. A proof of the Interpolation Lemma along those lines is given at the end of that Appendix. Here we will, as last item of this section, present a proof in which the same construction is used that is central to the completeness proof given in the main body of the text of Ch. 1 (see Section 1.2). This proof has an interest in its own right as a further application of the method used to prove completeness there, but it is more complicated than the one from the Appendix. (The central idea of this latter proof can be grasped immediately, although its technical details take up a certain amount of space.)

Proof of Beth's Theorem.

Beth's Theorem holds for arbitrary non-logical constants α. However, we will first give the proof for the case where α is an n-place predicate P. After completion of that proof we will then show how the case where α is a function constant can be reduced to the case where α is a predicate.

Suppose that L, T and α are as in the statement of the Theorem and that α is implicitly definable in T. Further assume that α is an n-place predicate P, that $L' = L\backslash\{P\}$ and that T' is the theory of L' defined by: $T' = T \cap \{A: A \text{ is a sentence of } L\}$. Let P_1 and P_2 be symbols not occurring in L and let L_1 and L_2 be the languages which result when we add, respectively, P_1 and P_2 as n-place predicates to L'. Let T_1 be the theory of L_1 which we get by replacing P in all theorems of T everywhere by P_1, and let, analogously, T_2. be the theory of L_2 which we get by replacing P in T everywhere by P_2. Let T_3 be the theory $CN_{L_3}(T_1 \cup T_2)$ in the language $L_3 = L_1 \cup L_2$. Then the following sentence (1) is a theorem of T_3:

$$(\forall x_1) \,..\, (\forall x_n)\ (P_1(x_1, \,..\,, x_n) \leftrightarrow P_2(x_1, \,..\,, x_n)) \qquad (1)$$

That (1) is a theorem of T_3 can be seen as follows. Suppose that M_3 is any model of T_3. Let M_1 be the reduction of M_3 to L_1, M_2 the reduction of M_3 to L_2 and M' the reduction of M_3 to L'. Then M_1 is a model of T_1, M_2 is a model of T_2 and M' is a model of T'. Since by assumption P is

implicitly defined in T, the same is evidently true of P_1 in relation to T_1 and of P_2 in relation to T_2. Since P_1 is implicitly defined in T_1, there is exactly one expansion M_1' of M' which is a model of T_1. So $M_1' = M_1$, which means that the extension of P_1 in M_1' is the same as it is in M_1. Since T_2 is just like T_1 except for renaming of the predicate P_1 as P_2, the unique expansion M_2' of M to L_2 that is a model of T_2 will assign to P_2 exactly the same extension as M_1' assigns to P_1. And, as before, the extension of P_2 in M_2' is the same as the extension of P_2 in M_2. So all these extensions are the same and in particular the extension of P_1 in M_1 is the same as the extension of P_2 in M_2. As these are also the respective extensions of P_1 and P_2 in M_3, P_1 and P_2 have the same extension in M_3. So it follows that (1) holds in M_3. Since this is true for arbitrary models M_3 of T_3, (1) is a logical consequence of T_3.

Since $T_3 \vdash$ (1), we also have $T_3 \vdash$ (2), where (2) is the result of dropping the universal quantifiers of (1) and replacing the variables $x_1,.., x_n$ by fresh individual constants $c_1,.., c_n$, which do not belong to L':

$$P_1(c_1,.,c_n) \leftrightarrow P_2(c_1,.,c_n) \qquad (2)$$

Since $T_3 = CN_{L_3}(T_1 \cup T_2)$ and $T_3 \vdash$ (2), there are finitely many sentences $D_{11},.., D_{1m}$ from T_1 and there are finitely many sentences $D_{21}, .., D_{2n}$ from T_2 such that

$$\{D_{11},.,D_{1n}, D_{21},.,D_{2m}\} \vdash P_1(c_1,.,c_n) \leftrightarrow P_2(c_1,.,c_n). \qquad (3)$$

We can choose the sentences $D_{11},..,D_{1n}, D_{21},..,D_{2m}$ in such a way that n = m and that D_{2i} is the result of replacing P_1 in D_{1i} by P_2. Forming the conjunction D_1 of the D_{1i} and the conjunction D_2 of the D_{1i} we get

$$D_1 \,\&\, D_2 \vdash P_1(c_1,.,c_n) \leftrightarrow P_2(c_1,.,c_n) \qquad (4)$$

and

$$D_2 = D_1\,[P_2/\,P_1]. \qquad (5)$$

(4) entails (6):

$$D_1 \,\&\, P_1(c_1,.,c_n) \vdash D_2 \rightarrow P_2(c_1,.,c_n) \qquad (6)$$

Note that in (6) the formula to the left of $\vdash$ belongs to L'_1 and the formula to its right belongs to L'_2, where $L'_1 = L_1 \cup \{c_1,.,c_n\}$, and

similarly for L'_2. So the Craig Interpolation Lemma applies: There is a sentence C from the common language $L' = L \cup \{c_1,., c_n\}$ such that

$$D_1 \,\&\, P_1(c_1,.,c_n) \vdash C \qquad (7)$$

and

$$C \vdash D_2 \rightarrow P_2(c_1,.,c_n). \qquad (8)$$

Since C does not contain any occurrences of P_2, the proof of $D_2 \rightarrow P_2(c_1,.,c_n)$ from C will turn into a proof of $D_1 \rightarrow P_1(c_1,.,c_n)$ from C when we replace all occurrences of P_2 by P_1. So we have

$$C \vdash D_1 \rightarrow P_1(c_1,.,c_n), \text{ or, equivalently:} \qquad (9)$$

$$D_1 \vdash C \rightarrow P_1(c_1,.,c_n), \qquad (10)$$

Also, (7) can be turned into

$$D_1 \vdash P_1(c_1,.,c_n) \rightarrow C, \qquad (11)$$

and (10) and (11) give us

$$D_1 \vdash P_1(c_1,..,c_n) \leftrightarrow C. \qquad (12)$$

Since D_1 is a sentence from L_1, it does not contain any of the constants $c_1,..,c_n$. So (12) entails:

$$D_1 \vdash (\forall x_1)..(\forall x_n)\ (P_1(x_1,.,x_n) \leftrightarrow C'), \qquad (13)$$

where C' is the formula of L which we get by replacing the occurrences of $c_1,.,c_n$ in C by the variables $x_1,.,x_n$. Replacing P_1 in (13) throughout by P gives us

$$D \vdash (\forall x_1) .. (\forall x_n)\ (P(x_1,.,x_n) \leftrightarrow C'), \qquad (14)$$

where D is a sentence from T' and C'' is a formula from L. So

$$T' \vdash (\forall x_1) .. (\forall x_n)\ (P(x_1,..,x_n) \leftrightarrow C') \qquad (15)$$

which shows that P is explicitly definable in T'. q.e.d.

This concludes the proof for the case where α is a predicate. Suppose now that α is an n-place function constant f. We can reduce this case to the case where α is a predicate by replacing f by an n+1-place predicate P, where "$P(x_1,.,x_n,x_{n+1})$" expresses that $f(x_1,.,x_n) = x_{n+1}$. Let P be a symbol not occurring in L and let 'L be the language $(L\backslash\{f\}) \cap \{P\}$. Corresponding to each model $M = \langle U,F\rangle$ for L there is a model $'M = \langle U,'F\rangle$ for 'L, where for any n+1-tuple $\langle u_1,.,u_n,u_{n+1}\rangle$ of elements of U, $('F(P))(\langle u_1,.,u_n,u_{n+1}\rangle) = 1$ iff $(F(f))(\langle u_1,.,u_n\rangle) = u_{n+1}$. Conversely, for any model 'M for 'L there is a model M for L such that 'M corresponds to M in the manner indicated.

Let $^+$ be the translation function from L to 'L defined in Exercise EA2 of the Appendix to Ch. 1. $^+$ translates terms τ into formulas $\tau^+(y)$ and formulas A of L into formulas A^+ of 'L. As shown in EA2, $^+$ has the property that for any model M for L, corresponding model 'M for 'L and assignment **a** in M, $[[\tau^+(y)]]^{'M,\mathbf{a}} = 1$ iff $[[\tau]]^{M,\mathbf{a}} = \mathbf{a}(y)$ and $[[A^+]]^{'M,\mathbf{a}} = [[A]]^{M,\mathbf{a}}$.

Let 'T be the deductive closure of the set of $^+$-translations of the sentences in T: $'T = Cl_{'L}(\{A^+: A \in T\})$. Then it follows from the above remarks about $^+$ that for any model M for L we have $M \models T$ iff $'M \models 'T$, where 'M is the 'L-model corresponding to M. Moreover, the "reduction" of T to L' = L\{f} - i.e. the theory $T' = T \cap \{A: A$ is a sentence of L'} - is the same as the "reduction" of 'T to L'. (Note that the language L' can also be written as 'L\{P}.) From these observations we can infer that P is implicitly definable in 'T. For suppose that M' is a model of T'. Then there is by assumption a unique way to expand M' to a model M of T. It follows from what we have said that the model 'M for 'L coresponding to M is a model of 'T. So there exists an expansion of M' to a model of 'T. Moreover, if there were two different expansions $'M_1$ and $'M_2$ of M' that were both models of 'T, then the corresponding models M_1 and M_2 for L would be also different and they would be expansions of M' that would be both models of T, which would contradict the assumption that f is implicitly definable in T.

Since P is implicitly definable in 'T we can apply Beth's Theorem for the case of predicates and obtain as theorem of 'T an explicit definition for P of the form given in (16).

$$(\forall x_1) .. (\forall x_n)(\forall x_{n+1})(P(x_1,.,x_n,x_{n+1}) \leftrightarrow A) \qquad (16)$$

where A is a formula of the language L'.

At this point we must refer once more to the properties of the translation function $^+$. One further property of $^+$ is that the formula $(f(x_1,.,x_n) = x_{n+1})^+$ is logically equivalent to the formula $P(x_1,.,x_n,x_{n+1})$ and that this equivalence is preserved by logical operations which combine these atomic formulas with each other and with formulas from L' (which are not affected by $^+$). This entails that (16) is logically equivalent to the $^+$-translation of (19).

$$(\forall x_1) .. (\forall x_n)(\forall x_{n+1})(f(x_1,.,x_n) = x_{n+1} \leftrightarrow A) \qquad (17)$$

So since (16) is a theorem of 'T, (17) is a theorem of T.

This concludes the proof of Beth's Theorem.

q.e.d.

There is a striking similarity between Beth's Definability Theorem and the Correctness-and-Completeness Theorem for first order predicate logic. Each theorem states an equivalence between (i) a syntactic and (ii) a semantic condition, and in each case the one condition is existential and the other universal. In our original formulation of the (Correctness and) Completeness Theorem the syntactic condition is existential - **there exists** a proof of B from the premises A_1, .., A_n - and the semantic condition universal - **every** model which verifies A_1, .., A_n also verifies B. Similarly, in the case of Beth's Theorem the syntactic condition - explicit definability, i.e. the existence of an explicit definition of α which is a theorem of T - is existential and the semantic condition - implicit definability, the unique expandability of every model of T' to a model of T - is universal. But we can also turn things around by taking contrapositives. The two conditions connected by the Completeness Theorem are then an existential semantic condition - **there exists** a model which verifies A_1, .., A_n but fails to verify B and a universal syntactic condition - **no** formally correct proof is a proof of B from A_1, .., A_n. Similarly, taking contrapositives in the case of Beth's Theorem turns it into an equivalence statement between an existential semantic condition - **there is** a model of T' that either cannot be expanded to a model of T at all or else can be expanded to a model of T in more than one way - and a universal syntactic condition - **no** explicit definition of α is a theorem of T.

When the Correctness-and-Completeness Theorem is stated as the equivalence between the negated conditions mentioned above - there

exists a "countermodel", in which A_1, .., A_n are true and B is false iff there is no derivation of B from A_1, ..,A_n -, then the hard part (completeness) is to prove that non-existence of a proof of B from A_1, .., A_n entails the existence of a countermodel. The converse - that the existence of a countermodel entails that there is no proof of B from A_1, .., A_n; in other words, the correctness of the given proof procedure - is generally easier (although how easy will depend somewhat on the proof procedure for which correctness and completeness are being proved). In the case of Beth's Theorem the difference between the two directions is even more striking. When there is a model of T' which either has no expansion or else more than one expansion to a model of T, then obviously it cannot be the case that T contains an explicit definition of α as a theorem. It was Beth's striking accomplishment to succeed in proving the converse of that.

In fact, the easy direction of the equivalence between implicit and explicit definition had been known for many years before Beth proved his Theorem. And it was one half of that easy direction - that the non-existence of an explicit definition of α in T can be established by finding a model of T' that can be expanded in more than one way to a model of T - which had gained currency under its own name, viz. as the "Method of Padoa", after the Italian mathematician Alessandro Padoa (1868-1937). It was by pursuing the question whether Padoa's Method was a necessary as well as a sufficient condition for the non-existence of an explicit edfinition of α in T that Beth was led to the proof of his definability theorem.

Internal definability questions - Is, for given T and $\alpha \varepsilon L_T$, α definable in T? - are sometimes easy to answer, but they can also be very hard. Examples of fairly easy questions of this kind we have observed earlier in this Chapter in connection with the Theory of Boolean Lattices and the Theory of Algebras. In the theory T_{bla} of Boolean Algebras given in Section 2.1.3 the operation $\cap$ is definable in terms of $\cup$ and - and, conversely, $\cup$ is definable in terms of $\cap$ and -. To show this is straightforward since in this case explicit definitions are easy to find: $\cap$ is definable in T_{bla} in terms of $\cup$ and - by the definition $(\forall x)(\forall y)(\forall z)(x \cap y = z \leftrightarrow z = -(-x \cap -y))$ and $\cup$ is similarly definable in terms of $\cap$ and - by a definition that is the "dual" of the one just given (i.e. one whose definiens is obtained by replacing in that of the given definition $\cup$ everywhere by $\cap$ and $\cap$ everywhere by $\cup$). We also saw that - is definable in T_{bla} in terms of $\cup$, $\cap$, 0 and 1, viz. by the definition $(\forall x)(\forall y)(-x = y \leftrightarrow (x \cup y = 1 \ \& \ x \cap y = 0))$.

In fact, there are even stronger definability results in this case: (i) the complement operation - is definable just in terms of $\cap$, for instance by the definition

(20) $(\forall x)(\forall y)(-x = y \leftrightarrow (x \cap y = 0 \ \& \ (\forall z)(x \cap z = 0 \rightarrow y \cap z = z)))$

where "$a \cap b = 0$" is short for: "$(\exists u)((\forall v)(u \cap v = u) \ \& \ a \cap b = u)$"

and (ii) - is definable just in terms of $\cup$, for instance by the definition

(21) $(\forall x)(\forall y)(-x = y \leftrightarrow (x \cup y = 1 \ \& \ (\forall z)(x \cup z = 1 \rightarrow y \cup z = z)))$.

(where "$a \cup b = 1$" is a similar abrreeviation as "$a \cap b = 0$")

The reason why (20) is a proper definition of - in T_{bla} is that it is one of the theorems of T_{bla} that for each x there is among the elements y such that $x \cap y = 0$ a unique largest one. Likewise, (21) is a proper definition of - in T_{bla} because T_{bla} has the theorem that for each x there is a unique smallest element y such that $x \cup y = 1$.

For the same reason the pseudo-complement - of pseudo-complemented lattices is definable in terms of $\cup$, $\cap$, 0 and 1. (See Section 2.2.1) For recall that one of the axioms of the theory of pseudo-complemented lattices says that for each x there is a unique largest y such that $x \cap y = 0$. But when the uniqueness requirement is dropped, the possibility of defining "-" in terms of these operations also disappears. More precisely, let T be the theory of the language $\{\cup,\cap,0,1,-\}$ which we get by adding to the axioms of T_{lata} the following sentence, which says that the meet of x and -x is always equal to the minimal element 0:

(21) $(\forall x)\ x \cap -x = 0$

In this theory there is no longer any guarantee that -x is unique and so there is no hope of defining - in terms of $\{\cup,\cap,0,1\}$.
That - is no longer definable can be seen as follows. Let $V = \langle U,\cup,\cap,\mathbf{0},\mathbf{1}\rangle$ be the lattice whose universe U consists of the elements $\{\mathbf{0},\mathbf{1},\mathbf{a}\}$ and the infinite set of elements $\{\mathbf{b}_n : n \in N\}$, where **1** is as always the largest and **0** the smallest element of the lattice and where the operations $\cup$ and $\cap$ are fixed by: (i) for all n, $\mathbf{a} \cup \mathbf{b}_n = \mathbf{1}$ and $\mathbf{a} \cap \mathbf{b}_n =$

0, and (ii) for all n, m such that n ≤ m, $\mathbf{b}_n \cup \mathbf{b}_m = \mathbf{b}_m$ and $\mathbf{b}_n \cap \mathbf{b}_m = \mathbf{b}_n$. Evidently V is a model of the theory T' consisting of those theorems of T that are expressible in the language {∪,∩,0,1}. We can extend V to a model of T in several ways by adding an extension for -. That is, we can choose $F_V(-)$ to be any of the following functions $-_n$ on U. The functions $-_n$ all coincide insofar as (i) $-_n(0) = 1$, (ii) $-_n(1) = 0$ and (iii) for all m, $-_n(\mathbf{b}_m) = \mathbf{a}$. But they differ from each other in the values they return for the argument **a**: for each n, $-_n(\mathbf{a}) = \mathbf{b}_n$. It is easily seen that each function $-_n$ yields a model of T when added to the model V of T'. So there is more than one way to expand V to a model of T.

Note that this argument is an application of Padoa's Method. In fact, to reach the conclusion that - is not definable in T it suffices to consider just two of the functions $-_n$, e.g. $-_0$ and $-_1$.

In the discussion above we have repeatedly used the phrase "α is definable in T in terms of ...", where the ... mention some of the other non-logical constants of L_T, but not necessarily all of them. We have so far only used this turn of phrase in connection with explicit definitions, and there it is immediately clear what is meant: a definition in which the definiens A contains only those non-logical constants that are mentioned in the dot part ...); thus α isn't merely claimed to be definable in the language L \ {α}, but in the sublanguage L' of L \ {α} which consists just of the symbols mentioned in the dor part. It is straightforward to also extend the characterisation of implicit definability to this more general case. All we need to do is to restrict the earlier characterisation of implicit definability to the sub-theories T' and T'' of T in the sublanguages L' and L'', where L' is the sublanguage just mentioned and L'' = L' ∪ {α}. To be precise, the characterisation of implicit definability of α in T in terms of the non-logical constants of L' now takes the following form:

(22) Let T be a theory of the language L. Let α be a non-logical constant of T, let L' ⊆ L\{α} and let L'' = L' ∪ {α}. Let T' = T ∩ {A: A is a sentence of L'} and T'' = T ∩ {A: A is a sentence of L''}.
Then α is said to be *implicitly definable in* T *in terms of* L' iff for each model M' of T' there is a unique expansion M'' of M' that is a model of T''.

It is left as an exercise to the reader to show that the corresponding version of Beth's Theorem holds:

(23) Let T, L and L' as in (22). If α is implicitly definable in T in terms of L', then there exists an explicit definition of α in terms of L' which is a theorem of T.

These generalised characterisations of implicit and explicit definability are convenient in particular in connection with a kind of application which we haven't yet mentioned, but of which there are many instances of the greatest importance. In such applications the focus is on particular structures - or, more precisely, on the descriptions of those structures in particular logical languages. Relevant examples that we have already entcountered are the structure of the rational numbers as described in the language $\{<\}$, and the Tarski Lattices for particular first order languages L as described in the language $\{\cup,\cap,0,1,-\}$.

Given a particular structure and a particular language in which it is described we can ask questions about the definability "within the given structure" of some of the notions represented in the describing language in terms of one or more of the others. Such questions can be phrased as definability questions of the kind we have been asking so far, i.e. as questions about the definability in a first order theory T of one non-logical constant α from the language of T, L_T, in terms of certain others. More specifically, they are questions of the form given in (24), where $\mathbb{S}$ is the structure in question, Th($\mathbb{S}$) is the set of all sentences of L_T that are true in $\mathbb{S}$ and L' is some sublanguage of $L_T \backslash \{\alpha\}$.

(24) Is α definable in the theory Th($\mathbb{S}$) in terms of the non-logical constants of L'?

In the next section we will study two structures that are at the very centre of mathematics. The first of these is "natural number arithmetic", i.e. the structure consisting of the natural numbers with the number null, the successor function S (where S(n) = n+1) and the operations of addition and multiplication; more explicitly, we will study the theory of natural number arithmetic as a theory of first order predicate logic formulated in the "language of Peano Arithmetic" - the first order language $L_{PA} = \{0,S,+,.\}$, where 0 is an individual constant, S a 1-place function constant and + and . are 2-place function constants. The second structure is that of real number arithmetic, i.e. the structure of the real numbers described in the first order language $\{+,.,<,0,1\}$, where + and . are 2-place function constants, < is a 2-place predicate constant and 0 and 1 are individual constants. About these and some other, related structures a range of questions of the general

form (24) can be asked - some easy, some hard and some with answers that have important further consequences.

The "Non-Circularity Requirement"

In the opening paragraphs of this section we promised a few words on the notion of definitional circularity. Many philosophical discussions of definitions make a big thing out of circularity, as something that is bad and should be avoided at all cost. Informally speaking, the basic cocern is something like this: Suppose you define a concept C in terms of certain other concepts C_1, ..., C_n. Suppose moreover that at the same time you define one of the C_i in terms of some further concepts one of which is C. That wouldn't be right, as the second definition would in all likelihood defeat the purpose of the first definition. For suppose you want to use the first definition to determine whether some given entities fall under C; then there is good chance that that will lead you consider whether certain entities, and quite possibly the same ones, fall under C_i. But to determine that you will, in all likelihood, be led to apply the second definition and that may get you involved in turn in questions about what falls under C; in particular, it may lead you back to the very same question that you started with.

We noted that circularity isn't really a topic that can be properly dealt with within the setting we have adopted - that of fully articulated theories formalised within first order logic. The difficulty can be illustrated at the hand of a very simple example. Consider the theory T_{lin} of arbitrary non-trivial linear orderings in the language $\{<,\leq\}$ according to which $<$ and $\leq$ stand in the familiar relation of a strict linear ordering and the corresponding weak ordering. We can axiomatise this theory by means of the axioms L1-L3 of Section 1.2.1 together with the sentences (25.i) and (25.ii).

(25) i. $(\exists x)(\exists y)\ x \neq y$

ii. $(\forall x)(\forall y)(x \leq y \leftrightarrow (x = y \vee x < y))$

Among the theorems of T_{lin} we find on the one hand the definition (25.ii) of $\leq$ in terms of $<$ and on the other - this is just as trivial to show - the definition (26) of $<$ in terms of $\leq$.

(26) $(\forall x)(\forall y)(x < y \leftrightarrow (x \leq y \ \&\ x \neq y))$

An obvious implication of this result is that for any given structure $\mathbb{S}$ which involves some linear ordering of its universe the weak ordering $\leq$ of the universe of $\mathbb{S}$ can be *defined in terms of* the strict ordering $<$ in the sense that (25.i) will be a theorem of the theory Th($\mathbb{S}$) for any first order language which includes the predicate symbols $<$ and $\leq$ and where these symbols are interpreted in $\mathbb{S}$ as $<$ and $\leq$. Conversely, in the same sense of 'define' $<$ can be defined in $\mathbb{S}$ in terms of $\leq$.

To repeat: $\leq$ can be defined for such structures in terms of $<$ and $<$ in terms of $\leq$. Does this mean that there is any circularity involved, of a sort that should be cause for worry? The answer would seem to be an obvious "no". You can define $\leq$ in terms of $<$ or you can define $<$ in terms of $\leq$; either is fine. What you *cannot* do, of course, is at the same time "define $\leq$ in terms of $<$ and $<$ in terms of $\leq$" - not at least if that were to mean that on the one hand you formulate the theory of linear orderings as one which uses $<$ as "primitive" - i.e. as a theory in the language $\{<\}$ - and then add $\leq$ as a defined concept (by extending the language $\{<\}$ to $\{<,\leq\}$ and adding, say, definition (25.ii) as a new axiom) - and also formulate the theory as a theory in the language $\{\leq\}$ and then extend that theory with a definition of $<$ (such as (26)). You have to make a choice: either formulate your theory in the language $\{<\}$ and then, if you wish, add $\leq$ by definition, or else formulate it in the language $\{\leq\}$ and then, if you wish, add a definition for $<$.

Surely the warning to avoid circularity can't be a warning against anything as obviously impossible as constructing a formal theory T whose language L_T is different from what it is. But then, what are the dangers of which we are being warned? To answer this it is important to realise that theory development is in general a very complex and protracted process, which typically runs through a number of successive stages. First, a body of data whose internal connections will often be quite poorly understood at hte outset must be structured into an organic, explanatory whole - into a "theory", in other words - and an essential part of that is to design the concepts in terms of which the central principles of the theory are to be stated. Exactly what these concepts stand for need not be fully clear from the start; often their true meaning will reveal itself only gradually, as the principles which make use of them become more firmly entrenched and their implications better understood (in particular those which link them to the data). Among the means of concept clarification that can be helpful

during this stage of theory development are definitions of one concept in terms of some of the others.[27] And such definitions may be useful even if some of the other concepts occurring in the definiens are still in need of further clarification in their turn. If, however, one then attempts to clarify one of those other concepts by means of a definition that employs the original concept C in its definiens, then that is a sign that something has gone awry. Trying to back a given definition of C with a further definition that makes use of C is a bit like putting up one piece of real estate as collateral when acquiring another, and then offering the second one as a collateral in an attempt to refinance the first. In business this is regarded as a form of fraud. Circular definitions won't land you in jail, but they too are violations of sound general principles and ought to be be avoided.

---

27 It is a remarkable fact that progress can be made in this way at all. Philosophers call this the "Paradox of Analysis": If we understand a conecpt C well enough to be able to judge a proposed definition as a correct definition of C, then how can that definition tell us anything about C that we didn't know already? There are, it would seem, just two possibilities: either we didn't know everything that the definition tells us, but then we are not in a position to recognise the definition as correct; or else we did already know all that it tells us, but then the definition cannot tell us anything about C that is really new to us; the best that it could do would be to give us something that we knew already in a different form. And yet it is undeniable that "explanatory" definitions - definitions of concepts we already have that seem right to us and that nonetheless reveal something new about the concepts they define - do play a sginificant part in theory development, and in concept formation generally.

How can a definition ever be explanatory in this sense? There are no easy answers to this question. But I think it is intuitively clear that any satisfactory answer must have to do in some essential way with the nature of human cognition. A person's thoughts form a complex web of propositional representations in which concepts are the principal building blocks. At the same time some of these concepts are linked to the external world by complex application criteria - criteria that determine for at least some real world entities that they belong to the extension of the concept, and for certain others that they do not, and which also enable us to recognise when this to the case. However, much of this - propositional representations as well as linking criteria - can be *implicit knowledge*: we can apply the criteria without being able to articulate them and we can draw inferences from the network of representations without necessarily being able to name or state all those parts of the network that serve as premises to the inference. Definitions which purport to reduce one concept to a number of others are among the most effective prompts for dragging to the surface of our awareness connections between two or more concepts that up to then were just implicit knowledge. In this way something that was known to us already in some hidden and nebulous way can acquire a new quality - become a "clear and distinct idea", to use Descartes' phrase. This may give us on the one hand the sense that we are learning something new while at the same time we can perceive that "new" piece of knowledge as agreeing with the implicit knowledge we already had As I said, this isn't much of an answer. But I think it indicates the direction in which we should look for one.

It doesn't follow from what has just been said that definitional circularity is a trap that it is easy to fall into. But it doesn't follow either that it is harmless altogether. There are at least two concomitant factors that contribute to the danger of being caught in it. First, definitional circles can be more concealed than they are in the simple case I have mentioned - they may involve not just two, but three definitions (D1 defines C with the help of C', D2 C' with the help of C'' and D3 C'' with the help of C) or even more than three. At a stage where one is still struggling for a better grasp of each of these concepts it is perfectly possible - and legitimiate - for all three definitions to be on the drawing board, each indicating a possible avenue of conceptual clarification. In this context the non-circularity principle can be seen as urging that a choice between those definitions will have to be made eventually: At least one of the definitions will have to be abandoned.

A second contributing factor is that theory development, and the conceptual analysis that is almost always an indispensable part of it in its earlier stages, is usually not a one-person enterprise but one that involves a group of investigators or even a whole scientific community. Different members of the group or community may come up with different definitions for different concepts. Taken together these definitions may well contain loops that no one member of the group or community is aware of; or else, individual members may not even be much concerned by such loops even if they see them, since they feel no commitment to one or more of the definitions involved. Once again, as a temporary state of affairs during the exploratory stage of theory development this situation need not be particularly objectionable. But, of course, by the time the theory has reached its definitive form all loops will have had to be eliminated.

When conceptual clarification has progressed to the point where logical formalisation becomes a meaningful option the explorations and debates that can lead to definitional circularity will normally have come to an end. At that point the hardest conceptual work that goes into developing the given theory will have been done as well. But this does not mean that logical formalisation should be seen as little more than a logician's pass time, from which nothing of substance can be learned that could not have been gathered just as easily from the theory before it is formalised. Within mathematics formalisation has led to numerous results that are not just of interest to formal logicians but are considered important by the community of mathematicians who deal with the branch of mathematics to which the given theory belongs, and who may have no particular interest in formal logic as such. Within the

empirical sciences formalisation has led to many important new insights too. Perhaps the single most important advance that has been achieved in this way within the general domain of empirical sicence is the formalisation of the concept of probability by Kolmogorov (1903-1987) (*About the Analytical Methods of Probability Theory*, 1831). Probability has become a central concept in all the empirical sciences, since it enters almost invariably in evaluating the truth or tenability of scientific hypotheses in the light of relevant data. Kolmogorov's axiomatisation has given us an understanding of the essentials of probability that, it seems fair to say, could not have been reached in any other way.

Proof of the Craig Interpolation Lemma.

Our last act in this section is the promised proof of the Craig Interpolation Lemma. (We remind the reader: an alternative proof can be found in the Appendix to Ch. 1.)

Proof of the Craig Interpolation Lemma.

Let A and B are as in the statement of the Interpolation Lemma and suppose that there is no C of L such that $A \vdash C$ and $C \vdash B$. We extend L to a language L' by adding an infinite sequence $\{c_i\}_{i \varepsilon N}$ of new constants. Similarly we extend, by adding this same set of constants, L_1 to L'_1 and L_2 to L'_2. Let $\{D_{i+1}\}_{i \varepsilon N}$ be an infinite sequence of sentences such that (i) the even-numbered sentences D_{2i} constitute a complete enumeration of the set of all sentences of L'_1 and the odd-numbered sentences D_{2i+1} a complete enumeration of the set of all sentences of L'_1. We proceed in a way reminiscent of the completeness theorem, extending once more given consistent sets in an infinite number of steps to maximal consistent sets. However this time we extend two sets in tandem and it is not just the consistency of the individual sets that we are interested in, but a kind of mutual consistency between them. More precisely, we generate two infinite sequences, a sequence $\{\Delta_{1i}\}_{i \varepsilon N}$ of finite but growing sets of sentences from L'_1 and a sequence $\{\Delta_{2i}\}_{i \varepsilon N}$ of finite but growing sets of sentences from L'_2. At each stage the pair $< \Delta_{1i}, \Delta_{1i}>$ is "compatible modulo L' " in the following sense:

(1) there is no sentence C of L' such that (i) $\Delta_{1i} \vdash C$ and (ii) $\Delta_{2i} \vdash \neg\neg C$.

Note that if (1) holds, then both Δ_{1i} and Δ_{2i} are consistent. For suppose e.g. that Δ_{1i} were inconsistent. Then $\Delta_{1i} \vdash \perp_{L'}$, where $\perp_{L'}$ is some logical contradiction of L'; but then we would also have $\Delta_{2i} \vdash \neg\neg \perp_{L'}$, which would contradict (1). Consistency of Δ_{2i} is entailed for the same reason.

Our initial sets are singletons: $\Delta_{1o} = \{A\}$ and $\Delta_{2o} = \{\neg\neg B\}$, and our first task is to verify that these two satisfy (1). This, however, follows directly from the reductio assumption we have made about A and B.

The construction of the sequences $\{\Delta_{ji}\}$ proceeds as follows: At the even steps 2.i we operate on the set $\Delta_{1,2.i}$ and at the odd steps 2.i + 1 we operate on the set $\Delta_{2,2.i+1}$. We will state the rules according to which the sets are modified only for the even steps. The case for the odd steps is entirely symmetric.

Step 2.i:

Consider $D_{2.i}$. (i) When (1) holds for $\Delta_{1,2.i} \cup \{D_{2.i}\}$ and $\Delta_{2,2.i}$, then we add $D_{2.i}$ to $\Delta_{1,2.i}$, and, as in the Completeness Proof, we add, in case $D_{2.i}$ is an existential sentence $(\exists v_i)E$, then we also add a "witness sentence" $E[c_k/ v_i]$, where c_k is a constant which does not occur in either $\Delta_{1,2.i}$ or $\Delta_{2,2.i}$. Much as in the case of the Completeness Proof we can show that (1) is preserved also in the case where $\Delta_{1,2.i+1} = \Delta_{1,2.i} \cup \{D_{2.i}, E[c_k/ v_i]\}$, given that it holds for the pair $\langle\Delta_{1,2.i} , \Delta_{2,2.i}\rangle$.

(ii) When (1) does not hold for $\Delta_{1,2.i} \cup \{D_{2.i}\}$ and $\Delta_{2,2.i}$, then we add $\neg\neg D_{2.i}$ to $\Delta_{1,2.i}$: $\Delta_{1,2.i+1} = \Delta_{1,2.i} \cup \{\neg\neg D_{2.i},\}$.

We need to show that in each of the three cases condition (1) is preserved. Case (i) is automatic in case $D_{2.i}$ is not an existential sentence. Suppose instead that $D_{2.i}$ is the sentence $(\exists v_i)E$. In that case is $\Delta_{1,2.i+1} = \Delta_{1,2.i} \cup \{(\exists v_i)E, E[c_k/ v_i]\}$, with c_k a constant not previously used. Suppose that (1) fails for $\Delta_{1,2.i+1}$ and $\Delta_{2,2.i+1} = \Delta_{2,2.i}$. Then there is a sentence C of L' such that

(2) (i) $\Delta_{1,2.i} \cup \{(\exists v_i)E, E[c_k/ v_i]\} \vdash C$, and
(ii) $\Delta_{2,2.i} \vdash \neg\neg C$

From (2.i) we get that there is a sentence G ε $\Delta_{1,2.i}$ such that

(3) $\{G, (\exists v_i)E\} \vdash E[c_k/ v_i] \rightarrow C(c_k)$

(Here we have made explicit that C may contain the constant c_k.)

Since c_k does not occur in $\{G, (\exists v_i)E\}$, (3) entails

(4) $\{G, (\exists v_i)E\} \vdash (\forall v_i)(E \rightarrow C(v_i)/c_k))$, and from this

(5) $\{G, (\exists v_i)E\} \vdash (\exists v_i)E \rightarrow (\exists v_i)C(v_i)/c_k)$,

(5) evidently entails

(6) $\{G, (\exists v_i)E\} \vdash (\exists v_i)C(v_i)/c_k)$.

On the other hand, since does not contain c_k, (2,ii) entails

(7) $\Delta_{2,2.i} \vdash (\forall v_i)\neg C(v_i)/c_k)$, and thus

(8) $\Delta_{2,2.i} \vdash \neg(\exists v_i)\, C(v_i)/c_k)$.

Thus $(\exists v_i)\, C(v_i)/c_k)$ is a sentence of L' which is provable from $\Delta_{1,2.i}$, while its negation is provable from $\Delta_{2,2.i}$. This contradicts the assumption that (1) holds for $\Delta_{1,2.i}$ and $\Delta_{2,2.i}$.

Case (ii) is also somewhat different from the corresponding argument in the completeness proof. The argument now takes the following form.

Suppose that (1) fails for $\Delta_{1,2.i} \cup \{\neg D_{2.i}\}$ and $\Delta_{2,2.i}$. Then

(9) there is a sentence C of L' such that $\Delta_{1,2.i} \cup \{\neg D_{2.i}\} \vdash C$ and $\Delta_{2,2.i} \vdash \neg C$.

Recall, however, that in this case we also have a failure of (1) for the pair $<\Delta_{1,2.i} \cup \{D_{2.i},\}, \Delta_{2,2.i}>$. So

(10) there is a C' of L' such that $\Delta_{1,2.i} \cup \{D_{2.i}\} \vdash C'$ and $\Delta_{2,2.i} \vdash \neg C'$.

(9) and (10) entail on the one hand that $\Delta_{2,2.i} \vdash \neg C$ and $\Delta_{2,2.i} \vdash \neg C'$ and thus that $\Delta_{2,2.i} \vdash \neg (C \vee C')$. On the other hand $\Delta_{1,2.i} \cup \{\neg D_{2.i}\} \vdash C$ entails $\Delta_{1,2.i} \vdash \neg D_{2.i} \rightarrow (C \vee C')$ and $\Delta_{1,2.i} \cup \{D_{2.i}\} \vdash C'$ entails $\Delta_{1,2.i} \vdash D_{2.i} \rightarrow (C \vee C')$.

These last two consequence relations jointly entail $\Delta_{1,2.i} \vdash (D_{2.i} \vee \neg D_{2.i}) \rightarrow (C \vee C')$ and thus also $\Delta_{1,2.i} \vdash (C \vee C')$. So there is a sentence C'' of L' (viz. C v C') such that $\Delta_{1,2.i} \vdash C''$ and $\Delta_{2,2.i} \vdash \neg C''$. So (1) fails for the pair $<\Delta_{1,2.i}, \Delta_{2,2.i}>$, contrary to assumption.

We now form $\Delta_1 = \cup \{\Delta_{1i}\}_{i \in N}$ and $\Delta_2 = \cup \{\Delta_{2i}\}_{i \in N}$. Much as in the proof of the Completeness Theorem, we can show that (1) holds for the pair $<\Delta_1, \Delta_2>$. This entails, as we have seen, that Δ_1 is a maximal consistent theory of L'_1 and that Δ_2 is a maximal consistent theory of L'_2. So, again as in the Completeness Proof, we can convert Δ_1 into a model M_1 for the language L'_1 which verifies precisely the sentences of Δ_1, and Δ_2 into a model M_2 for the language L'_2 which verifies precisely the sentences of Δ_2. We note the following: M_1 and M_2 have the same universe. For recall that if we proceed as in the Completeness Proof, then the universe U_1 of M_1 consists of equivalence classes $[c_i]_{\sim 1}$, where c_i is one of the new constants of L' and $\sim_1$ is the relation which holds between constants c_i and c_j iff the sentence $c_i = c_j$ belongs to Δ_1. Similarly, the universe U_2 of M_2 consists of equivalence classes $[c_i]_{\sim 2}$, where c_i is one of the new constants of L' and $\sim_2$ is the relation which holds between constants c_i and c_j iff the sentence $c_i = c_j$ belongs to Δ_2. It is to be stressed that in the present construction we take the elements of the universes U_1 and U_2 to consist *solely* of the new constants $c_i \in L' \setminus L$. (This means that we need a separate clause to deterine the denotations of the individual constants $c \in L_1 \cup L_2$. But this is unproblematic. For instance, assume that $c \in L_1$. Then there is a constant $c_i \in L' \setminus L$ such that $c_i = c \in \Delta_1$, i.e. $c \sim_1 c_i$. In this case we can unambiguously stipulate that $c_{M_1} = [c_i]_{\sim 1}$.) This entails that the two universes are in fact identical, since the relations $\sim_1$ and $\sim_2$ coincide. To see this, suppose for instance that $c_i \sim_1 c_j$. Then $c_i = c_j \in \Delta_1$. But then also $c_i = c_j \in \Delta_2$. For if not, then, by maximality of Δ_2, $c_i \neq c_j \in \Delta_2$. But then there would be a sentence C of L' (viz. $c_i = c_j$), such that $\Delta_1 \vdash C$ and $\Delta_2 \vdash \neg C$. So (1) would fail for $<\Delta_1, \Delta_2>$, which we

know already that (1) holds for these sets. Since $c_i = c_j \varepsilon \Delta_2$, $c_i \sim_2 c_j$. In the same way we show that if $c_i \sim_2 c_j$, then $c_i \sim_1 c_j$.

Not only do M_1 and M_2 have the same universes, they also assign the same interpretations to each of the non-logical constants of L'. For the new constants c_i this is immediate: $[c_i]_{M1}$ is the equivalence class $[c_i]_{\sim 1}$ and $[c_i]_{M2}$ is the equivalence class $[c_i]_{\sim 2}$, but these equivalence classes are the same. Now consider any non-logical constant α of L. Let us for simplicity assume that α is a 1-place predicate P. From the construction of M_1 we know that the extension of P in M_1, $[P]_{M1}$, consists of those equivalence classes $[c_i]_{\sim 1}$ such that the sentence $P(c_i) \varepsilon \Delta_1$. And by the same token, $[c_i]_{\sim 2} \varepsilon [P]_{M1}$ iff the sentence $P(c_i) \varepsilon \Delta_2$. But again we can infer from the fact that (1) holds for $<\Delta_1, \Delta_2>$ that $P(c_i) \varepsilon \Delta_1$ iff $P(c_i) \varepsilon \Delta_2$. For if not then we would have, say, $P(c_i) \varepsilon \Delta_1$ and $\neg P(c_i) \varepsilon \Delta_2$, so $P(c_i)$ would be a sentence C of L' contradicting (1). For non-logical constants of other types the argument is analogous.

We thus conclude that the reduction of M_1 to L' is identical with the reduction of M_2 to L'. This means that we can form the common expansion M_3 of M_1 and M_2 in that we add to their common reduction (i) the interpretations in M_1 of the non-logical constants of $L_1 \backslash L$ and (ii) the interpretations in M_2 of the non-logical constants of $L_2 \backslash L$. Since M_1 is the reduction of M_3 to L'_1, A, which is a sentence of L'_1, will have the same truth value in M_3 as in M_1. So A is true in M_3. An analogous argument shows that $\neg B$ is true in M_3. But this contradicts the assumption that $A \vdash B$. q.e.d.

2.6. Formalisations of Arithmetic

The first theory we looked at in this chapter aimed at giving as accurate a description as possible of one particular structure, viz. the ordering of the rationals. In that case our effort was as successful as a first order description of an infinite structure can be: the theory T_{rat} we formulated proved to be not only complete - in the sense that it captured as theorems all that can be said truly about that structure in the given first order language {<} in which T_{rat} was formulated - it even proved to be categorical in the cardinality of the target structure; every countably infinite model of T_{rat}, we found, is isomorphic to the ordering structure of the rationals.

The theories we have been looking at since then - lattices, distributive lattices, boolean algebras, groups - have for the most part been incomplete, and they were meant to be that. The aim of those theories was to capture what is common to a whole range of similar but non-identical structures, many of which differ from each other in ways that can actually be expressed in the language of the theory. In such cases the common core - the theory which consists of all sentences of the given language that are true in all the structures - is necessarily incomplete. It was only in a few cases - when we considered the theories of some particular orderings such as the ordering of the integers and that of the natural numbers or the theories of the Tarski Lattices of particular first order languages - that we were confronted once again with questions of the form: "What is the theory of *this* particular structure?

In this section we will focus once again on axiomatisation tasks connected with particular structures. We will be concerned with axiomatisations of two structures that occupy a central place in both pure and applied mathematics: (i) the structure of 'natural number arithmetic', i.e. the structure consisting of the natural numbers with the arithmetical operations + and . ; and (ii) the structure of 'real number arithmetic', i.e. the structure of the real numbers, also with the arithmetical operations + and . . The main results about axiomatisability of these two strutures are strikingly different, and at first sight they seem to contradict each other. The axiomatisation we will give for arithnmetic on the natural numbers will be, like any other axiomatisation for natural number arithmetic, incomplete and undecidable; these are the famous incompleteness and undecidability results for natural number arithmetic that we owe to Gödel. (Gödel's resuts will not be proved in this chapter). On the other hand, the axioms that we will give for arithmetic on the real numbers provide us with a complete axiomatisation of this kind of artihmetic. (For this result an explicit proof will be given here.)

How can this be, one might be tempted to ask? For it would seem obvious that airhtmetic on the real numbers is much richer than arithmetic on the natural numbers? and that the first includes the second as a part (and as a comparatively small and simple part at that). To put this intuition into a more concrete form: Couldn't one determine whether any arbitrary statement of natural number arithmetic is true by interpreting it as a statement of real number arithmetic (which speaks only of a small part of the real numbers, viz. the natural numbers) and then either derive or else refute this statement (as a statemrent about the reals) from our complete axiom

system for real number arithmetic? That would give as a decision method for number-theoretic truth; but that, Gödel proved, cannot be. The just mentioned results about natural and real number arithmetic thus entail that tstatements about antural number arithmetic cannot be interpreted as statements about the arithmetic of the real numbers. But why not? One of our tasks in this chapter will be to elucidate this apparent contradition.

2.6.1 The Natural Numbers and Peano Arithmetic.

The arithmetical structure $\mathbb{N}$ of the natural numbers consists of the numbers 0,1,2, ... ad infinitum, with the familiar operations of addition and multiplication. Our task in this subsection is to describe this structure by means of a first order theory.

Our first decision is to choose a suitable language. As we have seen repeatedly in this chapter, there usually is a certain freedom regarding this choice: We can choose one set of 'primitives' and then define the missing members of some other set in terms of them, or we can choose the other set and use those to define the missing members of the first set. Also it is not always desirable to keep the set of primitives as small as possible; it can be more perspicuous to choose a larger set, some members of which could also be defined in terms of the others and thus could have been dispensed with in principle.

This is the case for the language we will adopt for the description of $\mathbb{N}$. With the help of the operations + and . we can, given the right axioms, define a number of other notions, such as that of the number 0 (the unique number x with the property that for any number y, $y + x = y$); the number 1 (the unique number y such that $y \times y = y$); the successor function S, which assigns to each number the next one after it (this is the function which maps each number x onto $x + 1$), and the relation $<$ (which holds between x and y iff there is number $z \neq 0$ such that $y = x + z$). So none of these are absolutely indispensable. However, it has become standard practice to include both the constant 0 and the successor function S among the non-logical constants of the language of natural number arithmetic. Quite often the relation $<$ is included as well, but we won't do that here. So the language L_{PA} in which we will describe $\mathbb{N}$ has besides the 2-place function constants + and . the individual constant 0 and the 1-place function constant S, and that is it: $L_{PA} = \{0,S,+,.\}$.

In the literature on formal natural number arithmetic the following axiom set has gained wide currency. It is known as '(First Order) Peano Arithmetic', after the Italian mathematician Giuseppe Peano (1958-1932), who first formulated a set of axioms much like these. We refer to the theory axiomatised by PA1-PA7 simply as 'PA'.

PA1. $(\forall x)\ (x \neq 0 \leftrightarrow (\exists y)\ x = Sy)$[28]

PA2. $(\forall x)(\forall y)\ (Sx = Sy \rightarrow x = y)$

PA3. $(\forall x)\ x + 0 = x$

PA4. $(\forall x)(\forall y)\ x + Sy = S(x + y)\ )$

PA5. $(\forall x)\ \ x \ . \ 0 = 0$

PA6. $(\forall x)(\forall y)\ (x \ . \ Sy = (x \ . \ y) + x)$

PA7. $(\forall y_1)..(\forall y_n)((A[0/x] \ \& \ (\forall x)(A \rightarrow A[S(x)/x])) \rightarrow (\forall x)A)$, where y_1, .., y_n are all the variables other than x which have free occurrences in A.

The rationale behind these axioms is as follows. The first two concern only the constant 0 and the function S and say that 0 is the only element that is not in the range of S and that S is 1-1. These axioms guarantee that 0 is the first of an infinite series of elements 0, S0, SS0, .. all of which are different from each other, and thus that all models of the axioms will be infinite. The next four axioms 'recursively define' the operations of addition and multiplication - PA3 and PA4 do this for +, PA5 and PA6 build on this definition in the recursive definition for . . The specifications of these axioms can be regarded as recursive definitions in that they specify an algorithm for computing the results of these operations, reducing all instances ultimately of cases involving 0. Thus PA3 and PA4 define n + m for any two numbers n and m, by reducing the result via n + (m-1), n + (m-2), ... eventually to n + 0. Likewise for PA5, PA6 and the terms 'n.m'.

This leaves PA7. Here, for the first time, we are dealing not with a single axiom, but with an *axiom schema,* which can be instantiated to an infinite number of different axioms by substituting different formulas of L_{PA} for the schematic letter A. The idea behind this schema is the following. The structure of the natural numbers makes it possible to prove that all natural numbers have a certain property P by mathematical induction: Show (i) that 0 has P and (ii) that for any

---

28 Where there is no danger of confusion we will wríte 'St' instead of 'S(t)'. Note that in the literature one often uses a prime ' instead of S. Thus one writes ' t' ' instead of 'S(t)'. Thus, in partocular the term " 0' " will be a term denoting the number 1.

number x that has P Sx also has P. That it follows from (i) and (ii) that every natural number must have P can be argued in a number of different (if fairly closely similar) ways. One informal argument goes like this: (i) tells us that 0 has P. From this and one application of (ii), taking x to be 0, we get that 1 has P. From this and a second application of (ii), now taking x to be 1, we get that 2 has P, and so forth. In this way we eventually reach every number n and establish that n has P.

Peano recognised that the Principle of Induction - that (i) and (ii) suffice to show that all numbers have P irrespective of what P may be - is one of the central characteristics of the natural number system. And he made it into the corner stone of his axiomatisation of $\mathbb{N}$. PA7' states this principle with the force he intended it to have, but in the notation of formal logic as we know it today.

PA7' $(\forall P)(P(0) \ \& \ (\forall x)(P(x) \rightarrow P(Sx)) \rightarrow (\forall x)P(x))$

The problem with (1) is that it is not a formula of first order logic. It isn't because it quantifies over the predicate symbol P. This means that P is a predicate variable and predicate variables are not part of first order logic. They are part of what is called 'Second Order Logic', a very powerful extension of First Order Logic in which we can quantify not only over individuals but also over sets of individuals. Second Order Logic has formal properties that are very different from those of First Order Logic.

We can use PA7' to obtain an axiomatisation of $\mathbb{N}$ within second order logic in which the other axioms are PA1-PA6. In one sense this axiom system is the perfect answer to our desire for an exhaustive description of the properties of $\mathbb{N}$. For it has the property that any model of it is isomorphic to $\mathbb{N}$. To see that this is so, we first need to make explicit what is meant by a predicate quantification like that in PA7'. The standard semantics of quantifications over predicate variables is that for any set X of individuals of the model M in which the formula containing the quantification is evaluated there is a predicate that can be a value for the variable and which has X as its extension in M. This means that predicate quantification comes to the same thing as quantification over sets, more precisely: over arbitrary subsets of the universe of the model. In particular, PA7' can be stated equivalently in the form PA7''.

PA7'' $(\forall X)(0 \in X \ \& \ (\forall x)(x \in X \rightarrow Sx \in X) \rightarrow (\forall x)\ x \in X)$

Given this interpretation of the quantification in PA7', we can argue as follows. Let M = <U,F> and M' = <U',F'> be two models of {PA1-PA6, PA7'}. Consider the universe U of M. It contains denotations in M of all the terms 0, S0, SS0, .. of L_{PA}. (We will refer to these terms as the *numerals* of L_{PA}. Thus a numeral is a term in which the constant 0 is preceded by some number n of occurrences of the function constant S, where $n \geq 0$.) Let us denote the element of U that is denoted in M by the term 'S...S0', in which '0' is preceded by n occurrences of 'S', as n_M. Let N_M be the set of all $u \in U$ that are denotations of numerals:

$$N_M = \{u \in U: u = n_M \text{ for some natural number } n\}$$
$$(= \{u \in U: \text{there is numeral } \nu \text{ of such that } u = [[\nu]]^M\})$$

Then N_M is the extension in M of a possible value for the predicate variable P in PA7'. It is clear that when P ia assigned this extension in M, then the formulas P(0) and $(\forall x)(P(x) \rightarrow P(Sx))$, which form the antecedent of the conditional in PA7', are satisfied in M. So it follows that the consequent of the conditional is satisfied as well, i.e. $(\forall x)P(x)$. But that means that every element of the model belongs to N_M and thus is the denotation of some numeral.

This argument is just as applicable to M', so its universe too consists of all and only the elements that are denotations of numerals. Given this it is easy to define an isomorphism h from M onto M': for every numeral ν, $h([[\nu]]^M) = [[\nu]]^{M'}$. (It follows from the argument above that h is well-defined and onto, from PA1 and PA2 that h is 1-1, from the definition of 'numeral' that h preserves S and from PA3-PA6 that h preserves + and . .)

This means that the theory PA^2 axiomatised by {PA1-PA6, PA7'} is semantically complete: For any sentence A of L_{PA} we have either $PA^2 \models A$ or $PA^2 \models \neg A$. (For either $\mathbb{N} \models A$, but then, since all models of PA^2 are isomorphic to $\mathbb{N}$, for all M such that $M \models PA^2$, $M \models A$; or else $\mathbb{N} \models \neg A$, and so for all M such that $M \models PA^2$, $M \models A$. But unfortunately this is not much help in deciding which sentences are true in $\mathbb{N}$ and which are false. For second order Logic has no complete proof procedure - there is no completeness result for Second Order Logic comparable to the completeness of First Order Logic we proved in Ch. 1. In fact, it follows from Gödel's Incompleteness Theorems that there can be no sound and complete proof procedure for second Order Logic, for then we would have a decision procedure for natural number arithmetic: to decide

whether a sentence A is true or false in $\mathbb{N}$, launch a simultaneous search for a proof of A from PA^2 and a proof of $\neg A$ from PA^2 and go on until one or the other is found; this must happen at some point in a systematic proof search, since one of A and $\neg A$ must be derivable from PA^2. But what Gödel proved is that there cannot be such a decision procedure.

PA7' is thus too much of a good thing. If we want to stay within First Order Logic, which does have completeness, the best we can do is to save from PA7' as much as can be expressed in first order terms. Presumably PA7 is the best one can do towards this end (though it seems hard to turn this intuition into a well-defined statement that we might be able to demonstrate formally). PA7 saves from PA7' all those cases in which the value of the predicate variable P is a property that is defined by some formula A(x) of the language L_{PA}. We write 'A(x)' to indicate that we think of x as the 'predicate bearer': A(x) is to be understood as the predicate that is true of an individual d in a model M iff $M \models A(x)[d]$. This means that the interesting cases are those where A has free occurrences of x. (If x does not occur free in A, then the 'predicate' A is either true of all individuals in the model or else of none.) On the other hand we allow A to have other free variables besides x. This form of PA7 is more comprehensive and thus gives a stronger axiom system. In some inductive proofs this extra strength is actually needed and in many others, where it could strictly speaking be avoided, it can be quite convenient. We will see an example of this in our sample derivation below.

To prove general properties of the natural numbers from the Peano axioms almost always involves induction, and thus an appeal to one or more instances of PA7. As an example we derive the 'commutative law for +', i.e. the sentence $(\forall x)(\forall y)\ y + x = x + y$. This is a very simple statement, which most people - and in particular non-mathematicians - would be inclined to think hardly worth attention. But even the derivation of this intuitively simple law takes some doing.

The strategy we will follow is the following. We will apply induction to the property that is expressed by the formula $A(x) \equiv (\forall y)\ x + y = y + x$. That is, we use the following instance of PA7:

$$((\forall y)\ 0 + y = y + 0\ \ \&\ (\forall x)((\forall y)\ x + y = y + x\ \rightarrow\ (\forall y)\ Sx + y = y + Sx)) \rightarrow (\forall x)(\forall y)\ x + y = y + x \qquad (1)$$

To derive the consequent $(\forall x)(\forall y)\ x + y = y + x$ of (1) we must prove the two conjuncts that make up its antecedent. We begin with the first conjunct:

$$(\forall y)\ 0 + y = y + 0 \qquad (2)$$

According to PA3 $y + 0 = y$. But how do we prove that $0 + y = y$? This requires another induction, this time wrt. y. To this end we use the following instance of PA7. (Of course this involves renaming variables, but we know we can always do that in the sense that every sentence logically entails all of its alphabetic variants. See Section 1.2.? of Ch. 1)

$$0 + 0 = 0 + 0\ \&\ (\forall y)(0 + y = y + 0 \rightarrow 0 + Sy = Sy + 0) \rightarrow (\forall y)\ 0 + y = y + 0 \qquad (3)$$

To prove the antecedent of (3) first observe that its first conjunct - $0 + 0 = 0 + 0$ - is a logical truth. To prove the second conjunct,

$$(\forall y)(0 + y = y + 0 \rightarrow 0 + Sy = Sy + 0), \qquad (4)$$

assume that $0 + y = y + 0$. We must show that $0 + Sy = Sy + 0$. We argue as follows. $0 + Sy =_{(PA4)} S(0 + y) =_{(Ass)} S(y + 0) =_{(PA3)} Sy =_{(PA3)} Sy + 0$. This shows that $0 + y = y + 0 \rightarrow 0 + Sy = Sy + 0$ and so by Universal Generalisation we get (4). From (3) and (4) we get (2) by M.P.

We now turn to the second conjunct of (1):

$$(\forall x)((\forall y)\ x + y = y + x \rightarrow (\forall y)\ Sx + y = y + Sx) \qquad (5)$$

Suppose that

$$(\forall x)((\forall y)\ x + y = y + x \qquad (6)$$

We must derive from this

$$(\forall y)\ Sx + y = y + Sx \qquad (7)$$

Take any y. By PA4 we have $y + Sx = S(y + x)$, which by assumption (6) equals $S(x + y)$, which by another application of PA4 equals $x + Sy$. But unfortunately $y + Sx = x + Sy$ is not what we want; what we want is

y + Sx = Sx + y. There is nothing for it but to prove the missing equality, x + Sy = Sx + y, separately, and that requires yet another induction.

In other words we must prove

$$(\forall x)(\forall y)\ x + Sy = Sx + y \qquad (8)$$

It is now convenient to take some arbitrary x and prove by induction that

$$(\forall y)\ x + Sy = Sx + y \qquad (9)$$

This requires another induction and thus another instance of PA7, to wit

$$(\forall x)((x + S0 = Sx + 0\ \&\ (\forall y)(x + Sy = Sx + y \rightarrow x + SSy = Sx + Sy)) \rightarrow (\forall y)\ x + Sy = Sx + y)^{29} \qquad (10)$$

(10) entails the free variable formula (11)

$$x + S0 = Sx + 0\ \&\ (\forall y)(x + Sy = Sx + y \rightarrow x + SSy = Sx + Sy) \rightarrow (\forall y)\ x + Sy = Sx + y \qquad (11)$$

To prove (9) from (11) we have to prove the antecedent of (11). Its first conjunct is straightforward:

$$x + S0 =_{(PA4)} S(x + 0) =_{(PA3)} Sx =_{(PA3)} Sx + 0$$

To prove the second conjunct,

$$(\forall y)(x + Sy = Sx + y \rightarrow x + SSy = Sx + Sy), \qquad (12)$$

assume that x + Sy = Sx + y in order to show that x + SSy = Sx + Sy. We have:

$$x + SSy =_{(PA4)} S(x + Sy) =_{(Ass)} S(Sy + x) =_{(PA4)} Sx + Sy$$

This shows (12). From (12) together with the first conjunct of (11) we get (9) and from this by Universal Generalisation (8). We already saw

---

29 Here we make use of the strong form of PA7, according to which A(x) may have free variables other than the "induction variable" x,

that with the help of (8) we can complete our derivation of (7) from (6). This completes the proof of (5) and thus of the second conjunct of the antecedent of (1). (5) and (2) give us the desired conclusion $(\forall x)(\forall y)\ y + x = x + y$.

q.e.d.

Exercise: Give a complete formal derivation of this result from the axioms of PA, using the rules of MP and UG, together with the axioms of predicate logic and previously proved logical theorems.

Exercise: Prove from PA1-7 the following theorems:

(i) $(\forall x)(\forall y)\ (x + y = y + x)$

(ii) $(\forall x)(\forall y)(\forall z)\ ((x + y) + z = (x + (y + z))$

(iii) $(\forall x)(\forall y)\ (x \cdot y = y \cdot x)$

(iv) $(\forall x)(\forall y)(\forall z)\ ((x \cdot y) \cdot z = (x \cdot (y \cdot z))$

(v) $(\forall x)(\forall y)(\forall z)\ ((x + y) \cdot z = (x \cdot z) + (y \cdot z))$

(vi) $(\forall x)(\forall y)\ (x = y \vee (\exists z)(z \neq 0 \ \&\ x = y + z) \vee (\exists z)(z \neq 0 \ \&\ y = x + z))$

Exercise: In PA we can define the order relation between the natural numbers by: $(\forall x)(\forall y)(x < y \leftrightarrow (\exists z)\ x + Sz = y)$.

(a) Show that for any numbers n and m, n is less than m (in the standard sense) iff $\mathbb{N} \models ((\exists z)\ x + Sz = y)[n,m]$.

(b) Interpreting 'x < y' as an abbreviation for '$(\exists z)\ x + Sz = y$' prove that the following are theorems of PA:

(i) $(\forall x)(\forall y)(x < y \rightarrow \neg(y < x))$

(ii) $(\forall x)(\forall y)(\forall z)(x < y \ \&\ y < z \rightarrow x < z)$

(iii) $(\forall x)(\forall y)(x < y \ \vee\ x = y \ \vee\ y < x)$

(iv) $(\forall x)(\forall y)(x < Sy \leftrightarrow (x = y \ \vee\ x < y))$

(v) $(\forall x)(\forall y)(\forall z)(x < y \leftrightarrow Sx < Sy)$

(vi) $(\forall x)(\forall y)(\forall z)(x < y \leftrightarrow x + z < y + z)$

Induction and Well-Foundedness

The validity of the method of mathematical induction rests on the fact that the "less than" relation between natural numbers is *well-founded*. This means that every non-empty set of natural numbers has a smallest member, a number such that no other number in the set is less than it. WF, in which X is assumed to range over subsets of N, expresses this fact formally.

(WF) $(\forall X)\ (X \neq \emptyset \rightarrow (\exists z)\ (z \in X\ \&\ (\forall u)(u \in X\ \&\ u \neq z \rightarrow \neg\ (u < z))))$[30]

WF entails the Principle of Induction. Consider for instance the 'subsets of N' version of the principle PA7". That PA7" follows from WF is easy to show. Suppose that X is a subset of N such that (i) $0 \in X$ and (ii) $\forall x)(x \in X \rightarrow Sx \in X)$. Suppose for the sake of arriving at a contradiction that it is not the case that X = N. Then the set Y = N\X is non-empty. So according to WF Y has a smallest member y_o. Since by assumption $0 \in X$, $y_o \neq 0$. So y_o must be a successor, i.e. there must be a number z such that $y_o = Sz$; obviously this entaikls that $z < y_o$. Since y_o is the smallest number of Y, z is not a member of Y and therefore a member of X. But then by property (ii) of X Sz - i.e. y_o - must also be in X and thus not in Y; and with thia we have our contradiction.

The relation between well-foundedness and the validity of the method of proof by induction holds more generally. First, well-foundednes is a property that can be defined for arbitrary strict partial orderings.

Def. 15 Let <U,< > be a strict partial ordering.
<U,< > is *well-founded* iff every non-empty subset of U has a minimal element. Formally:

$$(\forall X \subseteq U)(X \neq \emptyset \rightarrow (\exists z)\ (z \in X\ \&\ (\forall u)(u \in X\ \&\ u \neq z \rightarrow \neg\ u < z)))$$

To this general notion of well-foundedness corresponds a more general induction principle on partial orderings, As in Def, 15 let <U,< > be a strict partial ordering.

(GIP) $(\forall X \subseteq U)((\forall x \in U)((\forall y \in U)(y < x \rightarrow y \in X) \rightarrow x \in X) \rightarrow U \subseteq X)$

Prop. 6 GIP holds for all well-founded strict partial orderings.

---

30 For the 'definition' of "<" see the Exercise at the end of the last section.

Prop. 6 can be proved in the same way as the special case we considered above where <U, < > was the ordering of the natural numbers.

The converse of Prop. 6 also holds: If GIP holds for <U,< >, then <U, < > is well-founded. The proof is left to the reader.

Well-foundedness is equivalent to the non-existence of infinite descending chains. An i*nfinite descending chain in* a strict partial ordering <U,< > is a function f from the set of the natural numbers N into U such that for all n f(n+1) < f(n). Clearly, well-foundedness of <U,< > entails the non-existence of such chains. For if there were such a chain, then Ran(f) would be a non-empty set without a first element. Conversely, if <U,< > is without infinite descending chains, then <U,< > must be well-founded. For suppose <U,< > were not well-founded. Then there would be a non-empty subset X of U without a minimal element. Let x_o be any element of X. We put $f(0) = x_o$. Since x_o is not a minimal element of X, there is an element x_1 in X such that $x_1 < x_o$. Put $f(1) = x_1$. Since x_1 is not minimal, there must be an element x_2 in X such that $x_2 < x_1$. So we can put $f(2) = x_2$; and so on ad infinitum. In this way we obtain an infinite descending chain f. (Warning: this second argument involves the Axiom of Choice. See Ch. 3 for discussion.)

Inductive proofs on well-founded partial orderings are very common in formal logic. We already encountered many examples of this, in particular in all those cases where we found it necessary or convenient to prove results by induction on the complexity of formulas. The partial order invoked in those proofs is that which holds between two grammatical expressions whenever the first is a constituent of the second. That such relations are always well-founded is plain: The easiest way to see this is to consider a well-formed expression together with all its syntactic constituents, Obviously there are no infinite descending chains of expressions, no infinite sequences of expressions in which each next element is a constituent of the last one. For each expression is built from basic expressions in a finite number of steps; so when we decompose an expression into its constituents, then we will get again to the bottom also in a finite number of steps.

As an example consider a language L of propositional logic with propositional constants p_o, p_1, .. and connectives $\neg$, &, v, $\rightarrow$, $\leftrightarrow$. The constituent relation between formulas of this language is of course

well-known by now, but for present purposes we will define it once again explicitly. We do this by first defining the relation of *immediate constituency*. The immediate constituency relation $<_{ic}$ for formulas of the given language consists of all pairs of the following forms:

<A, ¬A>, <A, (A & B)>, <B, (A & B)>, <A, (A v B)>, <B, (A v B)>,
<A, (A → B)>, <B, (A → B)>, <A, (A ↔ B)>, <A, (A ↔ B)>,

The general relation of constituency $<_{co}$, which holds also between A and B when A is not an immediate constituent of B, but, for instance, an immediate constituent of an immediate constituent of B, is defined as the *transitive closure* of $<_{ic}$. That is: $<_{co}$ holds between two formulas A and B iff there is a finite chain of formulas C_o = A, C_2, ... C_n = B, with n ≥ 1, so that for all i, $C_i <_{ic} C_{i+1}$.

Let U be the set of all formulas of L. Since <U,$<_{co}$>is well-founded, we can use GOP to prove that all formulas in U have a certain property P. Here is an example of such a property. Let NPC(A) be the number of occurrences of propositional constants in A and NBC(A) the number of occurrences of binary connectives in A. Then P is the property defined in (1)

$$NPC(A) = NBC(A) + 1 \qquad (1)$$

To prove that all formulas of L have P, suppose that X is the set of all formulas in U that have P. We show that

$$(\forall A \varepsilon U)((\forall B \varepsilon U)(B <_{co} A \rightarrow B \varepsilon X) \rightarrow A \varepsilon X) \qquad (2)$$

Suppose that A is any formula and that $(\forall B \varepsilon U)(B <_{co} A \rightarrow B \varepsilon X)$. We must show that A ε X. First suppose that A is a propositional constant. Then NPC(A) = 1 and NBC(A) = 0, so (12) is satisfied and A ε X.

Second suppose that A is of the form ¬C. Then C $<_{co}$ A. So C ε X and thus (12) holds for C. But then clearly (12) also holds for A, since adding a negation sign changes neither NPC nor NBC. So A ε X.

Finally suppose that A is built from two immediate constituents C and D, combined via a binary connective. For instance let A = (C & D). Then C $<_{co}$A and D $<_{co}$ A, so by assumption C, D ε X and therefore (12) holds for C and for D. Furthermore

$$NPC(A) = NPC(C) + NPC(D) \quad (3)$$

and

$$NBC(A) = NBC(C) + NBC(D) + 1 \quad (4)$$

So $NPC(A) = NPC(C) + NPC(D) =_{(Ind.Hyp.)} (NBC(C) + 1) + (NBC(D) + 1) = (NBC(C) + NBC(D) + 1) + 1 = NBS(A) + 1.$

So once more $A \varepsilon X$.

This concludes the proof of (2). With GIP we conclude that $X = U$, i.e. that all formulas have the property P and thus satisfy (1).

q.e.d.

Another way to justify the method of proof by induction on well-founded partial orderings is to reduce it to induction on the natural numbers via the notion of *rank*.

Def. 16 (of *rank*)

Suppose that $<U, <>$ is a well-founded strict partial ordering. Then we can assign elements x of U a rank by the following condition::

(i) If for no $y \in U$, $y < x$, then $rank(x) = 0$.

(ii) Otherwise $rank(x) = max(\{rank(y): y < x\}) + 1$

In general it is not clear that this will assign a rank to every element of U. For it is in principle possible that certain elements have 'infinite rank;. (For details see Ch. 3.). But in the case considered above, and similarly for all other cases where we have proved results by induction on partial orders so far in the Notes, every element of the ordering has 'finite rank', and in that case the interpretation of Def. 16 is unproblematic, and each element of U is assigned a finite number.

Given that all members of U have finite rank we can prove that all members of U have a certain property P by using induction over $\mathbb{N}$ to prove the following related property P' of natural numbers n, defined by

$$P'(n) \text{ iff } (\forall x \varepsilon U)(rank(x) \leq n \rightarrow P(x))$$

It is straightforwardly verified that the following two statements are equivalent:

(i) The instantiaton of GIP to the set X of all members of U that have P;

(ii) The instantiaton of PA7'' to the set X of all n that have P'.

Exercise: Check this for the example discussed above in which P is the property given by (1).

Extensions of PA and Non-Standard models of Arithemtic

It follows from Gödel's Incompleteness Theorems that PA is essentially undecidable: every consistent axiomatisable extension of PA is undecidable. Exactly what is meant by 'axiomatisable' here is something that we cannot properly account for with the means available to us. (Any account will presuppose a certain amount of Recursion Theory and as things stand, Recursion Theory is entirely missing from these Notes.) But for what we want to say here it suffices to note that finitely axiomatisable extensions of PA - extensions obtained by adding a finite number of axioms to those of PA - are axiomatisable extensions in the relevant sense. So it is true in particular that all finitely axiomatisable extensions of PA are undecidable.

This entails that every consistent finitely axiomatisable extension of PA must be incomplete. For suppose $T = CL_{L_{PA}}(PA \cup \{A_1,..., A_n\})$ were consistent and complete. Then we would have the following decision procedure for T: for any sentence B of L_{PA} start simultaneously a search for a derivation of B from T and a search for a derivation of $\neg$ B from T. A search for such a derivation can be set up in such a way that if there exists a derivation, then it will eventually be found: just enumerate all finite lists of sentences of L_{PA} and check whether they are correct derivations from T and whether they yield the target sentence as a theorem. When no finite list is left out by the search, the proof must be turned up at some point. Since by assumption T is complete, there must either exist a derivation from it of B or a derivation from it of $\neg$ B. So if both searches are carried out in tandem, then a proof of one of the two formulas will eventually turn up and that then tells us

whether B is a theorem of T: It is if the derivation that has been found is of B itself; it is not if the derivation is of its negation.

The fact that no finitely axiomatisable extension of PA is consistent and complete, means that the Tarski lattice $\mathbb{T}_{L_{PA},PA}$ is very rich. On the other hand, part of it admits of a comparatively simple characterisation. Let A_1, A_2,... be a complete enumeration of all sentences of L_{PA}. Pick the first sentence A from this list that is neither provable nor refutable in PA and form the to extensions $PA_{<0>} = CL_{L_{PA}}(PA \cup \{A\})$ and $PA_{<1>} = CL_{L_{PA}}(PA \cup \{\neg A\})$. (From now on we refer to a sentence that is neither provable nor refutable form a given theory as *independent from* T.) Both extensions will be consistent and incomplete. Consider $PA_{<0>}$. Since it is incomplete, there will be sentences that are neither provable nor refutable from it. Let $A_{<0>}$ be the first of these in our list. We form the extensions $PA_{<0,0>} = CL_{L_{PA}}(PA_{<0>} \cup \{A_{<0>}\})$ and $PA_{<0,1>} = CL_{L_{PA}}(PA_{<0>} \cup \{\neg A_{<0>}\})$ of $PA_{<0>}$. Similarly we can form consistent, but necessarily incomplete extensions $PA_{<1,0>} = CL_{L_{PA}}(PA_{<1>} \cup \{A_{<1>}\})$ and $PA_{<1,1>} = CL_{L_{PA}}(PA_{<1>} \cup \{\neg A_{<1>}\})$ of $PA_{<1>}$. Each of these four theories can then be extended in its turn into a pair of consistent, incomplete and mutually incompatible theories, and so on. In this way we obtain an infinitely branching binary tree all of whose branches are infinite.

Each branch B determines a theory T_B consisting of all sentences that belong to some node of the tree. (Exercise: prove that T_B is a theory.) Let us denote the successive nodes of B as $T_{B,1}$, $T_{B,2}$, .. . It is obvious that T_B is consistent. For its successive nodes are increasing in strength - for all n $T_{B,n} \subseteq T_{B,n+1}$. So if a contradiction were derivable from T_B, it would be derivable from some $T_{B,n}$, which is impossible since $T_{B,n}$ is consistent. Second, T_B is complete. For let C be any sentence of L_{PA}. Then C occurs somewhere in our list, say $C = A_k$. Then during the construction of the first k nodes $T_{B,1}$,.., $T_{B,k}$ of B C must have been considered at least once as a possible candidate for extending the theory $T_{B,i}$ that was up for extension. At that point there were two possibilities (a) C was independent from $T_{B,i}$. Then either $T_{B,i+1} = CL_{L_{PA}}(T_{B,i} \cup \{C\})$ or $T_{B,i+1} = CL_{L_{PA}}(T_{B,i} \cup \{\neg C\})$, so either C or $\neg C$ belongs to T_B. (b) C was not independent from $T_{B,i}$. That means that either C or $\neg C$ belongs to $T_{B,i}$; so again one of C and $\neg C$ belongs to T_B.

Furthermore, it is easy to show (i) that no T_B is finitely axiomatisable over PA - this follows from the fact that the theories are strictly

increasing in strength - and (ii) that if B and B' are different branches, then $T_B \neq T_{B'}$ - there must be some node T in the tree where B and B' part company and the two daughters R' and T'' of T that belong to B and B', respectively, will then differ in that for some sentence C, T' contains C and T'' $\neg$C. We conclude that there is a 1-1 correspondence between the complete consistent extensions of PA and the branches of our tree. From this we can infer that the cardinality of the set of all complete extensions of PA is that of the power set $P(\mathrm{N})$ of the set of natural numbers N.

So our tree gives a complete description of the complete extensions of PA. But it is not by any means an exhaustive representation of $\mathbb{T}_{L_{PA}, PA}$. For one thing the extensions it represents, by its nodes and by its branches, are either finitely axiomatisable over PA (the nodes) or else complete (the branches). However, there are also many incomplete extensions of PA that are not finitely axiomatisable. Also, which finitely axiomatisable extensions turn up as nodes of the tree depends on the enumeration $A_1, A_2, ..$ of the sentences of L_{PA}. And each enumeration will leave some of them out.

Exercise: Let the enumeration $A_1, A_2, ..$ and the tree *T* of extensions of PA constructed on the basis of that enumeration be as described above.

(a) Show that the cardinality of the set of branches of *T* is that of the power set $P(\mathrm{N})$. (Hint: Show that there is a 1-1 correspondence between the set of branches and the set of all denumerably infinite sequences of 0's and 1's. Note that there is a 1-1 correspondence between the countable sequences of 0' and 1' on the one hand and the subsets of N on the other.)

(b) Show that for every complete and consistent extension T of PA there is a branch B of *T* such that $T = T_B$.

(c) Show that there are incomplete extensions of PA that are not finitely axiomatisable over PA.

(d) Show (i) that there are finitely axiomatisable extensions of PA that do not occur as nodes of *T*.

The second topic of this section concerns the models of PA. Models of PA that are not isomorphic to the standard model $\mathbb{N}$ are usually referred to as *non-standard* models. Even when we stay within the realm of the

denumerably infinite, the variety of models is very great. First, since PA is incomplete, many models differ from $\mathbb{N}$ in that they do not even verify the same sentences. Such models will not be considered here. Instead we concentrate on non-standard models of the theory Th($\mathbb{N}$), consisting of all sentences of L_{PA} that are true in $\mathbb{N}$. Even of such models there exists a great variety. (The cardinality of the set of isomorphism types of denumerable models of Th($\mathbb{N}$) is again that of $P(N)$.) Here we will only show how certain non-standard models of Th($\mathbb{N}$) can be constructed with the comparatively simple methods that are available to us.

The geneRAL method we will use consists in adding new individual constants to the first order language of the theory of departure and adding new sentences involving those constants to the theory. In the case at hand the language is L_{PA} and the theory is Th($\mathbb{N}$).

First a matter of terminology. The *numerals of* L_{PA} are the terms 0, S0, SS0, ... - in other words, all terms of the form S...S0 consisting of the constant '0' preceded by zero or more occurrences of the symbol 'S'. Note that in the standard model $\mathbb{N}$ every individual is the denotation of some numeral: If $n \varepsilon N$, then $n = [[\nu_n]]^{\mathbb{N}}$, where ν_n is the term consisting of one occurrence of 0 preceded by n occurrences of S.

We begin by adding just a single constant c to L_{PA}, thus obtaining the language $L_{PA} \cup \{c\}$, which we will denote for simplicity as L(c). Let S be the set Th($\mathbb{N}$) together with all sentences of the form $c \neq \nu$, where ν is a numeral of L_{PA}: $S = Th(\mathbb{N}) \cup \{c \neq \nu: \nu \text{ a numeral of } L_{PA}\}$. It is easy to show that S is consistent. For this it suffices to show that Th($\mathbb{N}$) together with any finite subset of $\{c \neq \nu: \nu \text{ a numeral of } L_{PA}\}$ is consistent. Sol et S' be such a finite subset. Let k be the largest number n such that the numeral ν_n occurs in the sentences of S'. Expand $\mathbb{N}$ to a model $\mathbb{N}'$ for L(c) by adding that interpretation of c which assigns it as its denotation the number k+1. Then the sentences of Th($\mathbb{N}$) are true in $\mathbb{N}'$ for the same reason that they are true in $\mathbb{N}$ and the sentences in S' are true in $\mathbb{N}'$ since the numerals they contain all denote numbers that are distinct from the denotation of c. So S' has a model and thus is consistent.

Since S is consistent, S has a model. And since any model of S will be infinite - this is because all models of Th($\mathbb{N}$) are necessarily infinite - it has a denumerably infinite model. Let M be such a model. Then M is

not isomorphic to $\mathbb{N}$. To see this, let us consider what an isomorphism h from $\mathbb{N}$ into M would have to be like. We begin by observing that every numeral will denote a unique element of M. We refer to the denotations of 0, S0, SS0, .., S^n0,.. in M as 0_M, 1_M, 2_M,., n_M,.. . It is clear that the number 0, which is the element of $\mathbb{N}$ that is the denotation in of the constant '0', can only be mapped by h onto the element 0_M of M. For if h is to be an isomorphism from $\mathbb{N}$ to M, then it must preserve in particular the denotation of '0'. So we have: h(0) = 0_M. By the same token, the number 1 can only be mapped onto 1_M, since 1 and 1_M are the denotations in $\mathbb{N}$ and M, respectively of the term 'S0'; the same applies for the numbers 2,3,.. and the elements 2_M, 3_M of M; and so on ad infinitum. So we have for every natural number n in N that h(n) = n_M.

This specifies h for all of N. But the range of h will not consist of all of M. For the truth in M of all the sentences in S entails that the denotation of c is different from all denotations of numerals in M and thus from all elements in the range of h. It is not hard to verify that h is indeed an isomorphism. (The axioms PA3-PA6 fix the extensions of + and . in both $\mathbb{N}$ and M in terms of the extensions of 0 and S. So if the latter are preserved by h, then so are the former.) But h cannot be onto M. Since there can be no isomorphism from *onto* M, $\mathbb{N}$ and M are not isomorphic.

It would be natural to try and push this method further to show that there are more isomorphism types of denumerable models of Th($\mathbb{N}$) than the two we have so far identified. But that is not easy. Additional or alternative techniques are needed to make further progress on this particular question, and many others like that. We do not pursue this issue any further here.

2.6.2. Arithmetic on the Reals.

We now turn to arithmetic of the real numbers. We mentioned in the introduction that this arithmetic admits a complete axiomatisation. Again the choice of a first order language in which the axiomatisation is to be formulated leaves some latitude. We follow the standard in adopting as language the language L_{Rea} = {0,1,+,.,<} (where 0 and 1 are individual constants, + and . are 2-place function constants and < 2-place predicate). Let $\mathbb{R}$ be the structure of the real numbers cast in hte form of a model for the language L_{Rea}. That is, $\mathbb{R}$ = <R,**0**,**1**,**+**, **.** ,**<** >, where R is the set of real numbers and **0**,**1**,**+**, **.** and **<** are the number

zero, the number one, the operations of addition and multiplication on the reals and the standard ordering of the reals, respectively.

The theory T_{Rea} in the language L_{Rea} that we will consider is also standard. Its axioms are REA1-REA18.

REA1. $(\forall x)(x + 0 = x)$

REA2. $(\forall x)(\forall y)\ (x + y = y + x)$

REA3. $(\forall x)(\forall y)(\forall z)\ ((x + y) + z = (x + (y + z))$

REA4. $(\forall x)(\forall y)(\exists z)(\ x = y + z)$

REA5. $(\forall x)\ \ (x \cdot 1 = x)$

REA6. $(\forall x)(\forall y)\ (x \cdot y = y \cdot x)$

REA7. $(\forall x)(\forall y)(\forall z)\ ((x \cdot y) \cdot z = (x \cdot (y \cdot z))$

REA8. $(\forall x)(\forall y)((y \neq 0 \rightarrow (\exists z)(\ x = y \cdot z))$

REA9. $(\forall x)(\forall y)(\forall z)\ ((x + y) \cdot z = (x \cdot z) + (y \cdot z))$

REA10. $(\forall x)(\forall y)(x < y \rightarrow \neg (y < x))$

REA11. $(\forall x)(\forall y)(\forall z)\ (x < y \ \&\ y < z \rightarrow x < z)$

REA12 $(\forall x)(\forall y)(x = y \ \vee\ x < y \ \vee\ y < x)$

REA13. $0 < 1$

REA14 $(\forall x)(\forall y)(\forall z)(y < z \rightarrow x + y < x + z)$

REA15 $(\forall x)(\forall y)(\forall z)(x < 0 \ \&\ y < z \rightarrow x \cdot z < x \cdot y)$

REA16. $\forall x)(0 < x \rightarrow (\exists z)(\ x = z \cdot z))$

REA17. $(\forall a_o)\ ..\ (\forall a_{2n+1})(\ a_{2n+1} \neq 0 \rightarrow$
$(\exists x)(a_{2n+1}.x^{2n+1} + a_{2n}.x^{2n} + .. + a_o = 0),$

for all n, where " x^n " is short for x.x. x
(multiplication of x with itself n times)

REA18. $(\forall x_1)\ ..\ (\forall x_n)(x_1^2 + ... + x_n^2 \neq -1)$, for all n

The models of T_{Rea} are known among algebraists as *real-closed fields*.

Note that the last two axioms are, like PA7 in our formalisation of the arithmetic of the natural numbers, schemata: They are not single sentences of the language but infinite collections thereof. Once again this is essential; there are no finite axiomatisations of $\mathbb{R}$ in L_{Rea} that are equivalent to the axiomatisation presented.

As in the case of PA, it is not too difficult to see that all axioms REA1-REA18 are true in the model $\mathbb{R}$ that T_{Rea} is meant to describe. No more than standard high school knowledge is needed to verify all but REA17. REA17 expresses the fact, the proof of which requires a certain amount of algebra, that every polynomial in and 'unknown' x in which the highest occurring power of x is odd takes on the value 0. (This has to do with the fact that such polynomials always become negative for sufficiently large negative values of x and positive for sufficiently high positive values of x, together with the Mean Value Theorem for continuous functions on the reals. We do not go into the details here.)

We noted in Chapter 1 that the set of the real numbers is non-denumerable: There are as many real numbers as there are sets of natural numbers. Since T_{Rea} is a first order theory, it will also have denumerable models. In neither cardinality - that of the reals or that of the denumerable sets - is T_{Rea} categorical. For the case of the reals themselves this can be easily shown by the same trick which we used to show the existence of non-standard models of arithmetic. The standard model $\mathbb{R}$ of T_{Rea} has a property reminiscent of the property of $\mathbb{N}$ we used to show the existence of a non-standard model of Th($\mathbb{N}$) and which is known by the name *archemedean* (after the great Greek mathematician Archimedes.) $\mathbb{R}$ is archemedean in that for every real number r there is a natural number n such that $r < +n$, where '+n' is short for '1 + ... + 1 n times'. i. e. for the term of L_{Rea} in which '1' is followed by n-1 occurrences of '+ 1', and -n is the unique number such that $(-n) + (+n) = 0$. The existence of a non-standard model of T_{Rea}, which is not isomorphic to $\mathbb{R}$, follows from the fact that the following set S of sentences of the language $T_{Rea} \cup \{c\}$ is consistent:
$S = Th(\mathbb{R}) \cup \{+n < c: n \varepsilon N\}$. Clearly no model of S is archemedean. So, since S has models of any infinite cardinality, it will have a non-archemedean model M of the same cardinality as $\mathbb{R}$. M cannot be isomorphic to $\mathbb{R}$, since the denotation in $\mathbb{R}$ of any term +n must be mapped by any isomorphism h onto the denotation of +n in M. But that will mean that no matching element to c_M can be found in $\mathbb{R}$. Dor suppose that $h(r) = c_M$. Since is archemedean, there is an n such that r satisfies the formula "x < +n" in $\mathbb{R}$. But on the other hand the sentence "+n < c" is true in M. so r stands in the relation **<** to the denotation of +n in $\mathbb{R}$ whereas c_M does not stand in the relation $<_M$ to $(+n)_M$ in M. Thus h would not preserve <.

A similar argument is also possible for the denumerable case, provided we can show that T_{Rea} has denumerable models that are archemedean.

This can be done. But it requires some techniques we haven't developed. So we will let this matter rest.

The remainder of this section will be devoted to a proof of the completeness of T_{Rea}. This proof rests in part on deep properties of real-closed fields and otherwise on general results and arguments in general Model Theory. We closely follow the proof presented in (Hodges, 1993), which has the merit of separating the algebraic and model-theoretic components of the argument very clearly.

As in (Hodges,1993) we take the following two facts about real-closed fields for granted. Fact 2 is a 'deep' fact about real closed fields, and an algebraist would properly argue that that is really the crux of the entire argument. We also follow Hodges in using sometimes capital letters A, B, C, .. to denote models. Given the need that will arise more than once to talk about three models at once, this is somewhat more perspicuous than using M, M', M'', .. , as we have done so far.

Fact 1.

Let M be a model of T_{Rea} and let $p(x, y_1,..,y_k)$ be polynomial in x and the parameters $y_1,..,y_k$ - i. e. a term of L_{Rea} which has occurrences of x and $y_1,..,y_k$ (where $k \geq 0$, so the case without parameters in included) and which is of the form "$a_n.x^n + a_{n-1}.x^{n-1} + .. + a_o$", where the a_i are terms of L_{Rea} not containing x - and two elements u_1 and u_2 of M such that $u_1 <_M u_2$ and $M \models p(u_1).p(u_2) < 0$. Then there is an element u in M sich that $u_1 <_M u <_M u_2$ and
$M \models p(u) = 0$.

Fact 2.

Let A be real-closed field, i.e. A is a model for L_{Rea} such that $A \models T_{Rea}$ and C an ordered subfield of A, i.e. a submodel of A which satisfies the axioms of an ordered field, that is REA1-REA15. Then there exists an extension of C to a real closed field A' within A that is 'minimal' in the following sense:

(0) $C \subseteq A' \subseteq A$, $A' \models T_{Rea}$ and if B is a model of T_{Rea} such that $C \subseteq B$, then there is an isomorphic embedding f of A' into B which is the identity on C.

The strategy of the proof is as follows. We prove that T_{Rea} has Quantifier Elimination (QE) and from this that the theory is complete.

Def. 17 A theory T of a language L *has Quantifier Elimination* iff for every formula $A(y_1,..,y_k)$ of L there is a quantifier-free formula $B(y_1,..,y_k)$ such that $T \models (\forall y_1)..(\forall y_k)\,(A \leftrightarrow B)$.

To prove that T has QE it suffices to show

(1) For every quantifier-free formula $B(x,y_1,..,y_k)$of L there is a quantifier-free formula $C(y_1,..,y_k)$ such that

$$T \models (\forall y_1)..(\forall y_k)(\exists x)(B(x,y_1,..,y_k) \leftrightarrow C(y_1,..,y_k)).$$

That (1) entails that T has QE is easily verified. Let $A(y_1,..,y_k)$ be as in Def. 17, and let $(Q_1x_1)..(Q_mx_m)\,D(x_1,..,x_m, y_1,..,y_k)$ be a formula in prenex normal form that is logically equivalent to A, where D is quantifier free and for each i = 1,.., k, Q_i is either $\exists$ or $\forall$. We can of course always arrange for this to be so, by renaming.) Suppose first that Q_k is $\exists$. Then because of (1) there is a quantifier-free formula $D_m(x_1,..,x_{m-1}, y_1,..,y_k)$ so that

(2) $$T \models (\forall y_1)..(\forall y_k)((\exists x_m)D(x_1,.,x_m,y_1,..,y_k) \leftrightarrow D_m(x_1,.,x_{m-1},y_1,.,y_k)).$$

The equivalence (2) entails that in (3), where we have replaced $(\exists x_m)\,D(x_1,..,x_m, y_1,..,y_k)$ by $D_m(x_1,.,x_{m-1},y_1,.,y_k)$ in the normal form for A:

(3) $$T \models (Q_1x_1)..(Q_mx_m)\,D(x_1,..,x_m, y_1,..,y_k) \leftrightarrow (Q_1x_1)..(Q_mx_{m-1})\,D_m(x_1,.,x_{m-1},y_1,.,y_k)$$

In case Q_m is $\forall$, we proceed analogously, but making use of the equivalence between $\forall$ and $\neg\exists\neg$: According to (1) there is a quantifier-free $D'_m(x_1,.,x_{m-1},y_1,.,y_k)$ such that

(4) $$T \models (\forall y_1)..(\forall y_k)((\exists x_m)\neg D(x_1,.,x_m,y_1,..,y_k) \leftrightarrow D'_m(x_1,.,x_{m-1},y_1,.,y_k)).$$

So defining $D_m(x_1,.,x_{m-1},y_1,.,y_k)$ as $\neg D'_m(x_1,.,x_{m-1},y_1,.,y_k))$, we get

(5) $$T \models (\forall y_1)..(\forall y_k)((\forall x_m)D(x_1,.,x_m,y_1,..,y_k) \leftrightarrow \neg D_m(x_1,.,x_{m-1},y_1,.,y_k)).$$

Again (5) enables us to eliminate the innermost quantifier $(Q_m x_m)$ from the normal form. In this way we continue until all quantifiers have been eliminated and we have found a quantifier-free formula $D_1(y_1,..,y_k)$ that is provably equivalent in T to the normal form of A and thus also to A itself. So T has QE.

Before we go on, here is a brief comment on the two Facts we have stated. It is important to realise that these are facts about the theory T_{Rea}: What is claimed here is that the facts hold in any model of T_{Rea}, not just in its standard model, or perhaps one or two other models familiar from Field Theory as a branch of Analysis or Algebra. Thus there is an important difference in particular between Fact 1 and the appeal to the Mean Value Theorem that we made when discussing the truth of the axioms of T_{Rea} in $\mathbb{R}$. The proof we appealed to there can make use of any acknowledged form of argumentation that mathematicians as legitimate for proving results about the reals. Ín contrast, the claim made by Fact 1 is that there exists a formal derivation of the fact claimed from T_{Rea} - i.e. an axiomatic derivation in the sense of Ch. 1 or the construction of closed semantic tableau in the sense of the Appendix to Ch.1. So establishing these facts requires careful checking that all steps can be justified by the axioms REA1-REA18.

To prove that T_{Rea} has QE we need two intermediate steps. First we derive the following two properties of T_{Rea}:

(6) Let A, B be models of T_{Rea} and that $A \subseteq B$. Suppose that $D(x, y_1,..,y_k)$ is a quantifier-free formula of and that $a_1,..,a_k$ are elements of A. Then if $B \models (\exists x)D(x, y_1,..,y_k)[a_1,..,a_k]$, also $A \models (\exists x)D(x, y_1,..,y_k)[a_1,..,a_k]$.

(7) Suppose that $A \models T_{Rea}$ and C a submodel of A. Then the condition of Fact 2 is fulfilled: There exists an A' such that

(0) $C \subseteq A' \subseteq A$, $A' \models T_{Rea}$ and if B is a model of T_{Rea} such that $C \subseteq B$, then there is an isomorphic embedding of A' into B which is the identity on C.

Proof of (6) Suppose that A,B, D are described in (6) and that $B \models (\exists x)D(x, y_1,..,y_k)[a_1,..,a_k]$. We make use of the fact that because we are dealing with the language and theory of rela-closed fields,

$(\exists x)D(x, y_1,..,y_k)$ can be written in a special form. First, we note that, quite generally, $D(x, y_1,..,y_k)$ can be written in disjunctive normal form and the existential quantifier then distributed over the disjuncts. So it suffices to show that if B verifies on of the disjuncts, then A does too. Each disjunct will be of the form

(8) $(\exists x)(\alpha_1 \& .. \& \alpha_r)$

where the α_j are literals of L_{Rea} - atomic formulas or negations of atomic formulas. Note that atmoic formulas of are either equations $\sigma = \tau$ or inequalities $\sigma < \tau$, where σ and τ are terms of L_{Rea}.

Our next observation is that T_{Rea} allows us to replace the negations of atomic formulas by disjunctions of atomic formulas: $\neg(\sigma = \tau)$ is provably equivalent to $\sigma < \tau \vee \tau < \sigma$ and $\neg(\sigma < \tau)$ to $\sigma = \tau \vee \tau < \sigma$. When we substitute these disjunctions for the negative literals in (8), we get a conjunction of disjunctions following the quantifier $(\exists x)$. This conjunction can be transformed once more into a disjunctive normal form and the quantifier distributed once more over the disjuncts so that we end up with a disjunction of formulas of the form (8) where now the α_j are all positive literals.

It now helps to think of the terms σ and τ that occur in these atomic formulas as polynomials in x and to think of the elements $a_1,..,a_k$ from A as 'parameters' of these polynomials. (If we want to be very formal, we can extend the language L_{Rea} to a language $L' = L_{Rea} \cup \{\underline{a}_1,..,\underline{a}_k\}$, where $\underline{a}_1,..,\underline{a}_k$ are new individual constants and expand A and B to models of L' by adding the specification that $\underline{a}_i$ denotes a_i.) This means that the conjuncts of (8) are either of the form $p(x) = q(x)$ or of the form $p(x) < q(x)$, where p and q are polynomials in x with coefficients built from the constants 0,1, $\underline{a}_1,..,\underline{a}_k$. As a next step we can, by familiar algebraic manipulations of which it can easily be seen that they can be justified in T_{Rea}, Transform atomic formulas of the form $p(x) = q(x)$ into atomic formulas of the form $r(x) = 0$ (essentially by 'subtracting q from p or vice versa, though the matter is a little more involved, since we haven't introduced - as a separate operation into our language, and similarly reduce formulas of the form $p(x) < q(x)$ to formulas of the form $r(x) > 0$. This turns (8) into a formula of the form:

(9) $(\exists x)(p_1(x) = 0 \& .. \& p_m(x) = 0 \& p_{m+1}(x) > 0 \& .. \& p_r(x) > 0)$

We now distinguish two cases. (i) First suppose that there is at least one non-trivial equation among the conjuncts. Here 'non-trivial' means that not every element of A is a solution ot the equation, i.e. every possible assignment to x verifies the equation in A. It is a well-known fact of real-closed fields that if a polynomial equation is non-trivial inm this sense, then it has only finitely many solutions; moreover, any solution that exists in a real-closed field that extends A already exists in A. (This is really what 'real-closed' means, and it follows directly from the definitions of the notion that are found in mathematics. if real-closed fields are defined as the models of T_{Rea}, then more work is necessary here. of course that work needs to be done one way or another, for as we remarked above, models of is what we are concerned with, whatever we choose to call them. It too belongs to the results that we are taking for granted here, but that an exhaustive proof would have to supply. (As should be intuitively clear, the crucial part in demonstrating this fact is played by the solution axioms REA17.)

Suppose then that the equation $p_i(x) = 0$ ($1 \leq i \leq m$) is non-trivial. Since by assumption $B \models$ (9) there is an element b in B such that

(10) $B \models (p_1(x) = 0 \ \& \ .. \ \& p_m(x) = 0 \ \& \ p_{m+1}(x) > 0 \ \& \ .. \ \& p_r(x) > 0)[b]$

Since b is a solution of $p_i(x) = 0$ in B b must by the remark above belong to A. Furthermore, all the other equations and inequalities of (9) are also satisfied by b in B and thus, since they are all quantifier-free, will be equally satisfied by b in A. So we have

(11) $A \models (p_1(x) = 0 \ \& \ .. \ \& p_m(x) = 0 \ \& \ p_{m+1}(x) > 0 \ \& \ .. \ \& p_r(x) > 0)[b]$

From this we can conclude

(12) $A \models (\exists x)(p_1(x) = 0 \ \& \ .. \ \& p_m(x) = 0 \ \& \ p_{m+1}(x) > 0 \ \& \ .. \ \& p_r(x) > 0)$,

which concluds the first case.

The second case is that where there are no non-trivial equations in (9). In this case, (9) reduces to

(13) $\exists x)(p_{m+1}(x) > 0 \ \& \ .. \ \& p_r(x) > 0)$

We now make use of Fact 1. Let b_1, ... b_s be all the solutions of the equations $p_{m+1}(x) = 0$, .., $p_r(x) = 0$ in B, given in order of magnitude in B. (I.e. er have $b_1 <_B b_2$,. .) For the same reason that was mentioned in

the argument for case (i) all of b_1, ... b_s belong to A. Moreover, since these are all the solutions to these equations, there will bee no other switches from positive(negative to negative/positive values of any of the polynomials $p_{m+1}(x)$,.., $p_r(x)$. Since by assumption B verifies (13), there is a b in B such that

(14) $B \models (p_{m+1}(x) > 0 \ \& \ .. \ \& p_r(x) > 0)[b]$

this element b will be situated somewhere with regard to the sequence of elements b_1, ... b_s, e.g. between b_j and b_{j+1}; that is, $b_j <_B b <_B b_{j+1}$. This means that in B b verifies all the inequalities occurring as conjuncts in (14). Since b_1, ... b_s are all the places where any of the polynomials $p_{m+1}(x)$,.., $p_r(x)$ changes sign, the formula (14) will be satisfied by any element in the interval (b_j,b_{j+1}), whether in A or in B. There must be elements in (b_j,b_{j+1}) in A, since the order relation in any real-closed filed is dense. (This is yet another thing that must be derived from T_{Rea}, but this is quite straightforward.) Any such element a will satisfy the formula in (9) in A. So we get:

(15) $A \models (\exists x)(p_{m+1}(x) > 0 \ \& \ .. \ \& p_r(x) > 0)$

and so, in the light of the assumptions of case (ii), we have once more (12) and we are done.

We now proceed to the proof of (7)

Let A and C be as stated in (7). We must show that there exists A' such that

(0) $C \subseteq A' \subseteq A$, $A' \models T_{Rea}$ and if B is a model of T_{Rea} such that $C \subseteq B$, then there is an isomorphic embedding of A' into B which is the identity on C.

This is almost what Fact 2 tells us. The only difference is that our assumption is that in the assumption of Fact 2 C is an ordered subfield of A, whereas what we are given in (7) is only that C is a submodel of A. To bridge this gap we argue aas follows. Assume that A,B are models of T_{Rea} and that C is a submodel of both A and B. Here we must appeal to another general fact of real-closed fields: There is unique way of extending C to an ordered subfield C' of A. (C' can be obtained as the quotient field of C, a familiar construction which among other things leads to the arithmetical structure of the rationals starting from the integers.) This minimal field extension of C can be embedded also into

B, and in fact we may as well assume that C' is within the intersection of A and B. replacing elements in B by their originals from C' under the given embedding. This gives as the situation described in the assumptions of Fact 2. So there is a real-closed field A' such that $C \subseteq C' \subseteq A' \subseteq A$, such that A' can be embedded into B by a map which preserves C' and therefore also C.

Our next step is to prove from (6) and (7) the following condition (16):

(16) If A and B are models of T_{Rea}, and $<a_1, \ldots a_k>$, $<b_1, \ldots b_k>$ are k-tuples from A and B respectively such that

(i) $(A,a_1, \ldots a_k) \equiv_o (B,b_1, \ldots b_k)$,

then

(ii) $(A,a_1, \ldots a_k) \Rightarrow_1 (B,b_1, \ldots b_k)$

First we must explain the notation. For any models A, B for some language L and tuples $<a_1, \ldots a_k>$, $<b_1, \ldots b_k>$ from these models $(A,a_1, \ldots a_k) \equiv_o (B, b_1,\ldots b_k)$ means that the tuples satisfy the same quantifier-free formulas in A and B, respectively; and $(A,a_1, \ldots a_k) \Rightarrow_1 (B, b_1,\ldots b_k)$ means that every purely existential formula $(\exists x_1)..(\exists x_m)D(x_1,.,x_m,y_1,..,y_k)$ that is verified by $a_1, \ldots a_k$ in A is verified by $b_1, \ldots b_k$ in B.

Proof of (16) from (6) and (7).

Assume that $A, B \models T_{Rea}$, $(A,<a_1, .., a_k>) \equiv_o (B,<b_1, .., b_k>)$, and that $D(x_1,.,x_m,y_1,.,y_k))$ is a quantifier-free formula of L_{Rea} such that

(17) $A \models (\exists x_1)..(\exists x_m)D(x_1,.,x_m,y_1,.,y_k))[a_1, .., a_k]$.

We have to show that

(18) $B \models (\exists x_1)..(\exists x_m)D(x_1,.,x_m,y_1,.,y_k))[b_1, .., b_k]$.

Because of (17) there are elements $c_1, \ldots c_m$ of A such that

(19) $A \models D(x_1,.,x_m,y_1,.,y_k))[c_1, \ldots c_m,a_1, .., a_k]$.

We first show that there exists an elementary extension B_1 of B and an element d_1 in B_1 such that

(20) $(A,c_1,a_1, .., a_k) \equiv_o (B_1,d_1,b_1, .., b_k)$

Let $\Psi(x_1,y_1,.,y_k)$ be the set of all quantifier-free formulas satisfied by $<c_1,a_1, .., a_k>$ in A:

(21) $\Psi(x_1,y_1,.,y_k) = \{\psi(x_1,y_1,.,y_k): A \models \psi(x_1,y_1,.,y_k)[c_1,a_1, .., a_k]\}$

We infer that

(22) For each $\psi \varepsilon \Psi$, $A \models (\exists x_1)(x_1,y_1,.,y_k)[a_1, .., a_k]$.

Consider the subset $\{a_1, .., a_k\}$ of U_A. Since L_{Rea} contains function constants the restriction of A to $\{a_1, .., a_k\}$ will not be a submodel of A. But we can close $\{a_1, .., a_k\}$ under the operations of A and obtain in this way a (uniquely determined) extension A_o of this restriction which is a submodel of A. Since by assumption $(A,a_1, .., a_k) \equiv_o (B_1,b_1, .., b_k)$, the b's satisfy the same relations in B as the a's in A. This remains the case when we close $\{b_1, .., b_k\}$ to a submodel B_o of B. That is, we can extend the map $(a_1, .., a_k) \mapsto (b_1, .., b_k)$to an isomorphism f from A_o to B_o. We can rearrange things so that f becomes the identity function by taking an isomorphic copy B' of B which contains A_o as a submodel in lieu of $f(A_o)$. In other words we may assume that A_o is both a submodel of A and of B'.

We are now in a position to apply (7): There is a model A' of T_{Rea} such that $A_o \subseteq A' \subseteq A$ and such that A' has an embedding h in B' which is the identity on A_o. Since $A' \subseteq A$ and A' and A are both models of T_{Rea}, we can apply (6) and infer from (22) that,

(23) for each $\psi \varepsilon \Psi$, $A' \models (\exists x_1)(x_1,y_1,.,y_k)[a_1, .., a_k]$.

Since h is an embedding of A' in B' which preserves $a_1, .., a_k$ we can conclude that for each $\psi \varepsilon \Psi$, $B' \models (\exists x_1)(x_1,y_1,.,y_k)[a_1, .., a_k]$, and since B is an isomorphic copy of B' under an isomorphism which maps $a_1, .., a_k$ onto $b_1, .., b_k$, we get (24).

(24) for each $\psi \varepsilon \Psi$, $B \models (\exists x_1)(x_1,y_1,.,y_k)[b_1, .., b_k]$.

This entails that there is an elementary extension B_1 of B and an element d_1 in B_1 such that

(25) $B_1 \models \Psi(x_1,y_1,..,y_k)[d_1,b_1, .., b_k]$

(The argument is the same as in the proof of Thm. 8 of Ch. 1: We extend the language with names for all elements of B and form the theory Th'(B) of B in this language. Let $\Psi(\underline{d}_1,\underline{b}_1,..,\underline{b}_k)$ be the set of all sentences $\psi(\underline{d}_1,\underline{b}_1,..,\underline{b}_k)$, where $\psi \in \Psi$, $\underline{b}_1,..,\underline{b}_k$ are the names in the extended language for $b_1, .., b_k$ and $\underline{d}_1$ is an additional new constant (in yet a further extension of the language). It is then easily shown using (22) that Th'(B) $\cup$ $\Psi(\underline{d}_1,\underline{b}_1,..,\underline{b}_k)$ is consistent. Any model of this set will be an elementary extension of B in which the sentences of $\Psi(\underline{d}_1,\underline{b}_1,..,\underline{b}_k)$ are true.) Let d_1 be the denotation of $\underline{d}_1$ in B_1. Then for all $\psi \in \Psi$ $B_1 \models \psi(x_1,y_1,..,y_k)[d_1,b_1, .., b_k]$. So we have (23).)

Since contains all quantifier-free formulas ψ such that $A \models \psi(x_1,y_1,..,y_k)[c_1,a_1, .., a_k]$ we have (20).

We can now reiterate the argument above for A and B_1. in this way we obtain an elementary extension B_2 of B_1 and an element d_2 in B_2 such that $(A,c_1,c_2,a_1, .., a_k) \equiv_o (B_2,d_1,d_2,b_1, .., b_k)$; and, continuing, we eventually get an elementary extension B_m of B and $d_1, .., d_m$ in B_m such that

(26) $(A,c_1,..,c_m,a_1, .., a_k) \equiv_o (B_m,d_1,..,d_m,b_1, .., b_k)$

From (26) and (19) we infer that

(27) $B_m \models D(x_1,..,x_m,y_1,..,y_k)[d_1,..,d_m,b_1, .., b_k]$

So

(28) $B_m \models (\exists x_1)..(\exists x_m)D(x_1,..,x_m,y_1,..,y_k)[b_1, .., b_k]$

Since B_m is an elemrntary extension of B and $b_1, .., b_k$ belong to B we reach the desired conclusion:

(29) $B \models (\exists x_1)..(\exists x_m)D(x_1,..,x_m,y_1,..,y_k)[b_1, .., b_k]$.

q.e.d.

We now come to the last step in our proof that T_{Rea} has QE. We have seen that it suffices to show that T_{Rea} has the property (1).

Let $D(x,y_1,..,y_k)$ be a quantifier-free formula of L_{Rea}. We must show that there is a quantifier-free formula $E(y_1,..,y_k)$ such that

(30) $T_{Rea} \models (\forall y_1)..(\forall y_k)((\exists x)D(x,y_1,..,y_k) \leftrightarrow E(y_1,..,y_k))$

Let $\Psi(y_1,..,y_k)$ be the set of all quantifier-free formulas $\psi(y_1,..,y_k)$ of L_{Rea} such that $T_{Rea} \models (\exists x)D(x,y_1,..,y_k) \rightarrow \psi(y_1,..,y_k))$. We show that the set $T_{Rea} \cup \Psi \cup \{\neg (\exists x)D(x,y_1,..,y_k)\}$ is inconsistent. Suppose the set was consistent. Then there would be a model B of T_{Rea} and elements $b_1,..,b_k$ of B such that

(31) $B \models \Psi(y_1,..,y_k)[b_1,..,b_k]$ and $B \models \neg (\exists x)D(x,y_1,..,y_k)[b_1,..,b_k]$.

Let $Dia(B,b_1,..,b_k)$ be the set of all quantifier-free formulas $\chi(y_1,..,y_k)$ such that $B \models \chi(y_1,..,y_k)[b_1,..,b_k]$. Then $T_{Rea} \cup Dia(B,b_1,..,b_k) \cup \{ (\exists x)D(x,y_1,..,y_k)\}$ is inconsistent. For if the set were consistent, then there would be a model A of T_{Rea} with elements $a_1,..,a_k$ such that

(32) $A \models (\exists x)D(x,y_1,..,y_k)[a_1,..,a_k]$

(33) $A \models Dia(B,b_1,..,b_k)[a_1,..,a_k]$

But (33) entails that $(A,a_1,..,a_k) \equiv_o (B,b_1,..,b_k)$. So by (16) and (32), it follows that $B \models (\exists x)D(x,y_1,..,y_k)[b_1,..,b_k]$, which contradicts the assumptions about B.

The inconsistency of $T_{Rea} \cup Dia(B,b_1,..,b_k) \cup \{ (\exists x)D(x,y_1,..,y_k)\}$ entails that there is a finite conjunction $\chi(y_1,..,y_k)$ $(= \chi_1(y_1,..,y_k)$ & .. & $\chi_r(y_1,..,y_k))$ of formulas $\chi_i(y_1,..,y_k)$ from $Dia(B,b_1,..,b_k)$ such that

(34) $T_{Rea} \models (\exists x)D(x,y_1,..,y_k) \rightarrow \neg \chi(y_1,..,y_k)$

This means that $\neg \chi(y_1,..,y_k)$ belongs to the set $\Psi(y_1,..,y_k)$. But according to (33) $B \models \Psi(y_1,..,y_k)[b_1,..,b_k]$, so $B \models \neg \chi (y_1,..,y_k)[b_1,..,b_k]$. But this is impossible since on the other hand χ is a conjunction of members of $Dia(B,b_1,..,b_k)$.

This concludes the argument that $T_{Rea} \cup \Psi \cup \{\neg (\exists x)D(x,y_1,..,y_k)\}$ is inconsistent. From the inconsistency of $T_{Rea} \cup \Psi \cup \{\neg (\exists x)D(x,y_1,..,y_k)\}$

we infer that there is a finite conjunction $\psi(y_1,..,y_k)$ $(= \psi_1(y_1,..,y_k)$ & .. & $\psi_s(y_1,..,y_k))$ of formulas $\psi_i(y_1,..,y_k)$ from Ψ such that

(35) $T_{Rea} \models \psi(y_1,..,y_k) \rightarrow (\exists x)D(x,y_1,..,y_k)$

Since $\psi(y_1,..,y_k) \varepsilon \Psi$, we also have $T_{Rea} \models (\exists x)D(x,y_1,..,y_k) \rightarrow \psi(y_1,..,y_k)$.

So we get

(36) $T_{Rea} \models (\forall y_1)..(\forall y_k)(\psi(y_1,..,y_k) \leftrightarrow (\exists x)D(x,y_1,..,y_k))$

which concludes the proof of (1) and thus of the fact that T_{Rea} has QE.

q.e.d.

Our only remaining task is to derive the completeness of T_{Rea} from the fact that it has QE. This is easy. Let a be any sentence of L_{Rea}. Then there is a quantifier-free sentence B of L_{Rea} such that $T_{Rea} \models A \leftrightarrow B$. We already saw in the proof of (6) that any atomic formula of L_{Rea} can be transformed into a formulas of a very special form that is provably equivalent to it in T_{Rea}: every such formula is equivalent to a disjunction $V_j \gamma_j$ of conjunctions γ_j of atomic formulas. In the present case, moreover we are dealing with sentences. That is, our quantifier-free sentence B can be rewritten as an equivalent disjunction $V_j \gamma_j$ in which each atomic conjunct of each γ_j is a sentence that is either of the form $\sigma = \tau$ or of the form $\sigma < \tau$. The terms σ and τ occurring in these atomic sentences are all built up form the individual constants 0 and 1 with hte help of the functions constants + and . . It is not hard to verify that each such term τ can be transformed into a 'canonical' term τ' which is either 0 or 1 or a sum of the form 1 + .. + 1 involving two or more 1's. ('Transformed' in the sense that the equation "$\tau = \tau'$ " can be proved from T_{Rea}.) it is also straightforward to verify that T_{Rea} enables us to either prove or disprove any equation $\sigma' = \tau'$ and inequality $\sigma' < \tau'$, when σ' and τ' are both canonical.

This means that T_{Rea} will either prove or refute B. T_{Rea} will prove B iff there is at least one disjunct γ_j of its rewritten form $V_j \gamma_j$ such that T_{Rea} proves every conjunct of γ_j. Otherwise T_{Rea} refutes $V_j \gamma_j$, and with it B. The same is true for the sentence A we started with. So T_{Rea} either proves or refutes every sentence from L_{Rea}.

q.e.d.

In the introduction to Section 2.6 we remarked on the intuitively paradoxical result that the arithmetic of the reals admit formalisation as a complete and decidable explicitly axiomatised theory, whereas the arithmetic of the natural numbers does not. Now that we have shown the first of these two facts at the hand of the the theory T_{Rea} is, the paradox may seem even more striking. it is true that the axioms of T_{Rea} is that capture the behaviour of the operations + and . are quite different from those of PA. The latter cannot be used here, since - obviously- there is no way of reducing what happens when these operations are applied to numbers other than the natural numbers recursively to what happens when one of the arguments is 0. But on the other hand, it is not hard to see that the axioms of T_{Rea} force the behaviour of + and . on the natural numbers to be the way that PA describes them. Specifically, let M be any model of T_{Rea} and let N_M be the set of all elements of U_M that are the denotations of some closed canonical term τ' of L_{Rea}. Then the submodel $\mathbb{N}_M$ of M with universe N_M will be isomorphic to the standard model $\mathbb{N}$ of PA. This might suggest that it should be possible to translate every sentence A from the language of PA into a sentence A' of L_{Rea} which talks only about the submodels $\mathbb{N}_M$ of models M of T_{Rea}. However, that would give us a method to check for any A whether or not it is true in $\mathbb{N}$ and that is precisely what Gödel proved to be impossible.

What then is wrong with the suggestion? The answer is - and must be - that we cannot translate sentences from Peano Arithmetic intoi sentences of L_{Rea} that 'speak only about the submodels $\mathbb{N}_M$. And that in turn implies that there can be no formula N(x) of L_{Rea} that defines the set of natural numbers in T_{Rea}, in the sense that

(37) For all models M of T_{Rea}, $N_M = \{d \varepsilon U_M: M \models N(x)[d]\}$.

Exercise: Show that if there were a formula N(x) satisfying (39), then it would be possible to define a translation function tr from L_{PA} to L_{Rea} such that for every sentence A of L_{Rea} $\mathbb{N} \models A$ iff $T_{Rea} \models tr(A)$.

That the set of natural numbers cannot be defined in T_{Rea} in spite of the fact that in every model M of the theory it consists (modulo isomorphism) of precisely the denotations in M of canonical closed terms of L_{Rea}, is itself a surprising result, which has to do with deep properties of real-closed fields. (It is a result that is entailed by Gödel's Incompleteness Theorems and the completeness of but it does not in

any direct and obvious way entail one of those two results given the other.)

That the undefinability of the natural numbers within T_{Rea} is connected with special properties of real-closed fields is indicated by the fact that arithmetic on the rational numbers is crucially different in this respect. It is possible to give a (necessarily incomplete) axiomatisation $T_{\mathbb{Q}}$ of the arithmetic of the rational numbers - for instance in the language L_{Rea} - and to define a formula N(x) such that (39) holds for models of $T_{\mathbb{Q}}$:

(38) For all models M of $T_{\mathbb{Q}}$, $N_M = \{d \varepsilon U_M: M \models N(x)[d]\}$.

(This quite difficult result is due to (Robinson, 1949).)

(38) entails that $T_{\mathbb{Q}}$ must be undecidable and incomplete, just as PA and all its axiomatisable extensions.

2.2.6. Rooted Feature Structures.

Let A be a set. An n-*ary feature structure relative to* A is an algebra S = <U, $\mathbf{f}_1$,.., $\mathbf{f}_n$>, consisting of a *universe* U and n partial unary functions $\mathbf{f}_1$,.., $\mathbf{f}_n$ over U such that

(i) no feature $\mathbf{f}_i$ is defined on any element of U that belongs to A

The elements of U ∩ A are called *the atoms of* S. We refer to the members of U \ A as the *variables* of S. S is said to be *finite* whenever U is finite. Sometimes we will refer to the elements of U also as *nodes*.

We will be especially interested in *rooted* n-*ary feature structures*. Suppose that <U, $\mathbf{f}_1$,.., $\mathbf{f}_n$> is an n-ary feature structure relative to A, u ε U and u has the property:

(*) for each v ε U, v ≠ u, there is a composition $\mathbf{f}^1{}_o$ ....o$\mathbf{f}^j$ of features such that u = $\mathbf{f}^1{}_o$ ....o$\mathbf{f}^j$(u_o) and for each r = 1,...,k there is an i ≤ n such that $\mathbf{f}^r = \mathbf{f}_i$ (In other words, each element v of U can be reached from u via a "feature path").

Then u is called a*root of* <U, $\mathbf{f}_1$,.., $\mathbf{f}_n$>. By a *rooted n-ary feature structure relative to* A we understand an n+2-tuple <U, u, $\mathbf{f}_1$,.., $\mathbf{f}_n$> such that <U, $\mathbf{f}_1$,.., $\mathbf{f}_n$> is an n-ary feature structure relative to A and u is a root of <U, $\mathbf{f}_1$,.., $\mathbf{f}_n$>.

The relation "v can be reached from u' via some feature path" where u, v are elements of the universe U of a feature structure, is clearly a transitive relation. We denote this relation as $\leqslant_S$. S is called *well-founded* if $\leqslant_S$ is irreflexive (or "has no loops", as it is also put). Well-founded feature structures are also called *unfolded*. A well-founded feature structure S is called a *feature tree* if for no u, u' ε U there are distinct paths $\mathbf{f}^1{}_o$ ....o$\mathbf{f}^j$ and $\mathbf{g}^1{}_o$ ....o$\mathbf{g}^k$ such that u = $\mathbf{f}^1{}_o$ ....o$\mathbf{f}^j$(u') = $\mathbf{g}^1{}_o$ ....o$\mathbf{g}^k$(u').

For any rooted feature structure S = <U, u, $\mathbf{f}_1$,.., $\mathbf{f}_n$> and any v ε U, let S ⌈ v (*the restriction of* S *to* v) be the rooted structure <U', v, $\mathbf{f'}_1$,.., $\mathbf{f'}_n$> where U' consists of all w ε U such that there is a path from v to w and for i = 1, ..., n $\mathbf{f'}_i$ is the restriction of $\mathbf{f}_i$ to U' - i.e. for any v ε U' $\mathbf{f'}_i$(v) = $\mathbf{f}_i$(v), provided $\mathbf{f}_i$(v) is defined, and $\mathbf{f'}_i$ is undefined otherwise. (It is

easily verified that <U', $\mathbf{f}'_1$,.., $\mathbf{f}'_n$> is an n-ary feature structure relative to A and that v is a root of the structure <U', $\mathbf{f}'_1$,.., $\mathbf{f}'_n$>.)

Notation: It is common in the feature structure literature to write "u'$\mathbf{f}^1$....$\mathbf{f}^j$ in stead of $\mathbf{f}^1{}_o$ ....$_o\mathbf{f}^j$(u).

Often the root u_o of a feature structure S is the unique element of S which satisfies condition (iv). But this need not be so. It is not so, for instance, for the 1-ary structure S_o = < {1,2}, 1, $\mathbf{f}_1$>, where $\mathbf{f}_1$ is the function {<1,2>, <2,1>}. S_o can be graphically represented as follows:

$\mathbf{f}_1$

1 2

$\mathbf{f}_1$

Here not only 1 but also 2 satisfies condition (*) of definition of rooted feature structures; so < {1,2}, 2, $\mathbf{f}_1$> is a rooted feature structure as well.

Note however that if <U, u, $\mathbf{f}_1$,.., $\mathbf{f}_n$> is well-founded, then u will always be the unique element satisfying (*). (Show this!). So every well-founded rooted feature structure has a unique root.

The first language we choose to describe n-ary feature structures is L_n = {F_1, .. , F_n, At}, where the F_i (i = 1, ... ,n) are "partial one-place functors" and At (for "Atom"!) is a one-place predicate. (Partial one-place functors are really two-place predicate constants that are consistently interpreted as partial one-place functions; given this interpretative convention, it is possible to adopt a functor-like notation for them; see below.) An n-ary feature structure S = <U, $\mathbf{f}_1$,.., $\mathbf{f}_n$> relative to A can be regarded as a model <U,*F*> for L_n, where *F*(At) = U $\cap$ A and *F*(F_i) = $\mathbf{f}_i$.

Exercise: Formulate sentences of L_n which describe the following feature structures up to isomorphism. (Letters in the first half of the alphabet denote atoms - i.e. elements of A - letters in the second half denote variables.)

Often feature structures are described with the help of languages $L'_{n,B}$ that are minor variants of the languages L_n. The languages $L'_{n,B}$ differ

from their counterparts L_n in that they have, in lieu of the 1-place predicate At, a set B of individual constants. We will assume that these sets B are subsets of some given set of "canonical names" of the members of A. That is, each constant in B is taken to denote that atom in A of which it is the canonical name. It will be harmless, and simplify matters, to assume that the elements of A act as their own canonical names, so that B is simply a subset of A.

Exercise. For each of the feature structures of the last exercise, give a uniquely identifying description of it by a sentence belonging to some appropriate language $L'_{n,B}$.

As is always the case for finite structures, every finite n-ary feature structure can be uniquely described in L_n up to isomorphism. The same is true for models for L_n which consist of finite sets of disjoint finite feature structures. Unique characterization up to isomorphism is not possible, however, for "universal models" of finite feature structures, models for L_n in which all and only the finite feature structures are represented. Let us have a closer look at such models.

To define such a universal model we have to confine the universes of the feature structures it contains to some given set V. We assume that V is denumerably infinite. An easy set-theoretical argument shows that the set $\mathbb{S}(V)$ of all finite n-ary feature structures whose universes are included in V is also denumerable. To build a model in which all the finite n-ary feature structures are represented we have to proceed carefully. We cannot simply form the union of the structures in $\mathbb{S}(V)$, for then the elements in V would have to do multiple duty and that would lead to conflicts; for instance, an element u would have to act in one complex structure as a node on which the feature $\mathbf{f}_1$, say, is defined and in some other feature structure as a node on which $\mathbf{f}_1$ is not defined. Clearly we cannot have it both ways.

To avoid this difficulty we can proceed in one of two ways. The first way is to make the variable parts of the universes of all the represented finite structures disjoint. Note that $\mathbb{S}(V)$ contains many copies of what is intuitively just one feature structure. We can get rid of such spurious duplication by forming equivalence classes of ismorphic structures. Since the set of equivalence classes is again denumerable, it can be enumerated. Using this enumeration we can then replace each equivalence type in turn by an instance $S_i = \langle U_i, F_i \rangle$ of it such that the "non-atomic part" $U_i \backslash A$ of U_i consists of elements of $V \backslash A$ that do not

occur in any of the instances chosen for the equivalence types which, in the enumeration, occur before it. In this way we obtain representatives of all the equivalence types no two of which share any variables (i.e. elements that do not belong to A). We can now form the model $M_1(V)$ = <U,*F*> as the union of all the S_i: $U = \bigcup_i U_i$ and $F(F_j) = \bigcup_i F_i(F_j)$.

Exercise: Check whether the sentences you formulated in the two preceding exercises are true in $M_1(V)$. If not, then formulate other sentences which also describe the given graphs up to isomorphism and which are true in $M_1(V)$.

Just as one can construct a universal model of all finite n-ary feature structures we can also construct, by the same method, a universal model for all finite n-ary rooted feature structures.

The second way of constructing a universal model works smoothly only for rooted structures <U, u_o, $\mathbf{f}_1$,.., $\mathbf{f}_n$> with distinguished root u_o. This time we let the finite n-ary rooted feature structures themselves be the *elements* of the model. On this universe we must define interpretations of At and of the features F_i. We take as interpretation of At the set of all feature structures that consist of single atoms, i.e. all structures <{a}, *F*> such that a ε A and $F(At) = F(F_1) = F(F_2) = ... = \emptyset$. (Note that there is an obvious bijection between this interpretation of At and A.) Furthermore, we define $F(F_j)$ as follows. We put $F(F_j)$ = the set of all pairs < S, S'> such that S = <U, u_o, $\mathbf{f}_1$,.., $\mathbf{f}_n$>, S' = <U', u'_o, $\mathbf{f}'_1$,.., $\mathbf{f}'_n$>, u'_o ε U and S' = S⌈u'_o. We refer to the model thus constructed as $M_2(V)$.

It is easy to see that neither the model $M_1(V)$ nor the model $M_2(V)$ is identified up to isomorphism by the set of sentences true in it. The reason is a quite general one: If a first order theory has models of arbitrarily large finite size, it also has infinite models. By the same token, the sentences that are true in a model in which there are objects of any finite size (no matter how large), will also have models in which there are besides these finite objects also infinite objects which can be regarded as "limits" of chains of ever larger finite ones.

The proof of this fact rests on the compactness of first order logic. We consider the model $M_1(V)$. We extend L_n to a new language L_n' by adding a new individual constant c and $Th(M_1(V))$ to a new theory Th' by adding an infinite collection of sentences which together express that c is the root of an infinite feature structure. There are many different ways in which we can do this. Perhaps the simplest way is to

state that c is the root of an infinite path consisting exclusively of applications of the feature f1. That is, Th' = Th(M_1(V)) U $\{A_n\}_{n \in \omega}$, where A_n is the sentence

$(\exists x)$ (f1 o f1 o ... o f1(c) = x). Clearly Th' is consistent. For let G be a finite subset of Th'. Then G is consistent. For among the sentences A_n that it contains there is one with highest index, say A_{n_o}. This sentence will entail all other sentences A_n in the set. It is clear, however that A_{n_o} is true in the model M for L_n' which we get by adding to M_1(V) an interpretation for c which makes c denote a feature structure that has an f1-path of length at least n_o. In this model all sentences from G which belong to h(M_1(V)) will be true as well. So G is satisfiable. By compactness Th' is satisfiable. So it has a model M'. In this model the denotation of c will be the root of an infinite f1-path.

Exercises Ch. 2.

1. a. Let {} be the language {} which has no non-logical constants, and let T be any complete consistent theory of {} .
Show that T is κ-categorical for every cardinality κ (both finite and infinite)

b. Give a complete description of the complete theories of {}. Which of these are finitely axiomatisable?

c. Show that for any first order language L there are complete theories of L that have infinite models and that are κ-categorical for all infintie cardinals κ.

Moreover, how that there are such theories that finitely axiomatisable whenever L is finite.

2. Let L = {P}, where P is a 1-place predicate.

a. Define countably many complete theories of L that only have infinite models and that are categorical for all infinite cardinalities.

b. Specify a complete theory of L that is ω-categorical but not κ-categorical for uncountable cardinals κ.

Hint: One can express, by means of an infinite number of axioms of L, that (i) the extension of P is infinite, and (ii) that the complement of P's extension (the set of individuals that do not satisfy P) is also infinite. It is easy to show (i) that this theory has infinite models; (ii) that any model of it is infinite; (iii) that any two countable models of the theory are isomorphic; and (ii) that for any uncountable cardinality κ there are models of the theory of cardinality κ which are not isomorphic. (N.B. follows from the fact that M is a model of the theory and $|U_M|$ is uncountable, then the extension of P in M could be either countable or uncountable.)

c. Let L' = {R}, where R is a 2-place predicate.

Define countably many complete theories of L' that only have infinite models and that are ω-categorical but not κ-categorical for uncountable cardinals κ.

3. Show that the theory T_{rat} is not categorical for uncountable cardinalities.

Hint: In view of Morley's Theorem it suffices to show this for just one uncounbtable carfinality. Choose the cardinality 2^{ω} of the set $\mathbb{R}$ of real numbers.
Compare the following two models for the language {<}:

$M_1 = <\mathbb{R},<_{\mathbb{R}}>$, where $<_{\mathbb{R}}$ is the standard ordering of $\mathbb{R}$.
$M_2 = <\mathbb{Q} \otimes \mathbb{R},<'>$, where $\mathbb{Q}$ is the set of rational numbers and $<'$ is the "alphabetic ordering of $\mathbb{Q} \otimes \mathbb{R}$ induced by the standard orderings of $\mathbb{Q}$ and $\mathbb{R}$" - that is, for $q,q' \varepsilon \mathbb{Q}$ and $r,r' \varepsilon \mathbb{R}$ $<q,r> <' <q',r'>$ iff (i) $q <_{\mathbb{Q}} q'$ or (ii) $q = q'$ and $r <_{\mathbb{R}} r'$.
It follows from general facts of set theory that M_2 has cardinality 2^{ω} and thus that M_1 and M_1 are of the same uncountable cardinality.
Show that M_1 and M_1 are not isomorphic.

4. Let DS(T,L) be the lattice of all extensions of a given theory T of a some 1-st order language L.

Show: If DS(T,L) is a boolean algebra, then DS(T,L) is finite.

5. (Stone Representation Theorem for Boolean Lattices.)

Let $BL = <U, \leq>$ be any boolean lattice. For each $b \varepsilon U$, let $I_b = \{d \varepsilon U: d \leq b\}$. ($I_b$ is called the *prime ideal determined by* b.)

Show that BL is isomorphic to the structure $<U', \subseteq>$, where $U' = \{I_b: b \varepsilon U\}$ and $\subseteq$ is set-theoretical inclusion.

N.B. The intuitive significance of Stone's Representation Theorem is that all different types of boolean lattices (and thus also all types of all boolean algebras) are realised by set-theoretical structures, whose universe consists of subsets of some given set and in which lattice relation is set-theoretic inclusion.

(This is the purport of all representation theorems in mathematics: Every structure satisfying some general requirements (such as that of being a model of a given set of axioms) is isomorphic to - and thus can be "represented" as - a structure of some special form.)

6. Show that 0 and S are definable within PA in terms of +.

7. Show that S is not definable within PA in terms of . (multiplic.).

(N.B. intuitively this means: the successor operation on the natural numbers is not definable within PA just with the help of multiplication.)

Hint: Let $T_{PA,\{.\}}$ be the set of all sentences from the sublanguage {.} of L_{PA} that are theorems of PA.

i. Show that any denumerable model M of $T_{PA,\{.\}}$ is isomorphic to the model $\mathbb{N}_{\{.\}} = \langle N, .^N \rangle$, where $.^N$ is the multiplication operator from the standard model $\mathbb{N}$ of arithmetic.

To show this, first observe that "the number zero" and "the number one" are definable in PA from . alone (i.e. we can define in terms of . the predicate "is equal to the number zero" and the predicate "is equal to the number one"); and further that with . we can also define the predicate "is a prime". Once this has been established it is easily seen that among the things that $T_{PA,\{.\}}$ asserts is that there are infinitely many primes and that these are all different from both zero and one. This means that denumerable model M of $T_{PA,\{.\}}$ has a unique zero, a unique one and infinitely distinct primes. it is then easy to show that any bijection between the primes of M and the primes of the standard model of arithmetic N is an isomorphism between M and the model $\mathbb{N}_{\{.\}}$.

ii. Show that (i) entails the non-definability of S in $T_{PA,\{.\}}$.

8. Let L be the language (0, S}, with 0 a 0-place function constant and S a 1-place dunction constant. Let T be the L-theory that is axiomatised by the set { A_1,A_2}, where:

A_1 := $(\forall x)(x \neq 0 \leftrightarrow (\exists y)(x = Sy))$

A_2 := $(\forall x)(\forall y)(S(x) = S(y) \rightarrow x = y)$

Let *N* be the L-model <N, I>, where:

(i) N is the set of natural numbers;
(ii) I(0) = 0 (i.e. I(0) is the number zero); and
(iii) for every natural number nm I(S)(n) = n+1.

Evidently N is a model of T.

Show: There exist countably infinite models of T that are not isomorphic to N .

9. Let L be the language {0, 1, S, P}, where 0 and 1 are individual constants and S and P are 2-place predicate constants. LetT be the theory of L that is axiomatised by the following set of axioms:

A1. $(\forall x)(\forall y)\forall z)(S(x,y) \ \& \ S(x, z) \rightarrow y = z)$
A2. $(\forall x)(\forall y)\forall z)(S(y,x) \ \& \ S(z, x) \rightarrow y = z)$
A3. $(\forall x)((\exists y)(S(x,y)) \leftrightarrow x \neq 1)$
A4. $(\forall x)((\exists y)(S(y,x)) \leftrightarrow x \neq 0)$

A5. $(\forall x)(\forall y)\forall z)(P(x,y) \ \& \ P(x, z) \rightarrow y = z)$
A6. $(\forall x)(\forall y)\forall z)(P(y,x) \ \& \ P(z, x) \rightarrow y = z)$
A7. $(\forall x)((\exists y)(P(x,y)) \leftrightarrow x \neq 0)$
A8. $(\forall x)((\exists y)(P(y,x)) \leftrightarrow x \neq 1)$

A9. $(\forall x)(x \neq 1 \rightarrow (\exists y)(S(x,y) \ \& \ P(y,x))) \ \&$
$(\forall x)(x \neq 0 \rightarrow (\exists y)(P(x,y) \ \& \ S(y,x)))$

(Intuitively the conent of T as as follows:

(i) (A1 -A4) say that S sdenotes a partial 1-1 function, such that all elements of the universe U except for 1 belong to its domain and all elements of U except for 0 belong to its range;

(ii) (A5-A8) say that the same applies to P, except that in this case it is 0 that is missing from the domain and 1 that is missing from the range.

(iii) (A9) says that the function from U\{0} onto U\{1} that is denoted by P is the inverse of the function from U\{1} onto U\{0} that is denoted by S.)

1. Show that T has an infirnite model and that all models of T are infinite.

2. The constants 0, 1, P are all definable in T using just the constant S. (That is, for each of these three constants there is an explicit definition in which the only non-logical constant appearing on the right hand side is S.)

   Give explicit definitions for 0, 1 and P in terms of S in T.

10. Let L be the language {=, <, I, 0,S}, in which < is a 2-place predicate, I a 1-place predicate, 0 an individual constant and S a 2-place predicate. Let T be the theory of L that is axiomatised as follows:

A1 $\forall x \forall y\ (x < y \rightarrow \neg\ y < x)$
A2 $\forall x \forall y \forall z\ (x < y\ \&\ y < z \rightarrow x < z)$
A3 $\forall x \forall y\ (x < y\ v\ x = y\ v\ y < x)$
A4 $\forall x \forall y\ (x < y \rightarrow \exists z\ (x < z\ \&\ z < y)$
A5 $I(0)$
A6 $\forall x \forall y\ (S(x,y) \rightarrow I(x)\ \&\ I(y))$
A7 $\forall x \forall y \forall z\ (\ S(x,y)\ \&\ S(x,z) \rightarrow y = z)$
A8 $\forall x \forall y \forall z\ (\ S(x,z)\ \&\ S(y,z) \rightarrow x = y)$
A9 $\forall x \forall y\ (S(x,y) \rightarrow x < y)$

It is easily verified that T holds in the following model M_o:

(i) the universe of M_o is the set Q of rational numbers;
(ii) $<_{M_o}$ is the "less than"-relation between rational numbers;
(iii) I_{M_o} is the set of integers;
(iv) 0_{M_o} is the number zero; and
(v) S_{M_o} is the successor relation between integers.

Show that there is apart from M_o at least one ohter countable infinite model of T which is not isomorphic to M_o .

11. Let L_{PA} (= {0, s, +, .}) be the language of Peano Arithmetic. Let L_1 be the extension $L_{PA} \cup \{c_1,P\}$ of L_{PA} where c_1 is an individual constant and P a 2-place predicatem and let L_2 be the extension $L_1 \cup \{c_2\}$ of L_1 where c_2 is an individual constant . (So L_1 = {0, s, +, ., <, c_1} and L_2 = {{0, s, +, ., <, c_1, c_2}.)
Let T_1 = Cl(PA $\cup$ {$(\forall x)(\forall y)(x < y \leftrightarrow (\exists z)(\ z \neq 0\ \&\ x + z = y)\} \cup$
{$0 < c_1$, $S0 < c_1$, $SS0 < c_1$, ...}).

Let $T_2 = T_1 \cup \{c_1 < c_2, Sc_1 < c_2, SSc_1 < c_2, ...\}$. and let M_1 be any model of T_1.

Show that M_1 dcan be expanded to a model M_2 of T_2 by adding a suitable interpretation of the constant c_2.

12. Let L be the language {<}, where< is a 2-place predicate constant and öet L' = L ∪ {S}, with S a 1-place function constant. Let T' be the theory $Cn_{L'}(\{A.1, ..., A.4\})$, where:

A.1 $(\forall x)(\forall y)\ (x < y \rightarrow \neg\ (y < x))$

A.2 $(\forall x)(\forall y)(\forall z)\ ((x < y\ \&\ y < z) \rightarrow x < z)$

A.3 $(\forall x)(\forall y)\ (x < y\ \vee\ x = y\ \vee\ y < x\ )$

A.4 $(\forall x)(x < S(x)\ \&\ (\forall z)(x < z \rightarrow (S(x) < z\ \vee\ S(x) = z)))$

Show that S is definable in T' (i.e. in terms of <).

13. Deduce the fllowing statement from the axioms PA1-PA7:

$(\forall x)(\ (\exists\ y)(x = 2.y) \leftrightarrow \neg\ (\exists\ y)(x = 2.y + 1)\ )$

14. a. We extend the language of arithmetic L_{PA} with a new 1-place predicate G to the language $L' = L_{PA} \cup \{G\}$ and extend the theory PA to a theory T' of L' by adding as a new axiom the following definition D of G in terms of +:

(D) $(\forall x)(G(x) \leftrightarrow (\exists y)(\ x = y + y))$

(Intuitively D says that G denotes the property "is an even number".)

Show that the sentence $(\forall x)(G(x)\ \vee\ G(S(x)))$ is derivable from T'.

b. This time we extend L_{PA} with a new 2-place predicate < to the language $L'' = L_{PA} \cup \{<\}$ and extend PA to a theory T'' of L'' by adding as a new axiom the following definition D' of < in terms of + and 0:

(D') $(\forall x)(\forall y)(\ x < y \leftrightarrow (\exists z)(z \neq 0\ \&\ z + x = y))$

Show that the sentence (1) is deducible from T''.

(1) $(\forall x)(\forall y)(x < y \rightarrow x \neq y)$

(Hint: One way to show this is to prove first that (1) is equivalent to (2)

(2) $(\forall x)(\forall v)(x + Sv \neq 0)$

(Intuitively D says that G denotes the property "is an even number".)

Show that the sentence $(\forall x)(G(x) \vee G(S(x)))$ is derivable from T'.

b. This time we extend L_{PA} with a new 2-place predicate < to the language $L'' = L_{PA} \cup \{<\}$ and extend PA to a theory T'' of L'' by adding as a new axiom the following definition D' of < in terms of + and 0:

(D') $(\forall x)(\forall y)(x < y \leftrightarrow (\exists z)(z \neq 0 \ \& \ z + x = y))$

Show that the sentence (1) is deducible from T''.

(1) $(\forall x)(\forall y)(x < y \rightarrow x \neq y)$

(Hint: One way to show this is to prove first that (1) is equivalent to (2)

(2) $(\forall x)(\forall v)(x + Sv \neq 0)$

and then to prove (2) be mathematical induction.)

15. Let L and L' be languages of first order logic and let T and T' be theories of L and L', repsectively. Let I be a function from the sentences of L to the sentences of L'. (We may call such a function I a "translation" from L to L'.) We say that I *interprets* T i*n* T' iff for every sentence A of L such that $T \models A$, $T' \models I(A)$. Second, let $\mathbb{I}$ be a set of translation finctions from l to L'. Then we say that T is *interpretable in* T' *relative to* $\mathbb{I}$ iff there is an I in $\mathbb{I}$ which interprets T in T'. i

Let now T be the theory od stron partial orderings in the language L = {<}, whose axioms are

$$(\forall x)(\forall y)(\ x < y \ \rightarrow \neg \ y < x\ )$$
$$(\forall x)(\forall y)(\forall z)(\ x < y \ \&\ \ y < z \ \rightarrow \ x < \ z)$$

We can interpret T in the theory PA formulated in the language L_{PA} of Section 2.6.1 by means of the function I which is "based on" the following definition of "<" in L_{PA}:

$(D_<)$ $\qquad (\forall x)(\forall y)(x < y \ \leftrightarrow \ (\exists z)(z \neq 0 \ \&\ x + z = y))$

Here, when we say that I is "based on" $(D_<)$ what we mean is that for any sentence A of L, I(A) is the sentence which we get by replacing each subformula "u < w" of A by the right hand side of $(D_<)$, replacing x by u and v by w (and if necessary renaming z in order to avoid variable clashes).

Show that I interprets T in PA (and therewith that T is interpretable in PA relative to the set of all translations from L into L_{PA} that are based on possible definitions of "<" in L_{PA}).

16. Let T be a theory of some first order language L and let α be a non-logical constant of L. Let D be the set of all possible explicit definitions of α in terms of the remaining vocabulary of L. (That is, if α is an n-place predicate P, then D will be the set of all sentences of the form $(\forall v_1)..(\forall v_n)(P(v_1,.,v_n) \leftrightarrow A)$, where A is a formula of L \{α} in which only $v_1,.,, v_n$ may have free occurrences; and if α is an n-place function constant f, then D is the set of all formulas $(\forall v_1)..(\forall v_n)(f(v_1,.,v_n) = v_{n+1} \leftrightarrow A)$, where A is a formula of L \{α} in which the only free occurrences are of the variables $v_1,.,, v_{n+1}$.)

Let $\mathbb{I}$ be the set of all translations of L into L \{α} that are based on definitions in D, where for a definjition d ε D with right hand side A_d the translation I_d based on d is the one which replaces in any formula B of L all occurrences of atomic formulas involving α by the corresponding instantiations of A_d. (See also the previous exercise.)

Let T' be the theory T ∩ {C: C is a sentence of L \{α}}

Show: T is interpretable in T relative to $\mathbb{I}$ iff α is definable in T.

Lösungen von einigen Aufgaben.

9. Wir bezeichnen das zu beweisende Theorem
$(\forall x)((\exists y)(x = 2.y) \leftrightarrow \neg (\exists y)(x = 2.y + 1))$ als (*).

Wir verfahren nach Induktion und zeigen (*) indem wir zeigen:

(**) $(\exists y)(0 = 2.y) \leftrightarrow \neg (\exists y)(0 = 2.y + 1)$

(***) $((\exists y)(x = 2.y) \leftrightarrow \neg (\exists y)(x = 2.y + 1)) \rightarrow$
$((\exists y)(Sx = 2.y) \leftrightarrow \neg (\exists y)(Sx = 2.y + 1))$

(**): Einerseits haben wir PA $\vdash 0 = 2.0$. Also auch PA $\vdash (\exists y)(0 = 2.y)$. Andererseits gilt: PA $\vdash \neg (\exists y)(0 = 2.y + 1)$. Denn nehmen wir an, dass $(\exists y)(0 = 2.y + 1)$, dann gibt es ein y, so daß $0 = S(2.y)$, was dem PA-Axiom widerspricht, dass 0 nicht von der Form Sx ist.

(***): Nehmen wir an: $((\exists y)(x = 2.y) \leftrightarrow \neg (\exists y)(x = 2.y + 1))$.
Dann gilt also entweder

(i) $(\exists y)(x = 2.y) \& \neg (\exists y)(x = 2.y + 1)$ oder
(ii) $\neg (\exists y)(x = 2.y) \& (\exists y)(x = 2.y + 1)$

Im ersten Fall gibt es ein n, so daß $x = 2.n$. Also gilt $Sx = S(2.n) = 2.n + 1$ und deshalb auch $(\exists y)(Sx = 2.y + 1)$. Wäre es der Fall, daß $(\exists y)(Sx = 2.y)$, so gäbe es ein n, so daß $Sx = 2.n$. Offenbar kann n nicht gleich 0 sein. Also ist $n = Sm$ für irgendein m. Dann aber $Sx = 2. Sm = Sm.2 = Sm + Sm = S(Sm + m)$. Also ist $x = Sm + m = m + m + 1 = m.2 + 1 = 2.m + 1$. Also $(\exists y)(x = 2.y + 1)$, was dem zweiten Konjunkt in (i) widerspricht. Also führt die Annahme, daß
$(\exists y)(Sx = 2.y)$ zu einem Widerspruch. Somit haben wir
$(\exists y)(Sx = 2.y + 1) \& (\exists y)(Sx = 2.y)$ und damit
$(\exists y)(Sx = 2.y + 1) \leftrightarrow \neg (\exists y)(Sx = 2.y)$.

Der zweite Fall, (ii), erledigt sich ähnlich.

ii. Zu zeigen:

$$(\forall x)(\forall y)(Sx \cdot Sx = (x \cdot x) + y \rightarrow \neg (\exists u)(y = 2.u))$$

(Intuitiv besagt diese Formel, daß die Differenz zwischen zwei aufeinanderfolgenden Quadraten immer eine ungrade zahl ist.) Wir agumentieren wie folgt:

$Sx \,.\, Sx = (Sx \,.\, x) + Sx = (x \,.\, Sx) + Sx = ((x \,.\, .x\,) + x) + x + 1$
$= (x \,.\, .x\,) + (x + x + 1) = (x \,.\, .x\,) + (2..x + 1)$. Also, wenn
$Sx \,.\, Sx = (x \,.\, x) + y$, dann ist $y = 2..x + 1$. (Siehe unten!).
Wenn aber $y = 2..x + 1$, dann gilt auch $(\exists\, u)(y = 2.u + 1)$.
Dann gilt aber nach (i), daß $\neg\pi\ (\exists\, u)(y = 2.u)$.

(Wir haben hier von dem Prinzip Gebrauch gemacht, nach dem aus $x + y = x + z$ folgt, daß $y = z$. Dieses Prinzip läßt sich leicht nach Induktion beweisen:

(i) Wenn $o + y = o + z$, dann natürlich $y = z$.
(ii) Wenn gilt, dass (a) wenn $x + y = x + z$, dann $y = z$, dann gilt auch, dass (b) wenn $Sx + y = Sx + z$, dann $y = z$. Denn sei $Sx + y = Sx + z$. Dann $y + Sx = S(y + x) = z + Sx = S(z + x)$. Dann aber $y + x = z + x$. Also $x + y = x + z$ und nach Induktionshypothese $y = z$.)

Chapter III Set Theory as a Theory of First Order Predicate Logic.

Here is an appealing and apparently clear picture of the "universe of all sets": Suppose that a set A of "individuals" or "Urelements" is given. Then we can form sets from those individuals; these will be subsets of A. We can then form sets of which these subsets of A are in turn members;. In fact, it seems reasonable to hold that we can form not only such sets, but also sets which consist partly of subsets of A and partly of members of A; the sets which have only individuals as members and those which have only sets of individuals as members are special cases of this more general category. Having formed this second tier of sets we can then proceed to form a third tier, a collection of sets the members of which may be individuals, sets of individuals and sets which themselves count sets of individuals among their members. Carrying on in this manner ad infinitum we run through the so-called "cumulative hierarchy (of sets)". The structure which results in this way is the subject of the *theory of sets*. It is this structure that any axiomatic set theory should try to capture.

It isn't quite right to speak of *the* structure of set theory. For what the iterative process of forming sets produces evidently depends on the set A with which we start. But among the many different hierarchies which are generated by different sets of Urelements there is one that is special. This is the hierachy which results when we start with nothing, so to speak, i.e. when we begin with the empty set. It may not be immediately obvious that this will get us anything at all, but only a little reflection shows that it does. All that needs to be acknowledged is that the empty set is fit to act as a member of other sets. Once we accept this, we see that there is at least one other set besides the empty set, viz. the set whose only member is the empty set. (Clearly this set is different from the empty set, for it does have a member, whereas the empty set itself has none.) As soon as we have these two sets, it is possible to form more sets, e.g. the set which has both these two sets as members, etc. In fact, even if we start with the empty set of individuals, iterated set formation leads eventually to an unimaginably huge universe, and one that is certainly big enough to model any abstract structure - such as that of the real numbers, or of all functions from real numbers to real numbers, etc., etc. - that pure mathematics and the sciences which use mathematics as a tool ever made a topic of investigation. Because it gives us enough for these purposes, while on the other hand it apparently does without "extra-logical" assumptions (it does not involve the assumption of any "Urelements", which are themselves not sets), the hierarchy which starts from the empty set has

become the preferred object of study within mathematical logic. It is this structure that is usually referred to as *the cumulative hierarchy*.

The cumulative hierarchy, then, is that structure which we get when, starting from the empty set, we generate sets by the iterative procedure just sketched and carrying on "ad infinitum", as we just put it. But what is "ad infinitum"? It may be that what is meant by this appwars reasonably clear at first. But upon reflection the illusion of clarity quickly evaporates. The infinite, in all its different manifestations, is one of the trickiest abstract concepts there are, and this applies to the phrase "ad infinitum", as it figures in our informal description of the cumulative hierarchy, no less than to any other manifestation of it.

Set Theory was invented in large part to analyse the concept of infinity, and to develop systematic means of studying and describing its different manifestations in different contexts. Because of this it is in the curious situation that what it has to say about infinity is constitutive of the very structure of which it is meant to provide an accurate description. As a result there is, from the perspective we adopted in Ch. 2 a certain kind of circularity here, which is unlike anything we have found in connection with other theories discussed there that aim at the description of a single structure, such as the theory of the order of the rationals, or Peano Arithmetic, or the Theory of Real Closed Fields. In all those cases there was a well-defined, and independently definable, structure against which the axioms of the theory could be checked, so that various well-defined questions can be raised about the relation beteen structure and theory, e.g. whether the theory gives a complete, or a categorical characterisation of the structure. (And as we saw it is often possible, if rarely simple to answer such questions.)

Set Theory is different in this respect. The very question what the structure is like that it is its purpose to describe cannot be detached from the description that the theory itself provides; for part of what the theory asserts is what iteration of a given operation or set of operations *ad infinitum* comes to, and thus what the structure is that is the result of such an iteration ad infinitum.

One of the striking discoveries about infinity - which stood, one might say, at the cradle of Set Theory as we know it today - was that it comes in different 'degrees', or 'sizes'. As we noted in Ch. 1, Cantor. the founder of modern set Theory, showed that the power set $P(X)$ of a set X is of higher cardinality than X itself. This is true for any set X

whatever, and so in particular when X is infinite. Consequently each infinite set X is the starting point of an unbounded sequence X, $P(X)$, $P(P(X))$,.. of sets of ever larger infinite cardinality. But havin established that there is a multiplicity of different infinities, the set theorist sees himself confronted with further questions, concerning (a) the extent and (b) the structure of this multiplicity. Two such questions have dominated Set Theory for most of its history: (i) How many different sizes of infinity - how many 'cardinalites' - are there altogether? and (ii) are there any sets X whose cardinality |X| is between that of the set N of the natural numbers and that of its power set $P(N)$? (This second question is known as the issue of the *Continuum Hypothesis*. The Continuum Hypothesis (CH) is the statement that there are no such sets X: $\neg\, (\exists X)(|N| < |X| < |P(N)|)$.)

The investigations concerning the CH can be divided into three phases. At first, the goal was simply to decide whether or not the Continuum Hypothesis is true. This is the way Cantor, the one who introduced the issue of of the CH into Set Theory, understood it. (Cantor seems to have worked on this problem relentlessnly and the strain caused by his failure to settle the matter is said to have contributed to his eventual mental breakdown.) The second phase set in after, in the early parts of the 20-th Century, Set Theory had been formalised and characterised as a formal theory, given by a certain set of axioms. At that point the problem of the CH took on a correspondingly formal complexion: Can the CH be either proved or refuted from the axioms of formal Set Theory[1]? This question was settled in two stages. First Gödel proved in 1940 that CH is consistent with formal Set Theory, and thus that the axioms do not refute it. Then, in 1963, Cohen proved that CH is independent of this system, i.e. that it cannot be proved from its axioms either.

Cohen's result was not only the conclusion of the second phase, but also the point of departure for the third. This phase (which continues to the present day and will quite possibly never be concluded) is characterised by the search for new set-theoretical principles which settle the CH one way or the other, and which at the same time can be argued to be true on independent, intuitively persuasive grounds.

---

1 The formalisation of Set Theory didn't lead to just one set of axioms. However, it became clear fairly soon that the major proposals do not differ from each other as far as CH is concerned. So we can, without serious distortion to what actually happened, describe this phase in the history of Chas the question whether CH can be either proved or refuted from one of these axiomatic theories, viz from the theory ZF, or 'Zermelo-Fraenkel',which will be presented in this Chapter.

Though a number of formal results were achieved in the aftermath of Cohen's result, involving new axioms which settle the CH one way or thev other, none of the new axioms that were proposed seem to qualify as unequivocally true. So, from a conceptual point of view the CH is na open question to this day

For our present purposes the first question - What is the total range of infinite cardinalities? - is of more immediate importance. Work on this question has taken on a flavour much like that connected with CH: Various axioms have been proposed, each of which tells us something about the range of infinite cardinalities. Most of these axioms are 'Large Cardinal Axioms', which when added to ZF guarantee the existence of cardinalities larger then any that can be proved to exist without them. But the conceptual difficulty connected with these results is much like the one we just mentioned in connection with CH: In general it is difficult to persuade oneself that the proposed axioms must be true.

Connected with the question how large infinities can get is the question what should be understood by the phrase 'ad infinitum'. Even the multiplicity of cardinalites that is guaranteed by ZF by itself (i.e. without the addition of any further axioms) implies that many different answers are possible in principle here. One possible interpretation of *ad infinitum* is that "iteration ad infinitum" should be understood as iteration going up to the first, or 'lowest', degree of infinity, viz. that of denumerably infinity. The structure which is obtained by iterating, starting from the empty set, the set-forming operations up to this first level of infinity is known as the *Hierarchy of Hereditarily Finite Sets*. It goes by this name because all its elements are sets that are *hereditarily finite* in the sense that (a) they are finite themselves, and (b) their members are also finite sets, and likewise for the members of those members, and so on all the way down. It is clear, however, that this is *not* the structure that the axioms of Set Theory should try to capture. It is of the essence of the "real" structure of sets that some of the sets in it are inifinite. Since the Hierarchy of Hereditarily Finite Sets doesn't contain any such sets, not even the set of natural numbers, it cannot be the the one we are after.

Even apart from this consideration, the Hierarchy of Hereditarily Finite Sets should not qualify as the structure that Set Theory should describe on the grounds that what we intuitively want is the structure which results from iterating the set-forming operations through *all* infinite cardinalities; teh itereation shouldn't be stopped at any earier stage,

and stopping at the very first opportunity that offers itself is about as far removed from this general desideratum as possible

As we already noted, there is no way to determine the properties of the full structure of sets completely independently of what Set Theory says, for it is the theory which asserts how large and complex sets can become. In the light of all the work that has been donme on the question of large c ardinals there has been a growing impression that what can be said about this must to some extent remain a matter of stipulation. The upshot of this is that there may be no one 'true' structure of sets and therefore possibly also no one correct axiomatic set theory. The second question is complicated, however, by the c ircumstance that axiomatic set theories like the Theory of Zermelo-Fraenkel, or 'ZF', which we will present below, admit of so-called 'inner models' - structures which satisfy all the axioms of the theory but which are obtained be iterating the set formation operations only up to the cardinality of some set whose existence the axioms enable us to prove.[2] For this reason the quest for th right axiomatisation of Set Theory does not stand or fall with the quest for the true 'set-theoretical universe'.

Not only are first order axiomatic set theories like ZF exceptional from the perspective adopted in Ch. 2, they also hold a unique position within the landscape of logic, mathematics and the exact sciences in a different sense. As we noted in the Interlude on Set Theory in Ch.1, Set Theory is indispensible in the formalisation of mathematics. As we also noted there, the insight that it is needed for this purpose is certainly not self-evident; and as things actually happened, it was something that was learned the hard way: The insight emerged when Russell detected the error which had slipped into Frege's attempt to reduce arithmetic to 'pure logic' and which Russell exposed in the form of what has come to be known as 'Russell's Paradox'.

---

[2] This sounds paradoxical, for how can a structure which verifies all the axioms of the theory fail to contain sets that the theory claims to exist? The answer is that an inner model will in general not only lack the sets which would be reached only by carrying the iteration beyond the point where th inner model is reached, but also many of the functions which establish 1-1 correspondences between sets that are part of the inner model. This makes it possible for sets in the inner model to appear from a perspective internal to the inner model as if they had a larger cardinality than they can do from the external perspective of 'reality', - the functions that would establish them as being of the same cardinality as certain other sets of the internal model (and thus as having no larger cardinality than these), simply are not around.

The need for set-theoretical principles arises in the formalisation of any part of mathematics or science. It arises in particular in the formalisation of parts of *metamathematics*, i.e. of the discipline which deals with the general properties (such as completeness , consistency, soundness, compactness, etc.) of logical systems like the predicate calculus. And it is especially in such formalisations that the conceptual implications of its use are most important. For the point of such formalisations is to make certain that the general framework of mathematics and science does indeed have the general poperties of soundness and consistency which we attribute to it. [3]

When Set Theory is used as metatheory in formalisation, and especially in its role as metatheory in the formalisation of parts of metamathematics, it is of the outmost importance that its principles be ascertainable as true. For this reason formalisations in metamathematics should try to make as parsimonious a use of set-theoretic principles as possible, and to employ only those whose

---

[3] To give an idea of what formalisations of parts of metamathematics come to, here is an outline of the formalisation of the very first results we proved in Ch. 1, the soundness and completeness of first order logic. The formalisation of these results will involve, first, formal definitions within the language of ZF of the syntax, model theory and proof theory of first order predicate logic. This means that the languages of predicate logic, their symbols, formulas, and derivations as well as the models for those languages and the sequencs of formulas that constiotute correct derivations, are represented as set-theoretic objects, and that soundness and completeness are formulated as statements - pertaining to those objects - in the language of set-theory. Second, the proofs of soundness and completeness can then be turned into formal axiomatic derivations - in the sense defined in CH.1, Sn 1 - from the axioms of set theory together with the mentioned definitions.

Note that such attemtps at providing additional support for the soundness of our general logical framework are affected by an ineliminable element of ciruclarity. For the fact that the soundness theorem can be demonstrated in the form of a formal derivation provides support for its being true only to the extent that the formal method of derivationthat is used in the demonstration can be trusted. But that is precisely the issue that the soundness proof is trying to establish. It sould be noted that this circularity will be there independently of whether the formalisation of axioms of set-theory. These only add a further element of uncertainty insofar as there can be any doubt about *their* truth.

Of special significance is the fact that Set Theory is needed in the formalisation of the metamathematics of Set Theory itself. Here Set Theory plays the double role of object of investigation on the one hand and formalism within which the formalisation is being carried out on the other. This double role has given rise to forms of argumentation in which systermatic switches are made back-and-forth between the system as object- and as metaformalism.

validity is beyond controversy. As we will see, the axioms of ZF enjoy a coniderable degree of intuitive plausibility, though even among them it is possible to make out some differences in the kind or degree of self-evidence that attaches to them.

As a matter of fact the set-theoretical principles that are needed to formalise the more elementary parts of metamathematics (including all the results that were presented in Chs. 1 and 2) seem to be self-evident to a remarkable extent. Even if the combination of these principles with those of pure logic does go beyond what we now consider to be within the scope of pure logic, this does not seem to seriously affect the central purpose of the formalisation of metamathematics - to provide a proper foundation of scientific thought and reasoning.

Set Theory, then, can be seen as occupying a position halfway between logic and mathematics. On the one hand it seems to be about some particular matehamatical structure or structures, and as such it is on a par with other branches of mathematics. But on the other hand its central concepts, and the analyses of them that it has provided, come as close to what we would consider 'pure logic' as anything that doesn't actually lie squarely within it.

2. The Axioms of Set Theory.

In order to state the axioms of ZF we must first decide on a first order language in which they are to be expressed. We start with the assumption that this language has only one non-logical constant, the 2-place predicate ε, which designates the relation that holds between x and y when x is a member, or element, of the set y. As we go along, we will extend this language with new vocabulary, but always giving explicit definitions for the new notions in terms of the original ε. Thus each time a new predicate or function symbol is added to the language, the theory we are building is extended through the addition of a corresponding definition. As we have seen in Section 2.3, these additions always yield conservative extensions, which do not increase the set of theorems expressible in the original vocabulary $\{\varepsilon\}$.

The first principle that an axiomatic theory of sets should make explicit is the one which states what makes for the identity of a set. The principle we adopt, and which is in a sense definitory of the concept of set, is the *principle of extensionality*, according to which two sets are identical if and only if they have the same members:

SA 1. $(\forall x)(\forall y)\,(x = y \leftrightarrow (\forall z)(z \in x \leftrightarrow z \in y))$

The next three axioms tell us something about how to make new sets out of given sets. They testify to the possibility of forming *pairs*, *unions* and *power sets*, respectively

SA 2. $(\forall x)(\forall y)(\exists z)(\forall u)(u \in z \leftrightarrow (u = x \vee u = y))$

SA 3. $(\forall x)(\exists z)(\forall u)(u \in z \leftrightarrow (\exists v)(v \in x \,\&\, u \in v))$

SA 4. $(\forall x)(\exists z)(\forall u)(u \in z \leftrightarrow (\forall v)(v \in u \rightarrow v \in x))$

It is customary to denote the sets whose existence is asserted in SA2-SA4 as {x,y}, $\cup(x)$ and $P(x)$. Instead of '$\cup(x)$' and '$P(x)$' we also write '$\cup x$' and '$P\,x$'. {x} is short for {x,x},

N.B. these 'notational conventions' are our first examples of the mentioned practice in Set Theory to extend the language of set theory with new non-logical constants and the theory of set theory with axioms that have the form of explicit definitions for those constants. For instance, SA2 garantees the existence of an unordered pair for any two entities x and y, and it is easy to see that this pair is also unique. (This follows from the Extensionality Axiom SA1.) In other words the axioms so far adopted entail the following theorem:

(1) $(\forall x)(\forall y)(\exists z)((\forall u)(u \in z \leftrightarrow (u = x \vee u = y)) \,\&\, (\forall z')((\forall u)(u \in z' \leftrightarrow (u = x \vee u = y)) \rightarrow z' = z)))$

As we saw in Ch.2, (1) is the necessary and sufficient condition in order that adding the following definition (2) of the function constant {-,-} to any theory containing the axioms SA1 - SA2 yields a conservative extension.

(2) $(\forall x)(\forall y)(\forall z)(z = \{x,y\} \leftrightarrow (\forall u)(u \in z \leftrightarrow (u = x \vee u = y)))$

The same comment applies to the introduction of $\cup$ and P.

The next addition to our axiomatic theory is meant to capture the Comprehension Principle, the principle that for every property there exists a set which consists of just those entities which have the property (cf. Sn. 1.3.1). Here we encounter two difficulties. One of them is the problem that in this categorical form the Comprehension Principle cannot be true. (This is what Russell discovered when reading the ms.

of Frege's *Grundgesetze der Arithmetik* and explained in terms of 'Russell Paradox'.) So the best we can hope for is to adopt the principle in some weaker form.

In fact, there are two weakened versions of the Comprehension Principle which play a part in modern set theory. The first of these is due to Zermelo and the second to Fraenkel. Although the first version is logically entailed by the second, and thus the second sufficient by itself, we follow tradition in presenting both.

The first version is known as the *Aussonderungsaxiom*. This principle says that for any property P and any set x we can form the set *of those members of* x which have P. (That this is indeed a (weak) version of the Comprehension Principle follows if we assume that for each set there is the corresponding property of being a member of that set. For in that case we can form the complex property of (i) satisfying p and (ii) being a member of x; the set of all things satisfying this complex property is then the set which the Aussonderungsaxiom postulates for P and x.)

In trying to state the Aussonderungsaxiom within our language {ε} we encounter the second problem. Since we are working within first order logic, we do not have the means of quantifying over properties, and so wwe must make do with those properties which can be expressed within our language. So, just as for the Principle of Mathematical Induction in our formulation or Peano Arithmetic in Ch. 2, the best we can do is to specify the Aussonderungs-principle in the form of an axiom schema, i.e. as an infinite set of axioms, one for each formula A(u) of the language. As in the case of the Induction Schema PA7, we allow additional free variables $y_1,.., y_n$ in A. Thus the Aussonderungsaxiom takes the form given in SA5.[4]

SA 5. $(\forall x)(\forall y_1)...(\forall y_n)(\exists z)(\forall u)(u \in z \leftrightarrow (u \in x \ \& \ A(y_1,..,y_n,u)))$

---

[4] One might have thought that in the case of Set Theory there is no need to opt for an axiom schema: Instead of adopting an axiom for each formula A could we not quantify over sets, since sets are after all what Set Theory is about? Unfortunately this will not do. The claim - which would correspond to the categorical form of the Comprehension Principle - that for any set p there is a set z consisting of the members of p is a tautology; and the principle - corresponding to the Aussonderungsprinzip - that for any sets x and p there is a set z consisting of the members of x which are also members of p, while not actually tautologous, only asserts that the intersecion of two sets exists. This proves to be much weaker than the claim made by SA5 that every describable subset of a given set x exists.

Note that SA5. entails the existence of the intersection $x \cap y$ of two sets x and y. We obtain $x \cap y$ by applying SA5. to the formula '$u \in y$'. It is easily seen, moreover, that (4) satisfies the conditions for a definition of a 2-place function constant, and thus that we can extend our theory conservatively by adopting this defintion. (From now on we will adopt new vocabulary without making an explicit note that doing so is correct when this is obvious and/or the notation is familiar from informal treatments of Set Theory.)

(4) $(\forall x)(\forall y)(\forall z)(\forall u)(u \in z \leftrightarrow (u \in x \,\&\, u \in y)))$

The restriction which SA5 imposes on the Comprehension Principle is too severe and a set theory powerful enough to serve as framework for the formalization of mathematics and other areas of knowledge and reasoning needs something stronger. More specifically, we need a principle with the power to yield sets which are not subsets of sets that have already been constructed. The principle that has been adopted to this end, known as the "Replacement Principle"[5], is that the range of a function whose domain is a set is a set too. The Replacement principle too is a weakened version of the Comprehension Principle and one that (for all we know) is consistent.

In the formalisation of the Replacement Principle we have to deal with the same difficulty that we encountered in connection with the Aussonderungsaxiom. To state the principle we must speak about functions. But what is a function? Within set theory it is common to identify a function with its "course of values", i.e. with the set of all ordered pairs <a,b>, where a is an argument of the function and b is the corresponding value. Thus functions are sets, and if we make the usual identification of the ordered pair <a,b> with the unordered pair construct {{a}, {a,b}}, then functions are sets which are built out of their arguments and values by means that are entirely within the set formation repertoire we have already accepted in that it is entailed by the axioms SA1-SA5 already adopted.

If we were to formulate the Replacement Principle as involving functions in this sense, then we wouldn't get any sets whose existence cannot be proved from SA1-SA5. For suppose f is any function in this sense, i.e. a function-representing set of ordered pairs. Then the existence of a set consisting of the range of f is secured in any case by

---

5 The replacement Axiom is the axiom of ZF that is due to Fraenkel. It is also sometimes referred to as"Fraenkel's Axiom".

SA5, viz as the set of all u such that u ε x and A, where x is the set $\cup\cup f$ and A(v) is the formula $(\exists z)(\exists v)(z \in f \ \& \ z = \langle v,u \rangle)$. The existence of x is guranteed by SA3.

In order to get a version that enables to infer the existence of something whose existence isn't ascertainable in any case, we must once more make use of what can be described in our language of axiomatic set theory. This time what we want are descriptions of functions. That is, we need formulas A(u,v) with two free variables u and v, u for the argument of the function and v for the corresponding value. As in the case of SA5 we allow additional free variables $y_1,.., y_n$ in A.

This time we must be careful to make sure that our axiom schema does not overgenerate. If we allow aribitrary formulas A(u,v), then we are back at the contradiction that comes with the Comprehension Principle (e.g. by using for A(u,v) the formula '(u = u & ¬ v ε v)'. In order to avoid this we must restrict the A's that are permissible in the schema, to those which are 'functional for arguments which belong to the given set x':

(3) $(\forall u)(\forall v_1)(\forall v_2)(u \in x \ \& \ A(y_1,,y_n,u,v_1) \ \& \ A(y_1,,y_n,u,v_2) \rightarrow v_1 = v_2))$

Restricting the instances of A in the sense of (3) we obtain SA6 as formulation of the Replacement Schema

SA 6. $(\forall x)(\forall y_1)..(\forall y_n)((\forall u)(\forall v_1)(\forall v_2)(u \in x \ \& \ A(y_1,,y_n,u,v_1) \ \& \ A(y_1,,y_n,u,v_2) \rightarrow v_1 = v_2)) \rightarrow (\exists z)(\forall v)(v \in z \leftrightarrow (\exists u) (u \in x \ \& \ A(y_1,,y_n,u,v))))$

The axioms we have formulated so far represent a powerful set of principles to generate new sets from old ones. However, most of this power becomes relevant only when the sets involved are infinite. Generation of finite sets (more precsiely: the hereditarily finite sets) can be accomplished just with the axioms of pair formation and union, SA2 and SA3. However, there is nothing in the axioms we have so far adopted which entails the existence of any infinite set. One way in which this can be shown is to note that one the models for these axioms is the one we get when we iterate the operation $O(x) \equiv x \cup P\, x$ an infinite number of times, but stop at the first opportunity. As we observed earlier, the elements of this model are the hereditarily finite sets and it is straightforward to show that this is a model of SA1-SA6

So we need a further axiom - an "Axiom of Infinity" - to guarantee the existence of infinite sets. Interestingly, we need to postulate the existence of only one infinite set, for once such a set has been given, the axioms we have adopted generate a large (in fact, dazzlingly large) multitude of such sets.[6] There is a large number if different ways in which this requirement could be fulfilled. The form in which the axiom is usually giv en is as the claim that there is a set which (i) contains the empty set $\varnothing$ and (ii) contains for each of its members w also the 'successor' of w, i.e. the set $w \cup \{w\}$.

SA 7 $\quad (\exists y)(\varnothing \in y \ \& \ (\forall w)(w \in y \rightarrow (w \cup \{w\}) \in y))$

It should be intuitively clear that any set y which (i) contains $\varnothing$ and (ii) contains $w \cup \{w\}$ whenever it contains w must be infinite. In fact, we can prove that the sets $\varnothing, \varnothing \cup \{\varnothing\}, \varnothing \cup \{\varnothing\} \cup \{\varnothing \cup \{\varnothing\}\}$, ... are all members of such a y and also that they are all distinct from each other. In this way we can show that y has more elements than any finite number n.

It is easy to show that among the sets y which satisfy conditions (i) and (ii) there must be a minimal one. Let y_1 be any set satisfying (i) and (ii). If there is any other set y_2 which also satisfies these conditions, then the intersection $y_1 \cap y_2$ satisfies the conditions as well. So the smallest subset of y_1 which satisfies the conditions will necessarily be the smallest such set in absolute terms. Let S be the set of all subsets y of y_1 such that (i) $\varnothing \in y$ and (ii) $(\forall w)(w \in y \rightarrow (w \cup \{w\}) \in y)$ and let y_o be the set defined by

$(\forall v)(v \in y_o \leftrightarrow (v \in y_1 \ \& \ (\forall y)(y \in S \rightarrow v \in y)))$

Then clearly y_o satisfies (i) and (ii) and furthermore $y_o \subseteq y$ for every subset y of y_1 satisfying (i) and (ii).[7] So y_o is indeed the smallest set with these properties. An informal argument shows that y_o consists just of the sets $\varnothing$ (= "0"), $\varnothing \cup \{\varnothing\}$ (= "1"), $\varnothing \cup \{\varnothing\} \cup \{\varnothing \cup \{\varnothing\}\}$ (= "2"),... .

---

6 The need to postulate the existence of an infinite set was one of the disappointments of the so-called 'logicist programme', of which both Frege and Russell were advocates, to reduce mathematics to logic. It is hard too accept the existence of infinite sets as a principle that is valied for logical reasons.

7 Here of course "$y_o \subseteq y$" is short for "$(\forall z)(z \in y_o \rightarrow z \in y)$".

It should be clear that the set y_o is uniquely determined by the conditions we have used to define it. y_o is usually referred to as "ω". The set ω plays a pivotal role in Set Theory. We will soon meet it again when we will develop the concept of an ordinal. ω will be the *first transfinite ordinal.*

We are now in a position to give an impression of the importance of SA6. Given the existence of ω we can of course prove, using SA2 and SA3, that the sets $\omega \cup \{\omega\}$ (= "ω + 1"), $(\omega \cup \{\omega\}) \cup \{\omega \cup \{\omega\}\}$ (= "ω + 2"), ω +3, etc. exist as well. These sets form another infinite sequence, and it seesm reasonable to assume that this sequence too has a 'limit', just as the sequence 0, 1, 2, ... has the limit ω . But it is only with the help of SA6. that can show that this limiting set actually exists.

The argument goes as follows. Let A(x,y) be the formula:

$$(x = \varnothing \ \& \ y = \omega) \vee ((\exists u)(x = u \cup \{u\} \ \& \ (\forall w)\ (A(u,w) \rightarrow y = w \cup \{w\}))$$

It is easy to show that for all $n \in \omega$, (i) $(\exists v)$ A(n,v) and (ii) $(\forall v)\ (\forall w)(A(n,v) \ \& \ (A(n,w) \rightarrow v = w)$. To see this it is enough to observe that (a) ∅ is a set n satisfying (i) and (ii) and (b) if any set n satisfies (i) and (ii), then so does $n \cup \{n\}$. Since ω is by definition the smallest set S with the properties that $\varnothing \in S$ and that whenever $n \in S$ then $n \cup \{n\} \in S$, it follows that all members of ω satisfy (i) and (ii). To show (a) and (b) we proceed as follows. First, it is clear that there is exactly one set v such that A(∅,v), viz. ω. for when n = ∅ only the first disjunct of A is relevant. So (a) holds. Second, suppose that n satisfies (i) and (ii). Let y be the unique v such that A(n,v). To see that $n \cup \{n\}$ also satisfies (i) and (ii), note that now only the second disjunct of A is relevant. But from the second disjunct of A it is obvious that there is exactly one z such that A(n ∪ {n}, z), viz. the set $y \cup \{y\}$.

This shows that for all $n \in \omega$ there is exactly one w such that A(n,w). So we can apply SA6. with ω for x and the given formula A. The resulting instance of SA6. allows us to conclude that there is a set S which contains the sets "ω+ n" for all $n \in \omega$.

The Axioms SA1-SA7 make up what is often identified as "Zermelo-Fraenkel Set Theory" or ZF, after Ernst zermelo and Abraham Fraenkel,

the two mathematicians who were responsible for its formulation.[8] Often one adds to this system two additional axioms. The first seems to be evidently true of the structure of all sets as we intuit it, and so should, from our perspective, be included. This axiom expresses the idea that all sets are "built up from below". The idea is that when you take any set and try to make your way down to its "foundation" - by taking a member of the set, then a member of this member, then a member of that member, etc. - you must come to an end after a finite number of steps: There are no inifinite descending 'ε-sequences'.

That the following axiom expresses this intuition is not immediately obvious.:

SA 8. $(\forall x)\,(x \neq \varnothing \rightarrow (\exists y)(y \in x \,\&\, y \cap x = \varnothing))$

In fact, that SA8 does indeed prevent the existence of any infinite chain of sets s_n such that for all n $s_{n+1} \in s_n$, is quite involved and exploits deduction strategies that are specific to formal set theory and that it would carry us too far at this point to explain in sufficient detail. Sometimes one distinguishes explicitly between the theory axiomatised by SA1-SA7 ("ZF without Foundation“) and the one axiomatised by SA1-SA8 ("ZF with Foundation"). We will assume that SA8 is part of what we call ZF.

The last axiom - the *Axiom of Choice* (AC) - is generally regarded as more difficult to justify on intuitive grounds than those we have already considered. For this reason it is usually not considered as an integral part of ZF as such. But it has a reasonable degree of plausibility nonetheless, and it entails a large number of important set-theoretic results which cannot be proved without it. For this reason it has become standard practice to distinguish between ZF with and without AC. (The combination wird usually denoted as ZF+AC.)

The Axiom of Choice can be formulated in an astoundingly large number of different ways, some of which are very different from each other. But all of them can be shown equivalent on the basis of the axioms SA1 - SA7, so which formulation one chooses doesn't really matter in the end. In its perhaps most familiar form the axiom says that for any set x whose members are non-empty sets there exists a

---

[8] Fraenkel's only contribution to ZF is the Replacemernt Schema. We have just had a glimpse of the importance of this axiom, and we will soon have plenty of additional evidence. In fact the role of SA6 within ZF is so crucial, that it fully justifies the inclusion of Fraenkel's name in the designation of the theory.

function f with domain x which selects for each y in x a value f(y) that is an element of x. Note well that in this case the function which the AC asserts to exist is a function in the sense of set-theoretic object, i.e. a set of ordered pairs.

SA 9. $(\forall x)\,((\forall y)\,(y \in x \rightarrow y \neq \varnothing) \rightarrow (\exists f)(\text{function}(f) \,\&\, \text{Dom}(f) = x \,\&\, (\forall y)\,(y \in x \rightarrow f(y) \in y))$[9]

Experience with the theory ZF has shown that essentially all the theorems of set theory that have been proved by methods accepterd within mathematics can be formulated and formally derived within it. As an example, consider Cantor's Theorem, according to which there exists no injection of the power set $P(x)$ of a given set x into x.

Cantor's Theorem asserts that there exists no function of a certain kind. This involves quantification over functions. Since in ZF we can quantify only over sets we must once again make use of the set-theoretical concept of a function according to which it is a set of ordered pairs. Thus we come to the following formal statement (4) of the theorem.

(4) $(\forall x) \neg (\exists f)(\text{Dom}(f) = P(x) \,\&\, \text{Ran}(f) \subseteq x)$

(Here "Dom(f) = P(x)" is to be understood as in the explanation of SA9. and "Ran(f) $\subseteq$ x" is short for $(\forall v)(\exists u)\ \langle u,v\rangle \in f \rightarrow v \in x)$).

Within ZF the proof of Cantor's Theorem goes roughly as follows. Suppose that f were an injection of P(x) into x, for some set x. Let S be the set of all $u \in P(x)$ such that $\neg(f(u) \in u)$ - formally:

(5) $(\forall u)(\, u \in S \leftrightarrow u \in P(x) \,\&\, \neg(f(u) \in u)$.

That this set exists follows from SA5, taking P(x) as x and $\neg(f(u) \in u)$ as A(u). But now we can prove: $f(S) \in S \leftrightarrow \neg(f(S) \in S)$. Since this is a contradiction, the assumption that there exist x and f as hypothesized has been refuted; thus Cantor's Theorem has been proved.

---

9 The part beginning with "$(\exists f)$" would, in basic notation, be:
$(\exists f)((\forall u)(u \in f \leftrightarrow (\exists v)(\exists w)(v \in x \,\&\, u = \langle v,w\rangle) \,\&$
$(\forall y)\,(y \in x \rightarrow (\exists w)(\langle v,w\rangle \in f \,\&\, w \in y)))$

This 'proof' of Cantor's Theorem looks superficially very much like the proof that was presented in Ch. 1. But there is a difference of purport. The argument we have just presented is to be seen as an outline of what can be turned into a formal (i.e. axiomatic) derivation of the formal statement of Cantor's Theorem from the Axioms of ZF.
It should be emphasised that all proofs offered in this chapter should be understood in this way; they are all sketches of proofs that can be implemented as axiomatic derivations from ZF. In practice it hardly ever makes sense to carry out such derivations in full detail. Such derivations tend to conceal the ideas on which the proof is based behind a welter of formally necessary but intuitively trivial inference steps with which the intuitive ideas have next to nothing to do.

# #

Since ZF is a first order theory, it is subject to all the general results that apply to such theories. In particular, it is subject to the downward Skolem-Löwenheim Theorem. In the case of set theory this seems particularly puzzling. For suppose that ZF is consistent. (This is something we cannot prove. But now, after many decades of intimate experience with the theory which should have given much opportunit which should have given much opportunity to discover an inconsistency if indeed there was one, it seems very unlikely that the theory would be inconsistent after all.) Then ZF has a model (which, as can easily be shown, must be infinite) and so by Skolem-Löwenheim it must have a denumerable model - M, say. Clearly M is not the intended model of ZF. For the "real" structure of all sets is surely non-denumerable. For one thing, any model of ZF must, in view of the axiom of infinity, have a set "ω" and this set will be infinite, since it contains each of the sets $\varnothing, \varnothing \cup \{\varnothing\}, \varnothing \cup \{\varnothing\} \cup \{\varnothing \cup \{\varnothing\}\}$, ... and such sets will also be elements of the model and will all be distinct. But when ω belongs to the model, then so does $P(\omega)$ and this set is, by Cantor's Theorem, non-denumerable. In other words, there should be non-denumerably many elements in the model which all stand in the ε – relation to $P(\omega)$. But how can that be if M is only denumerable?

The paradox dissolves when we reflect on the exact meaning of Cantor's Theorem in the ZF formulation given above. In this formulation the theorem says that there is no "functional" set of ordered pairs which maps $P(\omega)$1-to-1 into ω. But does this really mean that $P(\omega)$ is non-denumerable? Well, it wouldn't if there weren't all that many functions within the model M, so that even if $P(\omega)$ is denumerable from an

external point of view, this fact could not be established within M for lack of the right function.

The existence of such models as M is thus no contradiction after all. It isn't a contradiction, because the axioms of ZF, while truly asserting the existence of such infinite sets as ω, do not succeed in truly asserting the existence of non-denumerably infinite powersets, such as $P(\omega)$. A denumerable set may behave, from the internal perspective of a given model, as non-denumerable simply because there are too few functions to expose it as a "fake non-denumerable" set, even though from an external perspective that is what it is, since an injection of it into ω does in fact exist.

Ordinals and Cardinals

We now proceed to develop the basics of an important part of set theory, the theory of ordinals and cardinals. We follow the now generally adopted approach originally due to Von Neumann.

Both the notion of an ordinal and that of a cardinal were invented by Cantor, as part of his attempts to develop a general consistent theory of infinite sets. Cantor was interested in particular in distinguishing between different kinds of infinity, something for which Cantor's Theorem provides the basis: The power set of any infinite set is x of a different, "higher" degree of infinity than is x itself. This distinction gives rise to the notion of *cardinality* and of *cardinal number*. Two sets have the same cardinality iff they can be injected into each other. Thus a set and its power set are always of distinct cardinality. Cantor then tried to develop a notion of cardinal number such that two sets have the same cardinal number iff they have the same cardinality.

Cantor also developed a more fine-grained method of counting infinite sets, which applies directly only to sets whose members are given in some order. The members of such sets would then be each assigned an ordinal number, and the set as a whole would be assigned the first ordinal number after all those assigned to members in it. Thus ordinal numbers were meant to be used as means of "counting" infinite sets in much the same ways as the natural numbers are used to count finite sets. This role that ordinal numbers were meant to play led to the idea that the class of all ordinals can be generated by the same kind of iterative procedure that is also assumed to generate the stucture of all sets: Each ordinal x gives rise to a next ordinal, the *successor* of x; and whenever a certain unbounded family of ordinals has been constructed, the limit of this family will once again be an ordinal, the first ordinal

after all the members of the family. The problem with such an inductive characterization of the generation process is that it is not quite clear how far it goes. For it is clear that not every unbounded family of ordinals will have an ordinal as limit. In particular the family of all ordinals - which is unbounded, as for each ordinal there is also its successor - cannot have such a limit. For if Ω were this ordinal, then Ω would be a member of the family of all ordninals and so would its successor. But then Ω would not come after all ordinals in the family: contradiction.

So, for which unbounded families of ordinals may it be assumed that limits exist? There seems no easy answer to this question. However, Von Neumann came up with a very ingenious solution, which consists in giving an explicit definition of a concept of "ordinal number", which apparently satisfies all the intuitive requirements that Cantor and the set theorists coming after him demanded of it. In this definition the successor of an ordinal x is defined by the operation we have already encountered a number of times, viz. as $x \cup \{x\}$. Von Neumann's explicit definition of the property of being an ordinal identifies the ordinals with those sets which are (i) linearly ordered by ε and (ii) are transitive - a transitive set being one which has the property that the members of its members are also members of it. Here is the formal definition:

Definition. A set x is an *ordinal* iff

(i) x is linearly ordered by ε, i.e. we have for all members u, v, w of x:

(a) $(u \in v \ \& \ v \in w) \rightarrow u \in w$

(b) $u \in v \ \vee \ u = y \vee \ v \in u$

(ii) x is *transitive*, i.e. for any y and z such that $y \in x$ and $z \in y$, we have $z \in x$.[10]

---

10 In the version of ZF we have presented here, in which the well-foundedness axiom SA8 is one of the axioms, this definition is adequate in the sense that oit supports all the theorems about ordinaly which follow. There also developments of set theory in which well-foundedness is not taken for granted - that is, SA8 is not adopted aas an axiom, or at least not from the outset. Withn such a weaker set-theory it is still possible to develop the theory of the ordinals on the basis of an explicit defintion, but now this definition must include the clause that an ordinal x is a set of sets which is well-ordered by ε - that is: if x is not empty, then there is a member of x which contains no member of x. (Exercise: Check that with this extra clause in the definition of 'ordinal' all the proofs which follow can be carried out without the use of SA8)

We write "Ord(x)" to express that x is an ordinal.

We can prove, in the order in which they are listed, the following theorems about ordinals:

Theorem O1. Ord(∅); Ord({∅}); Ord({∅,{∅}}); etc.

Theorem O2. $(\forall x)(Ord(x) \rightarrow Ord(x \cup \{x\}))$

Proof. Suppose that Ord(x). So x is transitive and linearly ordered by ε. We must show (i) that $x \cup \{x\}$ is linearly ordered by ε and (ii) that x U {x} is transitive. (ia). Let u, v, w ε $x \cup \{x\}$ such that u ε v ε w. When u, v, w ε x, then u ε w, since Ord(x). If u = x or v = x then we have a violation of axiom SA 8. So the only remaining possibility is that where u, v ε x and w = x. But then again u ε w. (ib) Suppose that u, w ε $x \cup \{x\}$. We want to show that u ε w v u = w v w ε u. If u, w ε x , this follows from the fact that Ord(x). If u = x & w = x then u = w; if u ε x & w = x, then u ε w; if w ε x & u = x, then w ε u. (ii) Let u ε w ε $x \cup \{x\}$. We want to show that u ε $x \cup \{x\}$. If w ε x , then u ε x because x is transitive, so u ε $x \cup \{x\}$. If w = x, then again u ε x and so u ε $x \cup \{x\}$.

Theorem O3. $(\forall x)(Ord(x) \rightarrow (\forall y)(y \in x \rightarrow Ord(y))$

Proof: Exercise

Theorem O4. $(\forall x)(\forall y)((Ord(x) \ \& \ Ord(y)) \rightarrow (x \in y \ v \ x = y \ v \ y \in x))$ (1)

Proof. Suppose the theorem does not hold. Then there is a counterexample to (1), i.e. there are x, y such that

(2) $(Ord(x) \ \& \ Ord(y)) \ \& \ \neg(x \in y) \ \& \ x \neq y \ \& \ \neg(y \in x)$.

With regard to x there are two possibilities: (a) there is no x' ε x such that (2) holds with x' for x and some y' or other for y. (b) there exists such an x'. In this second case we can form the set of all those x' ε x for which there is a y' so that x' and y' satisfy (2). Since this set is by

assumption non-empty, it has by SA8, a member x_o whose intersection with the set is empty. For this x_o we are then in case (a). Having thus obtained a minimal x_o we can now also find a minimal y_o among the y which jointly with x_o provide a counterexample to (1). Now let u be any member of x_o. Then, since $Ord(x_o)$, Ord(u). Since also $Ord(y_o)$ and x_o, y_o form a minimal counterexample to (1), we have: $u \in y_o \vee u = y_o \vee y_o \in u$. When $u = y_o \vee y_o \in u$, then $y_o \in x_o$, contrary to assumption. So $u \in y_o$. Since this holds for arbitrary $u \in x_o$, we have

(3) $(\forall u)(u \in x_o \rightarrow u \in y_o)$

Now let w be any member of y_o. Then as above we infer from minimality of y_o that $w \in x_o \vee w = x_o \vee x_o \in w$, and, again as above, that of these three possibilities only $w \in x_o$ is a live option. So we get

(4) $(\forall w)(w \in x_o \rightarrow w \in y_o)$

From (3) and (4) we get by extensionality: $x_o = y_o$, which contradicts the assumption that x_o, y_o satisfy (2). So (1) holds without exception.

Theorem O5. $(\forall x)((\forall y)((y \in x \rightarrow Ord(y)) \rightarrow Ord(\cup x))$

Proof: Exercise.

Theorem O6. $Ord(\omega)$

Proof. The strategy we will follow is to show that (a) all members of ω are ordinals and (b) that $\omega = \cup\omega$. Since by Theorem O5 and (a) $Ord(\cup\omega)$, (b) completes the proof.

(a) Let S be the set of all $x \in \omega$ such that Ord(x). (This set exists in virtue of SA5.) It is easy to show that S satisfies the conditions (i) $\varnothing \in S$ and (ii) $(\forall w)(w \in S \rightarrow w \cup \{w\} \in S)$. So, since ω is the smallest set satisfying these conditions, $\omega \subseteq S$. This concludes the proof of (a).

(b) First suppose that $u \in \omega$. Then $u \cup \{u\} \in \omega$. so there is a y such that $u \in y \in \omega$. So $u \in \cup\omega$. To show that $\cup\omega \subseteq \omega$ we proceed as under (a): Let S' be the set of all x in ω such that $(\forall w)(w \in x \rightarrow w \in \omega)$. Again we can show that S' satisfies the two conditions (i) and (ii) mentioned under (a). So $\omega \subseteq S'$. So if $u \in y \in \omega$, then $u \in \omega$. Now suppose that

$u \in \bigcup\omega$. Then for some y, $u \in y \in \omega$. So $u \in \omega$.

ω is our first example of an ordinal which is unbounded, in the set that for each $x \in \omega$ there is a $y \in \omega$ such that $x \in y$. Such ordinals are also called *limit ordinals*. If an ordinal is not a limit ordinal. it is, according to Thm O7 below, always of the form $w \cup \{w\}$. Such ordinals are called *successor ordinals*:

Definition. *LimOrd*(x) iff Ord(x) & $x \neq \varnothing$ & $(\forall w)(w \in x \rightarrow (\exists v)(w \in v$ & $v \in x))$

SuccOrd(x) iff Ord(x) & $(\exists v)(x = v \cup \{v\})$

Theorem O7. If Ord(x), then either (i) x = or (ii) SuccOrd(x) or (iii) LimOrd(x).

Proof: Exercise.

We already showed that with the help of SA7 we can prove the existence of the limit of the ordinals ω , $\omega + 1$, $\omega + 2$, ... (This is the ordinal we denoted as $\omega + \omega$.) In fact, SA7 makes it possible to prove the existence of a huge, barely surveyable, spectrum of limit ordinals beyond ω. Nevertheless, all ordinals that can be obtained by such methods are denumerable, i.e. stand in one-one correspondence with ω. To prove the existence of non-denumerable ordinals we have to appeal to a principle of a very different sort, which is implicit in the Axiom of Choice SA9. To establish this principle, the so-called Well-ordering Theorem, we need another, equally fundamental result, known as the Recursion Theorem.

The Recursion Theorem says, roughly, that recursive definitions along the ordinals constitute a valid means of defining functions. The theorem can be stated in a variety of ways. The one chosen here is inspired partly by the specific use to which we will put the theorem below.

In order to facilitate the statement of the theorem and the formulation of its proof, we introduce two notatonal devices. The first is a matter of strightforward definition. It will be conventient to have a compact notaton for the restriction of a function f to a certain set X. This restriction is the function whose domain is the intersection of X with the domain of f and which assigns to the arguments in its domain the

same values as f. To indicate restriction we use the symbol "⌈". Thus "f⌈X" stands for the set of all pairs <x, y> such that <x, y> ε f and x ε X.

The second bit of notation is a little more involved and needs to be handled with more care. One of the most common devices in natural language is the definite dscriptive term, such as "the King of France" or "the smallest perfect number" or "the empty set". The semantics of such terms is apparently that they denote the unique thing satisfying their descriptive content (i.e. the property expressed by their common noun phrase), provided there is just one such thing; but when there is no such thing, or if there is more than one, then there seems to be something wrong with the description - it is no longer clear what the description denotes; arguably it doesn't denote anything. Because of the danger of denotation failure, the device of definite description isoften excluded from the notational repertoire of formal logic, a policy which we have been following here too. But sometimes the device is handy and allows for more perspicuous formulas than would be availableotherwise. And since that will be the case in the Recursion Theorem to be stated presently, we introduce the device now.

For any variable x and formula A (typically, with free occurrences of the variable x, though strictly speaking we do not need to make this restriction) let "(Tx)A" stand for the unique x such that A(x). We will use this expression as a term, i.e. as occupying argument positions of predicates. Thus we will write for instance "P(c, (Tx)A)" to express the proposition that c stands in the relation P to the unique x such that A. However, we will only do so in contexts in which the unique existence of such an x is guaranteed, i.e. where the formula

(*) $(\exists x)\,(A(x)\ \&\ (\exists y)\,(A(y) \rightarrow x = y))$

holds. Note that where this conditon is fulfilled we can eliminate every occurrence of (Tx)A using notaton we already have. For instance, "

$$P(c,\ (Tx)A)$$

can then be rewritten as

$$(\exists x)\,(A(x)\ \&\ (\exists y)\,(A(y) \rightarrow x = y)\ \&\ P(c, x)).$$

When the formula in which the term "(Tx) A" occurs is complex, there are usually a numer of different ways in which its elimination might be carried out. For instance, we might get rid of the term from the

sentence $\neg P(c, (Tx)A)$ either by placing the quantificational complex inside the scope of or outside it, getting, respectively, (a) and (b):

(a) $(\exists x)\,(A(x) \,\&\, (\exists y)\,(A(y) \rightarrow x = y) \,\&\, \neg P(c, x))$.

(b) $\neg(\exists x)\,(A(x) \,\&\, (\exists y)\,(A(y) \rightarrow x = y) \,\&\, P(c, x))$.

But under the required conditions (i.e. that (*) holds) such alternative eliminations are provably equivalent.

Exercise. Show that

$$\vdash (*) \rightarrow (\,(a) \leftrightarrow (b)\,)$$

Equipped with these additrional means of notation we return to the Recursion Theorem. Suppose we want to define a function $f(\alpha, x_1,..., x_n)$, where α ranges over an ordinal γ and the x_i over some set X, and that we want to do this by (i) specifying, for arbitrary $x_1,..., x_n \in X$, the values of $f(0, x_1,..., x_n)$; (ii) specifying for arbitrary $x_1,..., x_n \in X$ and successor ordinal $\alpha + 1 \in \gamma$, the values of $f(\alpha + 1, x_1,..., x_n)$ on the basis of those of $f(\alpha, x_1,..., x_n)$; and (iii) specifying for arbitrary $x_1,..., x_n \in X$ and limit ordinals $\lambda \in \gamma$, the values of $f(\lambda, x_1,..., x_n)$ on the basis of the set of all $f(\beta, x_1,..., x_n)$ with $\beta < \lambda$, λ and $x_1,..., x_n$. Then a function f satisfying just those stipulations will indeed exist. (In fact, the proof of the theorem indicates a method for constructing an explicit definition of this function and prove of this definition that it is a proper defnition in the sense that it is satisfied by exactly one object, which satisdfies the imposed criteria. But this is an further aspect of the Recursion Theorem that we will notgo into here.)

More precisely, let $A(x_1,..., x_n, y)$ be a formula which is "functional in y" provided the $x_1,..., x_n$ are taken from $X_1,..., X_n$, i.e.

(1) $(\forall x_1)(\forall x_2)..(\forall x_n)(\forall y)(\forall z)(x_1 \in X_1 \,\&\, ...\&\, x_n \in X_n \rightarrow$
$(A(x_1,..., x_n, y) \,\&\, A(x_1,..., x_n, z) \rightarrow y = z)\,)$

Similarly, let $B(x_1,..., x_n, u, v, y)$ and $C(x_1,..., x_n, u, v, y)$ be formulas which express a functional dependency of y on any $x_1,..., x_n \in X_1,..., X_n$, arbitrary u and $\alpha \in \gamma$:

(2) $(\forall x_1)(\forall x_2)..(\forall x_n)(\forall u)(\forall \alpha)(\forall y)(\forall z)(x_1 \in X_1 \,\&..\&\, x_n \in X_n \,\&$

$\alpha + 1 \in \gamma \rightarrow (B(x_1,...,x_n,u, \alpha + 1, y) \ \& \ B(x_1,...,x_n,u,\alpha + 1, z) \rightarrow y = z))$

(3) $(\forall x_1)(\forall x_2)..(\forall x_n)(\forall u)(\forall \lambda)(\forall y)(\forall z)(x_1 \in X_1 \ \&..\& \ x_n \in X_n \ \& \ \lambda \in \gamma$
$\& \ \text{limord}(\lambda) \rightarrow (C(x_1,...,x_n, u, \lambda, y) \ \& \ C(x_1,..., x_n, u, \lambda, z) \rightarrow y = z))$

Then there is a unique function f which is defined on the $X_1,..., X_n$ and and which, for arbitrary $x_1 \in X_1,. .., x_n \in X_n$, and $\beta + 1, \lambda \in \gamma$ satisfies the following three conditions:

(i) $f(0, x_1,..., x_n) = \text{Ty } A(x_1,..., x_n, y)$
(ii) $f(\beta + 1, x_1,..., x_n) = \text{Ty } B(x_1,...,x_n, f\lceil(\beta+1), \beta, y)$
(iii) $f(\lambda, x_1,..., x_n) = \text{Ty } C(x_1,...,x_n,f\lceil\lambda, \lambda\rceil, y)$

Proof of the Recursion Theorem:

We begin by proving that

(*) For fixed $x_1 \in X_1,. .., x_n \in X_n$ there exists a function $f_{\{x_1,..., x_n\}}$ defined on γ such that the clauses (i), (ii) and (iii) hold for the given $x_1,..., x_n$ and arbitrary $\beta + 1, \lambda \in \gamma$.
(We omit the subscript $\{x_1,..., x_n\}$ for ease of notation).

We prove by induction on ordinals $\beta < \gamma$ the following statement:

(4) (1) There exists exactly one function f^β with domain equal to $\beta + 1$ and which, for ordinals belonging to $\beta + 1$ satisfies the clauses (i), (ii), (iii); and
(2) whenever $\delta < \beta$, then $f^\delta \subseteq f^\beta$.

We consider the three cases (a) $\beta = 0$; (b) $\beta = \alpha + 1$; and (c) $\beta = \lambda$, where limord(λ)

(a) Let

(5) $f^0 = \{<0, \text{Ty } A(x_1,..., x_n, y)>\}$.

It is easy to verify that (4.1) and (4.2) are both satisfied.

(b) Assume (4) for ordinals $< \alpha + 1$. Let

(6) $f^{\alpha+1} = f^\alpha \cup \{< \alpha + 1, \text{Ty } B(x_1,..., x_n, f^\alpha, \alpha + 1, y) >\}$.

It is easy to see that $f^{\alpha+1}$ satisfies the conditions (i)-(iii). To see that it is the only such function, suppose there are two such functions, g and g'. Then for some β, $g(\beta) \neq g'(\beta)$. Let δ be the smallest such β. If $\delta < \alpha + 1$ then $g\restriction(\delta + 1) \neq g'\restriction(\delta + 1)$. But it is easy to verify that both $g\restriction(\delta + 1)$ and $g'\restriction(\delta + 1)$ are functions with domain $\delta + 1$ which satisfy conditions (i)-(iii). So by induction hypothesis they are both identical to f^{δ}, and so must be identical to each other: contradiction. The remaining possibility is that $\delta = \alpha + 1$. But then $g\restriction(\delta + 1) = g'\restriction(\delta + 1) = f^{\alpha}$. Since g and g' also satisfy clause (ii) for the case where $\beta = \alpha$, it is easily verified that they are both equal to $f^{\alpha+1}$ as defined in (6).
Finally, let δ be any ordinal $< \alpha + 1$. Since $f\restriction(\delta + 1)$ has domain $\delta + 1$ and evidently satisfies (i)-(iii), it follows by induction that

(7) $f^{\delta} = f\restriction(\delta + 1) \subseteq f^{\alpha+1}$.

(c) Let λ be a limit ordinal $< \gamma$ and assume (4) for all ordinals $< \lambda$. We put

(8) $f^{\lambda} = \bigcup_{\beta<\lambda} f^{\beta} \cup \{< \lambda, \text{Ty } C(x_1,..., x_n, \bigcup_{\beta<\lambda} f^{\beta}, \lambda, y) >\}$.

Note that since for all $\delta < \beta < \lambda$, $f^{\delta} \subseteq f^{\beta}$, $\bigcup_{\beta<\lambda} f^{\beta}$ is a function. So f^{λ} is a function too. Again it is easy to verify that this function satisfies (i)-(iii), that its domain is $\lambda + 1$. To show that it is the only function with these properties and that for $\beta < \lambda$, $f^{\beta} \subseteq f^{\lambda}$, one proceeds as under (b).

To obtain the existence of a function f defined on $X_1 \times ... \times X_n \times \gamma$ which satisfies (i) - (iii) for arbitrary $x_1 \in X_1,..., x_n \in X_n$, and arbitrary $\alpha \in \gamma$, we observe that we could have proceeded just as well in the proof just given by adding at each stage pairs of the forms (5), (6) and (8), resepctively for all possible combinations of $x_1 \in X_1,..., x_n \in X_n$. It is easily seen that the above proof goes through essentially unchanged.

The recursion Theorem enables us to assert the existence of, among many other things, certain "arithmetical" operations on ordinals, in particular ordinal addition and multiplication. That is, for any ordinal γ there are 2-place functions $+_{\gamma}$ and $._{\gamma}$ defined on $\gamma \times \gamma$ such that the following holds for ordinals $\alpha, \beta < \gamma$:

(i_+) $\alpha +_{\gamma} 0 = \alpha$

(ii_+) $\alpha +_\gamma (\beta+1) = (\alpha+_\gamma \beta) +1$

(iii_+) $\alpha +_\gamma (\lambda) = \bigcup_{\beta \varepsilon \lambda} (\alpha+_\gamma \beta)$

$(i_.)$ $\alpha ._\gamma 0 = 0$

$(ii_.)$ $\alpha ._\gamma (\beta+1) = (\alpha._\gamma \beta) + \alpha$

$(iii_.)$ $\alpha ._\gamma (\lambda) = \bigcup_{\beta \varepsilon \lambda} (\alpha._\gamma \beta)$

For finite ordinals these operations are just the addition and multiplication familiar from ordinary artihmetic. To be precise, $+_\gamma$ is the set of all triples $\langle\langle n,m\rangle, n + m\rangle$, where n and m are finite ordinals and "+" is the operation of ordinary arithmetical addition on the natural numbers (which according to the set-theoretical perspective just are the finite ordinals); and similarly for $._\gamma$. However, for infinite ordinals the operations behave in a way which is quite surprising for someone used to the "plus" and "times" on the natural numbers. For instance, neither additoion nor multiplication are in general commutative. This is a consequence of a kind of absorption that happens when the left argument the operation is much smaller than its right argument. Thus we have in particular:

(OA.1) If n is finite and α is infinite, then
(i) $n + \alpha = \alpha$
(ii) $n \cdot \alpha = \alpha$

So we have for instance: $1 + \omega = \omega$ and $2 \cdot \omega = \omega$; and since $\omega \neq \omega + 1$ and $\omega \neq \omega \cdot 2$, the commutative laws "$\alpha + \beta = \beta + \alpha$" and "$\alpha \cdot \beta = \beta \cdot \alpha$" and are not generally valid.

Exercise: prove (OA.1) and the inequalities following it.

On the other hand the associative laws hold without exception:

(OA.2) (i) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$
(ii) $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$

Exercise: Of the following two putative laws one is generally valid while the other is not. Prove the validity of the valid one and give a counter-example to the other one:

(OA.3) (i) $(\alpha + \beta) \cdot \gamma = (\alpha \cdot \gamma) + (\beta \cdot \gamma)$
(ii) $\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$

Well-Foundedness and the Well-Ordering Theorem.

The next important theorem we need to establish is the so-called Well-ordering Theorem, which asserts that every set can be put into a 1-1 correspondence with some ordinal. We can also express this using the term *equipollent*.

Def. Let X and Y be sets. X and Y are *equipollent* iff there exists a bijection from X to Y.

So we can also express the Well-ordering Theorem by saying that every set is equipollent with some ordinal.

The Well-ordering Theorem implies - and this is what has given it its name - that every set X can be well-ordered, i.e. that there exists for X a binary relation (i.e. a set of ordered pairs) R which (i) is transitive, (ii) asymmetric and (iii) has the property that for every non-empty subset Y of X there is a $y \in Y$ such that for all $z \in Y$, if $z \neq y$ then yRz. (N.B a relation R with these three properties is in particular linear, i.e. for each x, y in the field of R, we have $xRy \vee x = y \vee yRx$. Show this.) For evidently the correspondence between X and some ordinal entails the existence of such a well-ordering. (Exercise: Show this.)

Well-ordering Theorem.

Every set X is equipollent to some ordinal.

Proof. Let X be any set. If X is the empty set there is nothing to prove. So we assume that X is non-empty. We proceed as follows. We consider the set $\mathbb{R}$ of all well-orderings of subsets of X. (That this set exists is easily seen. For each well-ordering of a subset of X is a set of ordered pairs of members of X. Since the ordered pairs of members of X form a definable subset Z of $P(P(X))$, the set of all well-orderings of subsets of X is a subset of $P(Z)$.) Moreover, this subset is definable (by the three properties (i), (ii), (iii) mentoned in the definition of well-ordering above). So $\mathbb{R}$ is a set.)

We first show that each such well-ordering R determines a unique order preserving map from R onto some ordinal α_R, i.e. a unique 1-1 function f_R onto α_R such that for all x, y in the field of R, xRy iff $f_R(x) \in f_R(y)$. We argue as follows. Let Y be the field of R. For each $y \in Y$ understand

by *the R-initial segment of* Y determined by y that subset Z of Y which consists of y and all z ε Y such that z R y. It is enough to show that for each y ε Y there exists a unique order-preserving map f_y from the R-initial segment determined by y onto some ordinal α_y and that moreover the f_y are nested, i.e. that if z R y, then $f_z \subseteq f_y$. (For either there is an R-last element u in Y, in which case Y is identical with the R-initial segment of Y determined by u; or else there is no last element, but then the union of all the functions f_y for y ε Y will be, since the f_y are nested, an order-preserving map from Y onto the union of the α_y.) Suppose there is a y for which there is no f_y as described. Then, since R is a well-ordering, there is a R-first such y. Either this y has an immediate R-predecessor z in Y. But then there is a unique order-preserving map f_z from the segment determined by z onto some ordinal α_z. So if $f_y = f_z \cup \{\langle y, \alpha_z \rangle\}$, then f_y is a unique order-preserving map from the segment determined by y onto $\alpha_z + 1$. If y does not have an immediate R-predecessor, then we put $f_y = \bigcup_{zRy} f_z \cup \{\langle y, \bigcup_{zRy} \alpha_z \rangle\}$. Again we conclude, now also using the nestedness of the f_z, that f_y is a unique order-preserving map from the segment determined by y to some ordinal. So in both cases we get a contradiction.

Let $\gamma = \bigcup_{R \in \mathbb{R}} \alpha_R$. We now make use of the Axiom of Choice, assuming that there exists a function g defined on the set of non-empty subsets of X such that for any such subset Z, g(Z) ε Z. We also use the Recursion Theorem. This allows us to assert that there exists a function f defined on γ + 1 which satisfies the following clauses:

(i) $f(0) = g(X)$

(ii) $f(\alpha+1) = g(X - \mathrm{Ran}(f\lceil(\alpha+1))$, if $X - \mathrm{Ran}(f\lceil(\alpha+1) \neq \emptyset$;
X otherwise;

(iii) $f(\lambda) = g(X - \mathrm{Ran}(\bigcup_{\beta<\lambda} f\lceil\beta))$, if $Y - \mathrm{Ran}(\bigcup_{\beta<\lambda} f\lceil\beta) \neq \emptyset$;
X otherwise.

Note that once $f(\alpha) = X$ then this will remain so for $\beta > \alpha$ - i.e. we also have $f(\beta) = X$. For "$f(\alpha) = X$" means that all of X has been exhausted by the time we reach α (i.e. $X \subseteq f\lceil\alpha$). Moreover, for each α such that $f(\alpha) \neq X$ the relation R_α defined by:

$\langle u,v \rangle \in R_\alpha$ iff there are δ, β such that $\delta < \beta$, $f(\delta) = u$ and $f(\beta) = v$

is a well-ordering and α is the ordinal $\alpha_{(R\alpha)}$ corresponding to this well-ordering in the sense of the first part of the proof. Therefore $\alpha \in \gamma$. So $f(\gamma) = X$ and consequently the first ordinal β such that $f(\beta) = X$ belongs to $\gamma + 1$. But this means that $X - Ran(f\lceil\beta)$ is empty. So $f\lceil\beta$ is a 1-1 map from β onto X. q.e.d.

The Well-ordering Theorem makes it possible to compare all sets according to size, in the following sense. For each set X let $|X|$ denote the smallest ordinal α such that α is equipollent with X. Since any two ordinals α, β are comparable as to size - we have either $\alpha \in \beta$ or $\alpha = \beta$ or $\beta \in \alpha$ - the relation "$X < Y$" defined by

$$X < Y \text{ iff } |X| \in |Y|$$

is a strict linear order on the totality of all sets. $|X|$ is also called *the cardinality of* X, or *the cardinal of* X. And by a *cardinal*, or *cardinal number*, we understand any ordinal that is equal to its own cardinality, i.e. any ordinal α such that $\alpha = |\alpha|$. Note that every finite ordinal is also a cardinal, but that among the infinite ordinals cardinals are extremely rare. For instance, ω is a cardinal, but $\omega + 1$, $\omega + 2$,..., $\omega + \omega$, $\omega.3$, ... $\omega.\omega$, ... are all of the same cardinality as ω and thus are not cardinals. Nevertheless we do know that there are also larger cardinals than ω. For according to Cantor's Theorem no set is equipollent with its power set. So in particular the cardinal number of $P(\omega)$ - it is often referred to as "$beth_1$" - is different from, and thus is larger than, ω; and the cardinal of the power set of the power set of ω is bigger than and so forth. But how much bigger is $beth_1$ than ω? In particular, is it the next cardinal after or are there other cardinals in between? This question, which was already raised by Cantor, can be said to have been the single most important question in set theory since Cantor, Dedekind and others first laid its foundations in the second half of the nineteenth century. (Cantor himself is said to have worked on this question with such desperation that it led, or at any rate significantly contributed, to a condition of clinical depression) The hypothesis that $beth_1$ is the first cardinal after ω is known as the *Continuum Hypothesis*. (It is called this because, as can be shown without too much difficulty, is also the cardinality of the "mathematical continuum", i.e. of the set of all real numbers.) After many fruitless attempts to prove the Continuum Hypothesis (from the Axioms of ZF, or from other, intuitively plausible axioms), Gödel succeeded in 1940 to prove at least that the Hypothesis was consistent with ZF (in fact,

with a somewhat stronger theory known after its architects as "Gödel-Bernays") It was not until 1961 that Paul Cohen showed that the Continuum Hypothesis is *independent from* ZF, i.e. that its negation is consistent with ZF. Since then various attempts have been made to think of intuitively valid principles which would settle the question, even if the search for such principles has produced many intreresting results about ZF and its possible models.

With the cardinal numbers comes a "cardinal arithmetic" which must be sharply distinguished form the ordinal artithmetic mentioned earlier. We give just two operations here, cardinal addition, $+$, and cardinal multiplication, $\odot$:

For any cardinals κ, μ

(1) $\kappa + \mu = |X \cup Y|$, where X and Y are any sets such that $|X| = \kappa$, $|Y| = \mu$ and $X \cap Y = \emptyset$.

(2) $\kappa \odot \mu = |X \times Y|$, where X and Y are any sets such that $|X| = \kappa$ and $|Y| = \mu$

Some results about cardinal arithmetic:

(3) For arbitrary cardinals κ and μ

(i) $\kappa + \mu = \mu + \kappa$
(ii) $\kappa \odot \mu = \mu \odot \kappa$

(4) For all infinite cardinals μ and arbitrary cardinals κ

(i) if $\kappa \leq \mu$ then $\kappa + \mu = \mu$
(ii) if X is a set of cardinality $\leq \mu$ and for each $x \in X$, x is of cardinality $\leq \mu$, then $\cup_{x \in X} x$ has cardinality $\leq \mu$.

(5) if $\kappa \leq \mu$, then $\kappa \odot \mu = \mu$

Of these only (3), (4) and (5) deserve careful attention. The other properties are left as exercises. We begin with the comparatively simple (3).

Our proof of (3) is based on the following three observations. The first is:

(6) Every cardinal is a limit ordinal.

(Exercise: Prove this.)

The second observation is closely related to the second:

(7) Every infinite ordinal α can be written in exactly one way as the ordinal sum $\lambda + n$ of a limit ordinal λ and a finite ordinal n.

(7) is proved by an easy induction on ordinals. For $\alpha = 0$ the assertion is trivial. Suppose that $\alpha = \beta + 1$ and (6) holds for β. Then there are unique λ and n such that $\beta = \lambda + n$. Then clearly $\alpha = \lambda + (n+1)$. Moreover, if $\alpha = \beta + 1$ for some other pair of a limit ordinal μ and a finite ordinal m, then (i), as α is a successor ordinal, $m = k + 1$ for some finite ordinal k. But then $\beta = \mu + k$. Since by assumption the decomposition of is unique, $\mu = \lambda$ and $k = n$. Finally assume that α is a limit ordinal. Then obviously
$\alpha = \alpha + 0$. Moreover, if for any λ and n, $\alpha = \lambda + n$, then $n = 0$; for otherwise α would be a successor ordinal. So $\alpha = \lambda + 0 = \lambda$.

The third observation requires the following definition. For any limit ordinal λ let *the* ω*-sequence generated by* λ be the set $\{\lambda + n\}_{n \in \omega}$. We denote this set as $\Omega(\lambda)$. Note that if λ, μ are distinct limit ordinals, then $\Omega(\lambda) \cap \Omega(\mu) = 0$. Using this notion, we claim:

(8) For every limit ordinal λ,

(i) $\lambda = \omega \cup \bigcup_{\beta \in Z} \Omega(\beta)$,
where Z is the set of limit ordinals $< \lambda$.

(8) is fairly obvious: The members of a limit ordinal are either limit ordinals or successor ordinals. Clearly every limit ordinal is the only limit ordinal in its ω-sequence, all the other members of the sequence being successor ordinals. The limit of the sequence is again a limit ordinal. The successor ordinals, moreover, are, according to (7), all of the form $\mu + n$, where μ is a limit ordinal and n is some finite ordinal > 0. So it should be evident that the right hand side of (i) exhausts λ.

Now let X and Y be a pair of disjoint sets of cardinal λ and let f and g be bijections from X and Y to λ, respectively. These functions assign edach member x of X and each member y of Y unique ordinals α_x and α_y belonging to λ. By (7) these ordinals have unique representations α_x

$= \mu_X + n_X$ and $\alpha_Y = \mu_Y + n_Y$. We must construct a bijection of $X \cup Y$ to λ. The trick is to map X onto the "even" members of λ and Y onto the "odd" members. That is, we let h be the function which maps each $x \in X$ to the ordinal $\mu_X + 2.n_X$ and each $y \in Y$ to the ordinal $\mu_Y + (2.n_Y + 1)$. It should be obvious (i) (using (7)) that h is a 1-1 and (ii) (using (8)) that h is onto λ.

(4) and (5) are proved together. In the proof we make use of the fairly obvious inequality:

(9) if X is a set of cardinality $\leq \kappa$ and for each $x \in X$, x is of cardinality $\leq \mu$, then $|\cup_{x \in X} x|$ has cardinality $\leq |\kappa \times \mu|$.

(Exercise: Prove this)

We prove by induction on infinite cardinals μ that whenever κ is a cardinal $\leq \mu$, then $\kappa \oplus \mu = \mu$. We distinguish between three cases; (a) $\mu = \omega$; (b) $\mu = \kappa^+$, where κ^+ is the first cardinal after κ; (c) μ is a limit cardinal, i.e. for each cardinal $\kappa < \mu$ we also have $\kappa^+ < \mu$. Case (a) is left as an exercise. We consider case (b). Let X be the set of all limit ordinals between κ and κ^+. X is well-ordered by $\in$ and so there is a (unique) ordinal β and 1-1 $\in$-preserving map f_X from β onto X. Using the Axiom of Choice (henceforth: AC) we assume that h is a function defined on all subsets Y of μ such that $|\mu - Y| = \kappa$ which assigns to each such Y a subset h(Y) of $\mu - Y$ of cardinality κ. Similarly, using AC together with the Induction Hypothesis, we assume that bi is a function which assigns to any pair $\langle Y,Z \rangle$ of subsets of μ both of which are of cardinality κ a bijection bi(Y,Z) from Y to Z. For any ordinals δ, γ such that $\delta < \gamma$ let $[\delta, \gamma)$ be the set of all ordinals α such that $\delta \leq \alpha < \gamma$.
We define the function g by recursion on β as follows:

(i) $g(0) = bi(f_X(0) \times f_X(0),\ f_X(0))$

(ii) $g(\alpha+1) = g(\alpha) \cup bi(((f_X(\alpha +1) \times [f_X(\alpha), f_X(\alpha +1)) \cup ([f_X(\alpha), f_X(\alpha +1)) \times f_X(\alpha +1))) ,\ h(Ran(g(\alpha)))$

(iii) $g(\lambda) = \cup_{\delta<\lambda} g(\delta)$

With regard to (ii) it is important to note that the two arguments of bi are indeed both of cardinality κ and that if $g(\alpha)$ is a bijection between $f_X(\alpha) \times f_X(\alpha)$ and some subset of μ, then $g(\alpha +1)$ is a bijection between

$f_X(\alpha+1) \times f_X(\alpha+1)$ and some subset of μ. With regard to (iii) we may note that $g(\lambda)$ is a bijection from $f_X(\lambda) \times f_X(\lambda)$ to some subset of μ of cardinality κ. The conclusion that the range of $g(\lambda)$ is of cardinality κ we use the Induction Hypothesis together with (9).
It is easily seen that, as $\cup_{\alpha<\beta} f_X(\alpha) = \mu$, g is a bijection from $\mu \times \mu$ to some subset of μ. It follows that $\mu \times \mu$ and μ are equipollent.

The proof for case (c) is similar to that for case (b). This time let X be the set of all infinite cardinals $< \mu$. Let f_X and β be defined as before. It is easily verified that $\beta \leq \mu$. Let h be a function which asigns to pair consisting of a subset Y of μ with $|Y| < \mu$ and an infinite cardinal $\alpha < \mu$ a subset of μ - Y of cardinality α, and let bi be a function which assigns to each pair of sets Y, Z of the same cardinality $\alpha < \mu$ a 1-1 map bi(Y, Z) from Y onto Z. (Again the existence of such a function is entailed by the Induction Hypothesis.) This time let g be the function with domain β defined by the clauses (i) and (iii) above together with the clause

(ii') $g(\alpha+1) = g(\alpha) \cup bi(((f_X(\alpha+1) \times. [f_X(\alpha), f_X(\alpha+1)]) \cup ([f_X(\alpha), f_X(\alpha+1)) \times. f_X(\alpha+1))), h(Ran(g(\alpha)), f_X(\alpha+1))$

It is easy to verify that in (ii') both arguments of bi are of cardinality $f_X(\alpha+1)$ and thus that if $g(\alpha)$ is a 1-1 function from $f_X(\alpha) \times f_X(\alpha)$ to some subset of μ (of cardinality $f_X(\alpha)$), then $g(\alpha+1)$ is a 1-1 function from $f_X(\alpha+1) \times f_X(\alpha+1)$ to some subset of μ (of cardinality $f_X(\alpha+1)$). With regard to (iii) note that since $\lambda < \beta \leq \mu$, $|\lambda| < \mu$. So, using (9) we can once again conclude that the range of $\cup_{\delta<\lambda} g(\delta)$ has a cardinality not greater than the maximum of $|\lambda|$ and $f_X(\lambda)$ and thus of cardinality $< \mu$.

The set-theoretical results we have mentoned here are only a small excerpt from the vast stock of theorems of this theory (some of them extremely difficult) that are known. Our selection has been governed primarily by the need to provide a certain impression of the two principal ways of "counting the infinite" which set theory has made precise and which are associated with the concepts of *ordinal* and *cardinal*, respectively. More specifically - and this is true in partiuclar for the last few results - we have aimed at providing the set-theoretical underpinnings for the following "converse" of the Downward Skolem-Löwenheim Theorem, which was preeented on p. . This converse, the

"Upward Skolem-Löwenheim Theorem", says that if a set of sentences Δ has a denumerably infinite model, then for every infinite cardinal κ, Δ has a model whose universe is of cardinality κ.

Theorem. (Upward Skolem-Löwenheim Theorem.)
Let Δ be a set of sentence of some language L, let M be a model for L such that $|U_M| = \omega$ and $M \models \Delta$. Let k be any infinite cardinal $> \omega$. Then there exists a model M' such that $M \models \Delta$ and $|U_M| = \kappa$.

Proof. The proof is similar to that of the Completeness Theorem. Let Δ, M and κ be as in the statement of the theorem. To show that Δ has a model M' of cardinality κ we extend L to the language L' by adding to it a set of cardinality κ of new individual constants. We shall show presently that the sentences of L' can be enumerated in a sequence the length of which is exactly κ. (To be precise, that there exists a 1-1 function from κ to the set of sentences of L'.) But before we do this, a remark is in order about "languages" with a non-denumerably infinite vocabulary. So far we have considered only languages whose voacabulary was at most denumerable. Even a denumerably infinite, as opposed to a finite, vocabulary may perhaps seem a little counterintuitive from the perspective of our experience with actual languages. For the vocabularies of those languages, as normally understood, do appear to be finite. However, it is clear how a denumerbly inifinite vocabulary can be "simulated" with the help of a finite number of signs. As an eaxmple we may consider the vocabulary consisting of all *numerals*, i.e. all canonical names of natural numbers. Our standard decimal notation provides such names as combinations of the ten signs "0", "1", ... , "9". Alternatively, we can use, as numeral for the number n, the complex sign consisting of a "0" followed by n "1"s.

But a non-denumerable vocabulary cannot be simulated in this way, for the set of all finite sequences over some finite "aphabet" of signs will always be denumerable. (Exercise: Show this.) So the concept of a language with a non-denumerable vocabulary is an abstraction, or extrapolation, from our intuitive concpet of a language in a way that languages with denumeranbly infinite vocabularies are not. So what should we understand by such a non-denumerable language?

To focus on this question, we should be clear of the kind of abstraction involved in the notion of a non-denumerable set - such as. for instance, the cardinal κ. The existence of such sets follows from our axioms of set theory; and set theory offers various constructs to form non-

denumerable sets out of others (as well, of course, as out of denumerable sets). But obviously we are never in a position to actually display or enumerate such a set explicitly - that is precisely what the term "non-denumerable" conveys. In the light of these considerations it is reasonable to see non-denumerable languages also as set-theoretical constructs, or, more accurately, to see the sentences and other well-formed expressons of such languages as constructs from finite subsets of their non-denumerable vocabularies. But in what sense can a well-formed expression of a language L - i.e. a sequence of "words" of L, i.e. of items from L's vocabulary - be a set-theoretic object? The natural answer to this question would seem to be: To the extent that sequences are, or can be considered, set-theoretical objects.

So what is a sequence in the set-theoretical sense? Set Theory suggests two possible asnwers to this question. According to the first answer a sequence of two elements will be an ordered pair- thus <a,b> is the sequence consistingof the elements a and b. Similarly a sequence consisting of three elements, a, b and c, say, will be a triple, e.g. the pair consisting of <a,b> and c: <<a,b>,c>, etc. The second answ<er is that a sequence is a functon the domain of which is an ordinal, and whose values are the members of the sequence. Thus the sequence consisting of a and b is the function {<0,a>,<{0},b>}, or {<0,a>,<1,b>} - a function the domain of which is the ordinal 2 (i.e. the set {0,{0}}). Similarly the sequence consisting of a,b and c is the function {<0,a>,<1,b>, <2,c>}, etc. This second notion of sequence has the advantages that it can be defined once and for all by a single, simple, explicit definition and (ii) that it generalizes straightforwardly to the infinite: a sequence in this sense can be finite or infinite according as the ordinal that is its domain is finite or infinite.

Adopting this second notion of sequence, we come to the following characterization of non-denumerable languages. As before a language L is a function from symbols to signatures (see p. 1), where the possibility that the domain of L is non-denumerable is explicitly included. The terms and sentences of L are then finite sequences of members of the domain of L, where "sequence" is to be understood in the set-theoretical sense just indicated, which satisfy the clauses (i) and (ii) of the definition of *term* and the clauses (i)-(v) of the definition of *formula* on p.1.

Now that we have made precise what should be understood by the language L' and its terms and sentences, we return to the proof of our theorem. We first divide the set C of new constants into two sets C_1 and C_2 , each of cardinality k. Let $\Delta' = \Delta \cup \{\neg(c = c'): c \text{ and } c' \text{ are}$

distinct constants in $C_1\}$. It is easily seen that Δ' is consistent. For let A be a finite subset of Δ'. A will consist of some finite subset of Δ together with finitely many sentences of the form "$\neg\Box(c = c')$". In the model M the former are true by assumption. Moreover, since U_M is infinite, it is possible to chose distinct denotations in U_M for each of the finitely many new constants that occur in sentences in Δ' of the second kind.

We now come to the point where we need some of the cardinal arithmetic we have presented here and all that was required to get that far. It is clear that the cardinality of the sentences of L' is at least κ, for even the sentences which have the form "$c = c_o$", where c_o is some particular new constant and c is any new constant, already has cardinality κ. But is the set of sentences of L' *exactly* of cardinality κ? To see that this is so, we first observe that the set of symbols of L' has cardinality κ. This follows directly from (3) on p.44. Our second observation is that for each n the set of n-place sequences of members of L' has cardinality κ. For n = 1 this is obvious. Suppose the claim is true for n = m. To see that it is then also true for n = m + 1, note that every m+1-place sequence of members of L' is decomposable, in a unique way, into (i) an m-place sequence of members of L' and (ii) a member of L'. Thus the set of all m+1-place sequences is equipollent with the cardinal product of the cardinal of the set of m-place sequences and the cardinality of L'. By induction hypothesis this is equal to $\kappa \cdot \kappa$, which according to (4) on p. 44 is equal to κ. Our last observation is that the set of all finite sequences of members of L' has cardinality κ. This follows from the fact that this set can be written as $\cup_{n \in \omega} X_n$, where for n = 1,2,... X_n is the set of all n-place sequences of members of L'. It follows from (5) on p. 44 that this set is again of cardinality κ. Since teh set of sentences of L' is a subset of this set, its cardinaltity is at most κ. We already know that its cardinality is at least κ. So it is exactly κ.

From here on the proof closely follows the completeness proof we gave earlier. Let $\{A_\beta\}_{\beta \in \kappa}$ be an enumeration of length κ of all the sentences of L'. We use this enumeration to construct a sequence $\{\Delta_\beta\}_{\beta \in \kappa}$ of extensions of Δ'. As in the completeness proof, the union Δ_κ of this sequence will determine a model M' of Δ' and this M' will be the model we are looking for. We define by means of the clauses:

(i) $\Delta_0 = \Delta'$

(ii) $\Delta_{\beta+1} =$ (a) $\Delta_\beta \cup \{A_\beta\}$, provided $\Delta_\beta \cup \{A_\beta\}$ is consistent and A_β is not of the form $(\exists v_j)B$

(b) $\Delta_\beta \cup \{A_\beta, B(c/v_j\}$, provided $\Delta_\beta \cup \{A_\beta\}$ is consistent, A_β is of the form $(\exists v_j)B$ and c is a constant from C_2 which occurs neither in Δ_β nor in A_β.

(c) Δ_β, provided Δ_β U $\{A_\beta\}$ is inconsistent.

(iii) $\Delta_\lambda = \cup_{\beta \varepsilon \lambda} \Delta_\beta$

Note (i) that for all $\beta \varepsilon \kappa$ there is a c not occurring in Δ_β or A_β. For only $|\beta|$ new constants can have been introduced into Δ_β and only finitely many such constants can occur in A_β. Since there are κ new constants in all and $|\beta| < \kappa$, it follows that there are still κ constants left. Note (ii) that by the Recursion Theorem the clauses (i)-(iii) define a function defined on κ. The range of this function is a set and so is its union. Call this union Δ_κ.

As in the completeness proof one shows that Δ_κ is consistent and complete in L'. Also, defining once more the relation $\sim$ between individual constants of L' by:

$$c \sim c' \text{ iff}_{def} \text{ the sentence } c = c' \text{ belongs to } \Delta_\kappa$$

we show as before that $\sim$ is an equivalence relation and that whenever $c \sim c'$ and $P(t_1,.., c,..,t_n) \varepsilon \Delta_\kappa$, then $P(t_1,.., c',..,t_n) \varepsilon \Delta_\kappa$.
Moreover, since for any pair c, c' of distinct new constants the sentence $\neg(c = c')$ belongs to , all new constants belong to distinct equivalence classes under $\sim$. So, if we define the model M' in the same way as in the completeness proof, then $|U_{M'}| = \kappa$. As before one shows that for every sentence A in Δ_κ, $M \models A$.

The Interpretation of Number Theory in Set Theory.

It is common to think of the members of the set ω as the "natural numbers". ω has a number of properties that suggest such an identification. For instance, as we have seen, ω is linearly ordered by ε and this order has the same structure as the set on natural numbers: (i) it begins with the empty set (which it is therefore natural to identify with the number 0), (ii) has the property that each "number" n has an immediate successor $n \cup \{n\}$, as well as, if it is different from $\emptyset$, an immediate predecessor, and (iii) it runs on forever. However, a proper identification of ω with the natural numbers requires that we interpret all operations and relations of number theory as operations and relations on ω, and in such way that number-theoretic laws turn into theorems of set theory.

In this section we formulate such an interpretation of number theory within set theory. It will have the property that for any theorem of our axiom system of Peano arithmetic the interpretation of that theorem (a sentence in the language of set theory) will be a theorem of the set-theoretical axioms SA1 - SA7.

Before we do this, we will define in more general terms the notion of an interpretation of a theory T1, formulated in a first order language L1, within a second theory T2, formulated within a first order language L2. Any such interpretation will be based on interpretations of all the non-logical constants of L1 by formulae of L2. For instance, if R is a 2-place relation of L1, then an interpretation of R in L2 will take the form of an L2 formula $A_R(v_1, v_2)$ in which v_1 and v_2 are the only free variables. An example which we have encountered already in a somewhat different context is the interpretation of the relation $\leq$ of the theory of Boolean lattices in terms of the operation U of Boolean Algebras. We can interpret the theory of Boolean lattices within the theory of Boolean Algebras by interpreting $\leq$ by means of the formula $v_1 \cup v_2 = v_2$.

For function constants of L1 the matter is a little more complicated. Since function constants form terms, and not formulas, interpreting an L1 function by means of an L2 formula makes no direct sense; rather the interpreting formula should be thought of as interpreting certain atomic formulae in which the function constant occurs. For instance, an interpretation of the theory of Boolean Algebras within the theory of Boolean lattices must be based on, among other things, an interpretation of the2-place function constant U. This interpretation is

to be understood as the interpretation of the atomic formula $v_1 \cup v_2 = v_3$. A natural choice (also encountered earlier) would be the formula

(1) $v_1 \subseteq v_3 \;\&\; v_2 \subseteq v_3 \;\&\; (\forall v_4)(v_1 \subseteq v_4 \;\&\; v_2 \subseteq v_4 \rightarrow v_3 \subseteq v_4)$

In general, to interpret an n-place function constant we need an n+1-place formula $A_C(v_1, \ldots, v_n, v_{n+1})$. Note well that in order that for $A_C(v_1, \ldots, v_n, v_{n+1})$ to be suitable as the intepration of an n-place function constant, the last argument most be functional in the first n arguments, that is, we must have that for all relevant values of the ariables the following open formula is satisfied:

(2) $A_C(v_1, \ldots, v_n, y) \;\&\; A_C(v_1, \ldots, v_n, z) \rightarrow y = z$

In general, interpretation of T1 within T2 involves yet another L2 formula, viz one which demarcates the universe of T1 within the universe of T2. The case before us, the interpretation of number theory within set theory, is an example. It is only the members of ω that are to be the "natural numbers" in our interpretation, not the entire universe - consisting of all sets - that our set theory talks about. The interpretation of the "universe of T1" is a formula $A_U(v_1)$ with only v_1 free. In the interpretation of Peano Arithmetic within ZF this formula should of course say that v_1 belongs to ω. We will give this formula as "$v_1 \in \omega$"; but of course, if the target language of our interpretation is our original, "minimal" language of set theory whose only non-logical constant is $\in$, then this formula must be seen as abbreviation of a much more complicated formula from which the "ω" has been eliminated, using the defintions by means of which it was introduced.

Intuitively, $A_U(v_1)$ should define a non-empty universe, i.e. the sentence $(\exists v_1)A_U(v_1)$ ought to be true. As for the unique condition on interpretations for function constants, we will impose this condition when we wil need it.

These preliminaries should suffice to make sense of the following definition:

Def. 1 Let L1 and L2 be first order languages. A *translation base for interpreting* L1 *in* L2 is a pair consisting of (i) a formula $A_U(v_1)$ of L2 with only v_1 free and (ii) a function which maps each non-logical constant C of L1 onto a formula $A_C(v_1, \ldots, v_k)$ of L2 in which $v_1, \ldots, v_k$

are the only free variables and where (a) if C is an n-ary predicate constant, then k = n and (b) if C is an n-ary function constant, then k = n+1.

Each translation base for interpreting L1 in L2 induces a function which maps arbitrary formulas of L1 onto formulas of L2, so that in particular sentences form L1 turn into sentences of L2. In case L2 has only predicate, but no function constants, the translation is quite straightforward: Basically all one needs to do to translate any formula B of L1 is to replace each atomic subformula $P(x_1, \ldots, x_k)$ by $A_P(v_1, \ldots, v_k)$ (making sure to rename bound variables where necessary). But when L1 contains function constants the matter is more complicated. For how are we to translate an atomic formula $P(t_1, \ldots, t_k)$ where all or some of the t_i are terms other than variables. To see what the problem is, consider once more the above interpretation of the union operation of the language of Boolean Algebras given in (1). Suppose we want to translate the formula

(3) $(x \cup y) \cup z = x \cup (y \cup z)$.

Here we have a predication involving the special predicate symbol = and two complex terms. Since (1) applies directly only to atomic formulas of the form $x \cup y = z$, there is no direct way in which it can be applied to (3). One way in which we can make it apply is to rewrite (3) into an equivalent formula in which all atomic subformulas are of the form to which (1) can be applied directly:

(4) (3) $\Rightarrow$

$(\exists u)(u = x \cup y \;\&\; u \cup z = x \cup (y \cup z)) \Rightarrow$

$(\exists u)(\exists v)(u = x \cup y \;\&\; v = y \cup z \;\&\; u \cup z = x \cup v) \Rightarrow$

$(\exists u)(\exists v)(\exists w)(u = x \cup y \;\&\; v = y \cup z \;\&\; w = x \cup v \;\&\; u \cup z = w)$

The last formula of (4) can now be translated by replacing its atomic formulas with suitable variants of (1).

An alternative way of dealing with this problem is to associate with each term t a formula $A_t(v_1)$ of L2 which represents t in the sense that, intuitively speaking, it is satisfied uniquely by the "value of the intrerpretation of t"; thus $A_t(v_1)$ serves as the translation of the formula $t = v_1$. As can be seen from Definition 2 below, the definition of $A_t(v_1)$ has teh reduction illustrated in (4) built into it.

Def. 2 Let T1 be a theory of the first order language L1 and T2 a theory of the first order language L2. Let $< A_U, I>$ be a translation base for interpreting L1 in L2.

1. $< A_U, I>$ is *suitable according to* T2 iff

(i) $T2 \models (\exists v_1)A_U(v_1)$

(ii) For each n-place function constant F of L1

(5) $$T2 \models (\forall v_1) .. (\forall v_n)(\forall y)(\forall z)(A_C(v_1, ... ,v_n,y) \ \& \ A_C(v_1, ... ,v_n,z) \rightarrow y = z)$$

2. Suppose that $< A_U, I>$ is suitable for T2. The *interpretations* of terms and formulas of L1 in L2 *based on* $< A_U, I>$ are defined as follows:

1. Terms. For each term t the interpretation of t based on $< A_U, I>$ is the formula $A_t(v_1)$ defined as follows

i. If t is the variable x, then $A_t(v_1)$ is $v_1 = x$

ii. If t is the term $F((t_1, ... ,t_n)$, then $A_t(v_1)$ is the formula $(\exists x_1).. (\exists x_n)(A_{t_1}(x_1) \ \& \ .. \ \& \ A_{t_n}(x_n) \ \& \ A_F(x_1, ... ,x_n,v_1))$

2. Formulas. For each formula B the interpretation of B based on $< A_U, I>$, I*(B), is defined by:

i. $I^*(P(t_1, ... ,t_n)) =$
$(\exists x_1)..(\exists x_n)(A_{t_1}(x_1) \ \& \ .. \ \& \ A_{t_n}(x_n) \ \& \ A_P(x_1, ... ,x_n))$

ii. $I^*(\neg B) = I^*(\neg B)$; $I^*(B \ \& \ C) = I^*(B) \ \& \ I^*(C)$; $I^*(B \vee C) = I(B) \vee I^*(C)$; $I^*(B \rightarrow C) = I^*(B) \rightarrow I^*(C)$; $I^*(B \leftrightarrow C) = I^*(B) \leftrightarrow I^*(C)$;

(iii) $I^*((\forall v_i)B) = (\forall v_i)(A_U(v_i) \rightarrow I^*(B))$;
$I^*((\exists v_i)B) = (\exists v_i)(A_U(v_i) \ \& \ I^*(B))$

3. The translation base $<A_U, I>$ is an *interpretation of* T1 *within* T2 iff (i) $<A_U, I>$ is suitable according to T2; and

(ii) For any sentence B of L1, if $T1 \models B$, then $T2 \models I^*(B)$.

N.B. If T1 is given by a set of axioms, then to check that 3.ii. is satisfied it suffices that each of these axioms translates into a theorem of T2.

We now turn to the interpretation of Peano Arithmetic in ZF Set Theory. After the general foundations we have just discussed, this is now quite straightforward. All we need to do is define a translation base for interpreting the language L_{PA} into the language of set theory $\{\varepsilon\}$, and then check that it satisfies the conditions (i) and (ii) of Def. 3.2.

In defining the translation base, we will continue with the convenient device of specifying the interpretations of the non-logical constants of Peano Arithmetic in the definitionally extended language of set theory we have been using. As with the formula $v_1 \varepsilon \omega$, an interpretation in the language $\{\varepsilon\}$ can be obtained from the fomula thus specified by elimination of the defined function constants and predicates.

The interpretation of the constants = and S is straightforward. But those of + and · require some thought. What needs to be done is to mimick the recursive defintions of + and · given by the Peano axioms PA3- PA6. We accomplish this by using the familiar trick of quantifying over finite functions which encode initial segments of the relevant recursion. Thus the interpretation of + has the following form.

(6) $(\exists f)(Fn(f) \;\&\; Dom(f) = v_2 \cup \{v_2\} \;\&\; f(\varnothing) = v_1 \;\&\; (\forall n)(n \varepsilon v_2 \rightarrow f(n \cup \{n\}) = f(n) \cup \{f(n)\}) \;\&\; f(v_2) = v_3)$

The function f defined in the quantifier-free part of (6) is intuitively the function which assigns to each of the numbers n from 0 to v_2 as values the numbers $v_1 + n$. This has the effect that in particular v_3 is the number $v_1 + v_2$. The interpretation of · is constructed along the same lines; the formula looks a little more complicated because the recursive clause for · makes use of +.

Def. 3 Translation Base for interpreting Peano Arithmetic in ZF:

(i) $A_U(v_1) := v_1 \varepsilon \omega$

(ii) $I(0) := v_1 = \varnothing$

(iii) $I(S) := v_2 = v_1 \cup \{v_1\}$

(iv) $I(+) := (\exists f)(Fn(f) \;\&\; Dom(f) = v_2 \cup \{v_2\} \;\&\; f(\varnothing) = v_1 \;\&\; (\forall n)(n \varepsilon v_2 \rightarrow f(n \cup \{n\}) = f(n) \cup \{f(n)\}) \;\&\; f(v_2) = v_3)$

(iv) $I(\cdot)$:= $(\exists f)(Fn(f) \;\&\; Dom(f) = v_2 \cup \{v_2\} \;\&\; f(\emptyset) = \emptyset \;\&$
$(\forall n)(n \in v_2 \rightarrow I(+)(f(n), n, f(n \cup \{n\})) \;\&$
$f(v_2) = v_3)$

Theorem. The translation base of Def. 3 is an interpretation of Peano Arithmetic within ZF, in the sense of Def. 2.3.

##

Let us call an interpretation of T1 within T2 *absolute* iff the first member of its translation base (i.e. the formula $A_U(v_1)$) is true of all things in the "universe of T2", that is, if T2 $\vdash (\forall v_1)A_U(v_1)$. The situation where there is an absolute interpretation of T1 in T2 can also be described as follows: For each non-logical constant C of T1 there is an explicit definition B_C of C in T2, such that, if T2 is the theory in the language L2 $\cup$ L1 which we obtain by adding all these definitions to T2, then T2' $\vdash$ T1.

An important relationship between theories T1 and T2 is when each is absolutely interpretable within the other. In such a situation T1 and T2 can be regarded as different formalizations of the same "conceptual structure" - whether one starts from the notions that are primitive in T1 (i.e. the non-logical constants of L1) or from those that are primitive in T2, the other notions can always be obtained from these by explicit definition so that the axioms of the other theory become theorems of the first. A very simple (and quite uninteresting) example is provided by the theory of partial order, which can be formulated either in the language $\{<\}$ with the axioms PO1 and PO2 above - let this theory be TPO1 - or in the language $\{\leq\}$, with the axioms (PO1') $(\forall x)(\forall y)\forall z)(x \leq y \;\&\; y \leq z \rightarrow x \leq z)$ and (PO2') $(\forall x)(\forall y)(x \leq y \;\&\; y \leq x \rightarrow x = y)$ - let this theory be TPO2. Then TPO1 is absolutely interpretable wihin TPO2 and TPO2 is absolutely interpretable within TPO1.

Exercise. Prove this by formulating definitions of $\leq$ in TPO1 and $<$ in TPO2 and then showing that each definition turns the axioms of one theory into theorems of the other.

A more interesting example is provided by the theory of groups. The formalization that we gave here, with $\cdot$ and $^{-1}$ as primitives, constitutes only one of many possibilities. Another version one often encounters in the literature starts with $\cdot$ and e as primitives and takes as axioms

(for instance) (i) $(\forall x)(x \cdot e = x)$; (ii) $(\forall x)(e \cdot x = x)$. (iii) $(\forall x)(\forall y)(\forall z)((x \cdot y) \cdot z = x \cdot (y \cdot z))$; (iv) $(\forall x)(\exists y)(x \cdot y = e)$.

It is not hard to show that (i) - (iv) entail that the y of (iv) is unique. (Argument: Suppose that $x \cdot y = e$ and that $y \cdot u = e$. Then $x = x \cdot e = x \cdot (y \cdot u) = (x \cdot y) \cdot u = e \cdot u = u$. So $y \cdot x = y \cdot u = e$. Now suppose that y and z are both such that $x \cdot y = e$ and $x \cdot z = e$. Then $y = y \cdot e = y \cdot (x \cdot z) = (y \cdot x) \cdot z = e \cdot z = z$.) So we may define $(\forall x)(\forall y)(x^{-1} = y \leftrightarrow x \cdot y = e)$. it is eaasy to check that with this definiton all axioms of the version of group theory given in the text follow from (i) - (iv) above.

Exercise. It is also possible to formulate the theory of groups with just one 2-place operation / as primitive. Intuitively x/y means the same as $x \cdot y^{-1}$.

(i) Show that if we add to our original formulation of the theory of groups (8) as additional axiom, then the sentences (9) - (12) are derivable as theorems

(8) $(\forall x)(\forall y)(\forall z)(x / y = z \leftrightarrow z = x \cdot y^{-1})$
(9) $(\forall x)(\forall y)(x/x = y/y)$
(10) $(\forall x)(\forall y)(x = x/(y/y))$
(11) $(\forall x)(\forall y)((x/x)/(x/y) = y/x)$
(12) $(\forall x)(\forall y)(\forall z)((x/y)/z = x/(z/((y/y)/y)))$

(ii) Let TG' be the theory given by (9) - (12). Show that the formulas (13) - (15) are definitions in TG' (i.e. show that the relevant existence and uniqueness conditions for the definientia of (13) - (15) are theorems of TG')

(13) $(\forall z)(e = z \leftrightarrow (\forall y)(z = y/y))$
(14) $(\forall x)(\forall y)(x^{-1} = y \leftrightarrow y = e/x))$
(15) $(\forall x)(\forall y)(\forall z)(x \cdot y = z \leftrightarrow z = x/y^{-1}))$

(iii) Show that all axioms of our original formulations of the theory of groups are derivable from (9) - (15).